NEW CENTURY SENIOR PHYSICS

Concepts in context

SECOND EDITION

Richard Walding

Greg Rapkins

Glenn Rossiter

OXFORD

Prelims-Walding 4th 25/8/04 1:55 PM Page i

NEW CENTURY SENIOR PHYSICS

Concepts in context

SECOND EDITION

Richard Walding

Greg Rapkins

Glenn Rossiter

OXFORD

OXFORD
UNIVERSITY PRESS

253 Normanby Road, South Melbourne, Australia

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Taipei Tokyo Toronto

OXFORD is a registered trade mark of Oxford University Press in the UK and in certain other countries

© Richard Walding, Greg Rapkins, Glenn Rossiter 2004

First edition published 1999
This edition published 2004

This book is copyright. Apart from any fair dealing for the purposes of private study, research, criticism or review as permitted under the Copyright Act, no part may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission. Enquiries to be made to Oxford University Press.

Copying for educational purposes
Where copies of part or the whole of the book are made under Part VB of the Copyright Act, the law requires that prescribed procedures be followed. For information, contact the Copyright Agency Limited.

The National Library of Australia Cataloguing-in-Publication entry for your forthcoming publication is as follows:

Walding, Richard.
New century senior physics: concepts in context.

2nd ed.
Includes index.
For year 11 and 12 students.

ISBN 0 19 551777 6.

1. Physics—Textbooks. I Rapkins, Greg. II. Rossiter, Glen. III. Title.
530

Typeset by Currency Communications Australia
Printed through Bookpac Production Services, Singapore

CONTENTS

Table of contexts and chapters		
Preface		vi
Acknowledgments		vii

UNIT 01 MEASUREMENT AND PHYSICAL QUANTITIES

CHAPTER 1	Measurement and physical quantities	2

UNIT 02 FORCES AND MOTION

CHAPTER 2	Motion in a straight line	26
CHAPTER 3	Vectors and graphing	56
CHAPTER 4	Forces in action	74
CHAPTER 5	Projectile, circular and periodic motion	108
CHAPTER 6	Astrophysics	134
CHAPTER 7	Hydrostatics: The physics of fluids	164

UNIT 03 ENERGY AND MOMENTUM

CHAPTER 8	Momentum	184
CHAPTER 9	Work and energy	211

UNIT 04 THERMAL PHYSICS

CHAPTER 10	Heat and temperature	240
CHAPTER 11	Heat and matter	254
CHAPTER 12	Heat transfer	271

UNIT 05 WAVE MOTION

CHAPTER 13	Wave motion in one dimension	284
CHAPTER 14	Wave motion in two dimensions	305
CHAPTER 15	Light — A wave?	320
CHAPTER 16	Sound, music and audio technology	348

UNIT 06 OPTICS

CHAPTER 17	Reflection of light	382
CHAPTER 18	Refraction	400
CHAPTER 19	Lenses	418
CHAPTER 20	Optical instruments	429

UNIT 07 ELECTRICITY AND ELECTRONICS

CHAPTER 21	Electrostatics	444
CHAPTER 22	Electric circuits	466
CHAPTER 23	Electronics	501
CHAPTER 24	Electronic systems	524

UNIT 08 MAGNETISM AND ELECTROMAGNETISM

CHAPTER 25	Magnetism and electromagnetism	548
CHAPTER 26	Electromagnetic induction	576

UNIT 09 ATOMIC AND NUCLEAR PHYSICS

CHAPTER 27	Atomic structure	596
CHAPTER 28	Nuclear physics	616
CHAPTER 29	Quantum physics and fundamental particles	654

UNIT 10 EXTENSION TOPICS

CHAPTER 30	Special and general relativity	680
CHAPTER 31	Designing practical electronic circuits	710
CHAPTER 32	Solar physics	731
CHAPTER 33	Medical physics	745

Answers	761
Glossary	770
Appendices	776
Index	783

This text has been written to support a variety of popular contexts. The following table shows the link between these contexts and the chapters that support them.

CONTEXT	CHAPTER
Amusement parks	1, 2, 3, 4, 5, 8, 9
Ancient technologies	1, 2, 4, 7, 8, 9
Atmospheric physics	7, 11, 12, 15, 16, 21, 32
Automobile’s electrical system	22, 23, 24, 25, 26, 31
Charge it!	21, 22
Designing practical electronic circuits	31
Electric effects on humans	22
Electronic systems	24
Fluids and floating	4, 7
Food and cooking	10, 11, 12
Forensic science	8, 9
Gravity and space physics	4, 6, 30
Heat and the environment	10, 11, 12
History of measurement	1
Household electricity	22, 26
Medical physics	33
Music and audio production	15
Nuclear bombs	27, 28
Nuclear power	27, 28
Optical instruments	20
Particle physics and cosmology	27, 29, 30
Revolutionary and landmark developments	30
Robotics	31
Scientific and industrial uses radioisotopes	27, 28
Sight and seeing	17, 18, 19, 20
Solar physics	32
Sport	1, 2, 4, 5, 8, 9
Sporting collisions and explosions	8, 9
Transport and safety	2, 3, 4, 8, 9

PREFACE

This is a fully revised edition of *New Century Senior Physics* and is designed to complement the 2004 Queensland Senior Physics Syllabus. The new syllabus is about learning in context. This book continues to provide a rich source of contextual detail as the key concepts are developed. The research literature suggests that concepts are best understood when they are presented in more than one context, and we have done this over and over again. We have also tried to maintain the arrangement of material so that teachers and students have little difficulty in finding what they want. This text will be a great resource for students and teachers alike as they seek to understand the world from a physics perspective.

Students

- Don't be alarmed at the amount of work in this book. There's more than enough for a two-year course. Your teacher will often be saying 'we're not doing this for assessment in our course'. The rest of the text you may well treat as extra background material or read just for your own interest.
- You might think that some questions are too easy or repetitive. We intended this. Expert problem solvers practise the easy work until it becomes automatic. Become that sort of person.
- You might also think that our worked examples are laborious. As you learn physics, you'll develop your own shortcuts. Remember — there is no one right way to solve a problem. Developing these techniques is what physics is about.
- If you get stuck, have a look at our web page on the Internet. You'll find worked solutions to selected questions. Find us at: *http://www.mbc.qld.edu.au/oxford/physics.html*

Teachers

- Choosing a text is the easy part; knowing what to put in your course is harder. This text should support most of the contexts you would want, as it is based on the most popular contexts chosen by teachers for their school work programs.
- Any suggestions are welcome from you or your students. Please e-mail us at school. The e-mail addresses are on the web page at www.mbc.qld.edu.au/oxford/physics.html
- Examples of a wide range of contexts, work programs, sample assessment tasks, discussion papers and networking opportunities can be found on the 2004 Senior Physics Syllabus web page at www.mbc.qld.edu.au/physics/sp.html
- We have included a huge range of questions and stimulus material, providing both practice and assessment opportunities for students. They include open and/or closed tasks inviting open or closed responses. Questions and tasks presented are suitable as practice and exemplars of **written tests**, **extended response tasks (assignments and stimulus-response items)** and **extended experimental investigations**.
- Please make your students aware of both web addresses. Students have found them very useful in the past.

Richard Walding, Greg Rapkins and Glenn Rossiter

ACKNOWLEDGMENTS

The authors would like to thank:

- their families for their help and perseverance during the production of this book
- students from dozens and dozens of schools who contacted us with comments about improvements to the text, queries about questions and alternative solutions
- colleagues on the QSA Science Subject Advisory Committee, the Physics Syllabus Sub-committee and the Physics State and District Panels for their thoughts on what a textbook should be like if it is to support their school's work program
- the physics teachers in the Trial Pilot schools whose discussions about choices of learning experiences and the development of interesting and useful contexts gave us great ideas for inclusion in this text.

The authors and publisher wish to thank the following authors and copyright holders for granting permission to reproduce their material:

Extract from 'Journal of the American Veterinary Medical Association', Vol. 191, No. 11, ©1987, reprinted by permission of the American Veterinary Medical Association; Extract from *Fundamentals of Physics 3rd Edition*, ©1988, by David Halliday and Robert Resnick, this material is used by permission of John Wiley and Sons Inc; Extract from *Advanced Physics 4th Edition*, ©1994, by Tom Duncan, published by John Murray, reprinted in adapted form by permission of Hodder Murray.

The publishers and authors would like to thank the following copyright holders for permission to reproduce copyright photographs:

ANT Photo Library/I. R. McCann, photo 17.3;
Dale Mann/Retrospect, photos 17.1, 17.2, 20.1;
Dale Mann/Retrospect and Physics Department, Glen Waverley Campus, Wesley College, photos 14.2, 14.3, 14.4, 14.5, 15.3, 18.1, 18.3;
Dick Smith Electronics, colour resistor code, centre colour pages;
Jan Curtis, images of the aurora, centre colour pages;
Kanehara & Co. Ltd, Tokyo, Japan, Ishihara colour blindness card, inside cover;
Magellan/Thales, 30.4;
NASA, photos 4.1, 6.2, 6.5, 6.6, 30.3, 30.6;
National Measurement Laboratory, photo 1.1;
Photodisc, inside covers;
Sport the Library/Sporting Pictures UK Ltd, 2.4 and centre colour pages;
Stock Photos, photos 7.3, 8.1, 12.2;
The Photo Library, Sydney, photos 6.3, 6.4, 7.2, 15.5 and inside cover, 18.2, 28.1, 28.2, 30.1, 32.1, 33.1;
The Picture Source/Terry Oakley, inside covers.

Every effort has been made to trace the original source of copyright material contained in this book. The publisher would be pleased to hear from copyright holders to rectify any errors or omissions.

KEY

Experimental Investigations

Stimulus Response

Non-experimental Investigations

Ch01-Walding 4th 25/8/04 9:19 AM Page 1

UNIT 01

Measurement & Physical Quantities

CHAPTER 01

Measurement and Physical Quantities

1.1 MEASUREMENT

Physics, like the other sciences, is all about explaining the natural world. Measurement is at its very heart. Ever since humans have been thinking about their place in the universe, they have been making measurements. Have you ever wondered about any of these:

- What would have been the first sort of measurement made by humans?
- When you use the unit of length, *foot*, whose foot was the standard?
- What is the shortest length of time that can exist? Is there no limit?
- Time passes but why can't it go backwards?
- Just how heavy is the universe? How did they weigh it?
- Is cream more dense than milk and, anyway, who invented *density*?

Questions like these have always intrigued people. As you study physics some of them will become clearer. But hopefully you will ask your own questions and make your own measurements, for this is what the study of physics is all about.

Activity 1.1 ESTIMATING

EI

1. Estimate the length of this page to the nearest millimetre. Now measure it. Were you over or under?
2. Now that you've had practice, estimate the length of this line:

   ____

   Were you any more accurate?
3. How far is it from the floor to the ceiling? Write down your estimate and then find the actual value.
4. Can you estimate 30 seconds? Look at your watch, cover it and uncover it when you think 30 seconds is up. Repeat it until you are accurate to within 1 second. How did you count off the seconds? How did others in the class count off the seconds?
5. How good are you at estimating mass? Estimate the mass of this book in grams without lifting it and then again after lifting it. Did you lift it up and down to estimate mass? Why?
6. Feel the thickness of one page of this book. How many pages do you estimate this book has? Check.

NOVEL CHALLENGE

Here are a few 'Fermi' questions (named after US physicist Enrico Fermi, who used to drive his students nuts with them).

A How quickly does hair grow?
B How many piano tuners are there in your capital city?
C How many ping-pong balls can you fit in a suitcase?
D How quickly does grass grow?

NOVEL CHALLENGE

The four compass directions North, East, South, West are derived from old foreign words. Can you match up the original meanings with the compass directions:

A Indoeuropean *wes* = Sun goes 'down'.
B Italian *nerto* = 'to the left' as one faces the Sun.
C German *suntha* = region in which the 'Sun' appears in the Northern Hemisphere.
D Indoeuropean *aus* = Sun 'rises'.

NOVEL CHALLENGE

If you were transported in a time machine to an unknown date in Australian history, how could you work out the date? See our Web page for some suggestions.

Estimating measurements is important. You can see whether answers are reasonable or nonsense if you have a feeling for some of the common units of measurement in physics. The three quantities you've measured in the activity are the most basic measurements in physics: **length, time** and **mass**. But your estimates probably differed from others in your class and that's why standards were developed. The importance of measurement grew as human societies became more complex.

The first measurement the earliest humans are believed to have used was the 'day'. Hence, the 'day' became the first unit of measurement, well before any concept of length or mass. Which unit do you think came next? Perhaps the 'month' — from one new moon or full moon to the next; and then perhaps the 'year' when people noticed that the Sun rose again in the same constellation of stars after many new moons.

Neanderthal burial sites from 50 000 years ago suggest that people were conscious of the past, the present and the future — something that most other animals are believed to be unaware of.

As humans have progressed, so too has their need for new units of measurement. The need for a unit comes before a unit is invented. Only recently have units like the *barn* been invented. The size of a nucleus as seen by a high speed atomic particle is as big as the side of a barn, hence the name. One barn equals 10^{-28} m^2. There was no need for this unit until Einstein produced the 'theory of relativity' and physicists applied it to atomic structure.

PHYSICAL QUANTITIES 1.2

There are a number of things in the world we want to measure. As well as the three mentioned above (length, time and mass), there are others, such as temperature, electric current and weight. These measurable features are called **physical quantities**. There are also some non-physical quantities, for example intelligence, beauty and personality, that are difficult-to-measure. Attempts have been made to devise measurements for quantities such as these but have always ended up in disagreement and, in many cases, failure.

The **international system** of units called **SI** (from the French name for the system, *Système International d'Unités*), is now commonly used around the world. It is sometimes called the **metric system** (from the Greek *metron* to 'measure').

The seven fundamental (or base) units of this system are shown in Table 1.1.

Table 1.1 SI UNITS

PHYSICAL QUANTITY	SYMBOL OF QUANTITY	NAME OF UNIT	SYMBOL FOR UNIT
Length	l	metre	m
Mass	m	kilogram	kg
Time	t	second	s
Electric current	I	ampere	A
Temperature	T	kelvin	K
Amount of substance	n	mole	mol
Luminous intensity		candela	cd

To get multiples of the base units, prefixes are added. Table 1.2 lists some of these prefixes that will be used throughout your physics course. You should remember from nano to mega. Check with your teacher if you need any others.

Table 1.2 PREFIXES

PREFIX	SYMBOL	MEANING	VALUE	FACTOR
Pico	p	one million-millionth	0.000 000 000 001	10^{-12}
Nano	n	one thousand-millionth	0.000 000 001	10^{-9}
Micro	μ	one millionth	0.000 001	10^{-6}
Milli	m	one thousandth	0.001	10^{-3}
Centi	c	one hundredth	0.01	10^{-2}
Deci	d	one tenth	0.1	10^{-1}
Kilo	k	one thousand	1 000	10^{3}
Mega	M	one million	1 000 000	10^{6}
Giga	G	one thousand million	1 000 000 000	10^{9}
Tera	T	one million million	1 000 000 000 000	10^{12}

Example of using a prefix with a unit: 1 millimetre = 10^{-3} metre = 0.001 metre.

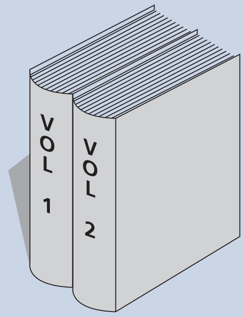
Rarely used prefixes are:

- 10^{-15} femto (f) — radius of a proton is 1 fm
- 10^{-21} zepto (z) — charge on the electron is 160 zC
- 10^{-24} yocto (y) — mass of the hydrogen atom is 1.66 yg
- 10^{-27} xenno (x) — magnetic moment of a proton is 14 xJ T^{-1}
- 10^{21} zetta (Z) — distance to Andromeda galaxy is 20 Zm
- 10^{24} yotta (Y) — mass of the Earth is 5977 Yg.

Others you'd never use are vendeko (v) 10^{-33} and vendeka (V) 10^{33}. Can you think of any practical use of these prefixes? Mathematicians also use the term *googol* to represent 10^{100} and *googolplex* for 10 raised to the power of a googol: $10^{10^{100}}$. The biggest number in the world (apart from infinity) is Grahams' number. If all the material in the world was turned into paper there still wouldn't be enough paper to write it down. Now that's big!

NOVEL CHALLENGE

You have two 100-page volumes of a dictionary on your shelf. A worm eats its way from Volume 1 page 1 through to Volume 2 page 100.
How many pages does it eat through?



Standards

Standards have to be agreed upon for units to be useful throughout the world. For instance, the temperatures in different countries couldn't be compared until a universal temperature scale was devised. The following shows how some of these units have developed.

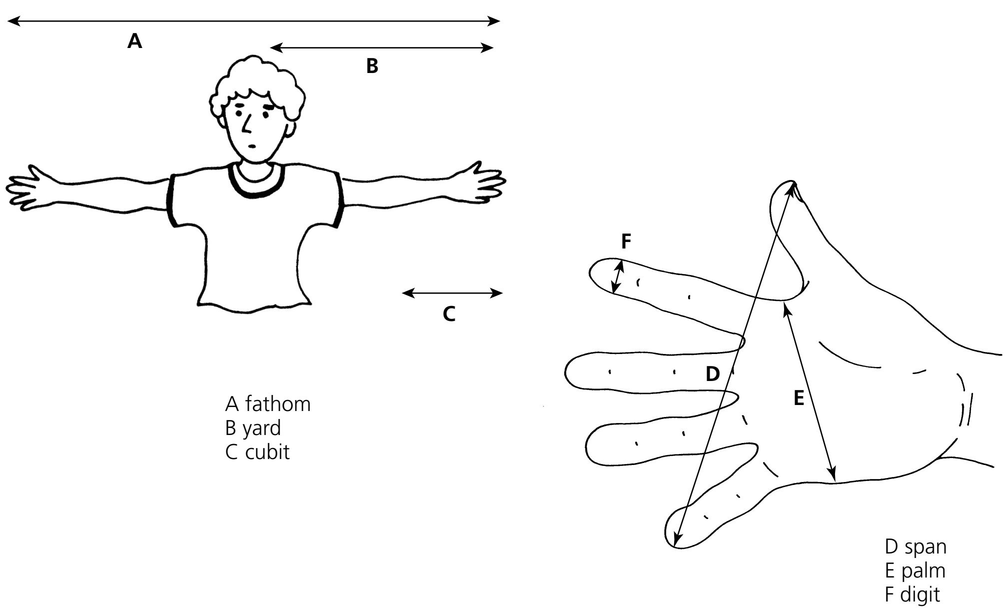
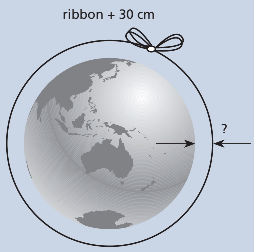


Figure 1.1
Body measurements.

Ch01-Walding 4th 25/8/04 9:19 AM Page 5

NOVEL CHALLENGE

Consider the Earth to have a circumference of 40 000 km and a ribbon to be put tightly around it. If you cut the ribbon and inserted a 30 cm piece, how far would the ribbon be from the earth if it was evenly spaced?



Length

As with most early units, people used the most convenient measures — themselves. The length of a foot or a stride was a convenient measure. So was the span of a hand or the thickness of a thumb. But as civilisations grew, these ways of measuring became inadequate. How could a foot be used as a measure when one person's foot was so much longer than another's? Hands and thumbs were different too. In ancient times a measure that was used in one country was often later adopted by others through trade or invasion. Roman measures spread throughout Europe, Asia, England and Africa as the Romans conquered and occupied these lands but gradually, through mistakes in copying and figuring, the standards became so confused that most of them dropped out of use. By the sixteenth century most people in Europe had returned to the old body measurements and we still use some of these today.

The shortest unit of length was the *digit*, the width of a finger, or three-quarters of an inch. An inch is the width of a thumb; a hand is four inches and the span is nine inches. To try to standardise these units, Edward II of England ruled that one inch 'shall be equal to three grains of barley, dry and round, placed end to end lengthwise'.

The *foot* was about 11.5 inches in Greece, 12 inches in England and other English speaking countries, and 11 to 14 inches anywhere else. The earliest attempt to standardise the foot was in 2100 BC when it was decreed that a foot was the length of the foot of the statue of the ruler of Gudea of Lagash in Babylonia. It was 10.41 inches long and divided into 16 parts.

The *pace* was another common measure. It was about 5 feet — the length of two complete steps. Roman soldiers paced off the miles as they marched. A thousand paces made up a mile, just a little less than the modern mile, which is 5280 feet. Now we measure a pace as a single step — about 2.5 feet.

Lastly, the *yard*. The yard was defined in two ways: in northern Europe it was the length of an Anglo-Saxon's belt whereas in the south it was a double cubit. A cubit is 18 inches — the distance from the elbow to the wrist. Henry I, at the beginning of the twelfth century, fixed the yard as the distance from his nose to the thumb of his outstretched arm.

The metric system (SI) was invented by the French in 1790, following the French Revolution. It was a part of a plan for a new beginning, a whole new social and economic life in France, without any ties to the past. The metre, the basic unit of length, was supposed to be one ten-millionth part of the distance from the North Pole to the Equator. But in the eighteenth century instruments were not as accurate as they are today, so there was a measurement error. By the time the error was realised, the metre was so well established at 39.37 inches that it was left at that.

A platinum–iridium bar exactly this distance long was made and this became the standard for the metre. In 1960 the standard metre was redefined to be the length equal to 1 650 763.73 wavelengths in a vacuum of the red-orange light emitted by the krypton-86 atom. Since 1983, however, the metre has been redefined as the length of the path travelled by light in a vacuum during a time interval of 1/299 792 458 of a second.

NOVEL CHALLENGE

Quick now — is a physics lesson longer or shorter than a microcentury?

Table 1.3 SOME LENGTHS

LENGTH	METRES	LENGTH	METRES
To furthest quasar	10^{26}	Thickness of a page	10^{-4}
To nearest star	10^{16}	Radius of H atom	10^{-10}
To Pluto	10^{13}	Radius of a proton	10^{-15}
Radius of Earth	10^{7}		

Time

The first way of measuring time was to keep a record of the repetition of natural events. From sunrise to sunrise was the most fundamental of periods as it was so easy to measure and hence the day became the first unit of time. We do not know how long ago people started using the idea of days but it would certainly have been tens of thousands of years ago.

People realised that the Sun and the Moon were the best timekeepers of all. They called the time taken for the Earth to make one orbit of the Sun a year. We now know it to be 365 days, 5 hours, 48 minutes and 45.7 seconds long. The extra hours, minutes and seconds are collected together every 4 years to make the additional day we have in a leap year. To keep the timing more accurate, the start of a century is classified as a *non*-leap year.

The length of a day is fixed by the Earth's rotation on its axis. But at different times of the year the length of the day varies from place to place so we have to take this into account. What we end up with is the mean solar day but its duration is now standardised in terms of the hour, minute and second.

It was the ancient Babylonians who divided their measurements into sixty parts and we have kept their divisions for the hour and minute. A minute is 60 seconds and an hour is 60 minutes. The fundamental unit, the second, is now defined as the time for 9 192 631 770 vibrations of light (of a specified frequency) emitted by a caesium-133 atom. In principle, two caesium clocks would have to run for 6000 years before they differed by 1 second. In practice, atomic clocks do better than that. The latest Hewlett Packard 5071A caesium clock achieves an accuracy of 1 second in 1.6 million years and only costs about $90 000. Some experimental clocks are within 1 second in 30 billion years. Every physics lab should have one.

But a fundamental question about time has always bothered physicists. What does the passage of time mean? What is the difference between the past and the future apart from the passage of time? Nobel prize-winning physicist Richard Feynman said, 'We physicists work with time every day but don't ask me what it is. It's just too difficult to think about'.

Table 1.4 SOME TIME INTERVALS

TIME INTERVAL	SECONDS
Lifetime of a proton	10^{39}
Age of the universe	10^{18}
Age of the pyramid at Cheops	10^{11}
Human life expectancy	10^{9}
Length of a day	10^{5}
Physics lesson	10^{4}
Lifetime of a muon	10^{-6}
Shortest time interval known*	10^{-43}

* This is known as Planck time — the earliest time after the 'Big Bang' at which the laws of physics as we know them can be applied.

Mass

Measurements of mass and weight came a long time after measurements of length and time. An early way of thinking about weight was the amount a person could carry. At first, people compared weight by balancing small objects, one in each hand, and estimating whether one was heavier than another. About 7000 years ago, the Egyptians devised a crude scale — a stick hanging by a cord tied around its middle acting as a balance. By 3000 BC small stone weights were used as a comparison, but as trade developed, different weights were used for different objects. Honey, medicine and metal all had different units of weight, many of which have persisted into modern times. For example, the avoirdupois system of weights includes ounces, pounds and the ton. Grain was measured in bushels; liquids were measured in pints and gallons or in the case of oil, in barrels. But no mention in early history has been made of the quantity known as mass.

Mass and weight are different quantities but people use them as if they mean the same thing. In Chapter 4 you will see the difference. Mass is a measure of an object's resistance to motion when being pushed or pulled. A 1 kilogram mass will be just as hard to push around no matter where in the universe it is. Weight, on the other hand, is a measure of the force of gravity acting on an object and will vary depending on how strong gravity is in that place. But weight has always been the quantity people have associated with heaviness; after all,

NOVEL CHALLENGE

In the first paragraph of Charles Dickens' *The Pickwick Papers* he states that he was at the bottom of a deep well and could see the stars in the daytime. Aristotle made the same claim in *On the Generation of Animals* in 350 BC.
Is this possible? Propose points for and against this idea. See the Web page for an answer.

PHYSICS UPDATE

The Time Service Department, US Naval Observatory, Washington, DC provides time signals for use throughout the USA and other parts of the world. You can access their clock on the Internet at http://tycho.usno.navy.mil/what.html and even set your computer's clock against their master signal.

NOVEL CHALLENGE

The world is broken up into many time zones based on the longitude of the various regions. Queensland is 1/2 hour ahead of South Australia, for instance.
But what time zone is the South Pole? No emailing Casey Station to find out.

PHYSICS FACT

The word 'hour' comes from the Greek word meaning 'season'. The length of daylight depends on the season. The word 'day' comes from the Saxon word 'to burn', referring to the hot days of summer.

NOVEL CHALLENGE

Under the system of measurement adopted during the reign of Queen Elizabeth I:

1 mouthful	= 1 cubic inch
1 handful	= 2 mouthfuls
1 jack	= 2 handfuls
1 gill	= 2 jacks
1 cup	= 2 gills
1 pint	= 2 cups
1 quart	= 2 pints

If 1 cubic inch = 14.7 mL, how many cups to 1 litre?

gravity is fairly constant all over the world and it hasn't been until the twentieth century that humans have left the Earth to go into space. More importantly though, the concept of mass was not developed until the 1680s when English scientist Isaac Newton proposed a relationship between force and acceleration that profoundly affected the new science of mechanics. This idea will be developed further in Chapter 4.

Mass is a fundamental quantity and the *kilogram* has been adopted as the fundamental unit of mass in the SI or metric system. The standard kilogram is a platinum–iridium cylinder kept at the International Bureau of Weights and Measures near Paris. Accurate copies have been sent to other standardising laboratories in other countries and the masses of other bodies can be determined by balancing them against a copy.

Photo 1.1
The standard kilogram.

Table 1.5 SOME MASSES

OBJECT	KILOGRAMS
Universe	10^{53}
Our galaxy	10^{41}
Sun	10^{30}
Moon	10^{23}
Ocean liner	10^{8}
Human	10^{2}
Grape	10^{-3}
Speck of dust	10^{-9}
Penicillin molecule	10^{-17}
Uranium atom	10^{-26}
Proton	10^{-27}
Electron	10^{-30}
Neutrino	10^{-30}
Uranium atom	10^{-26}

NEI Activity 1.2 BIGGEST, LONGEST AND OLDEST

Use the *Guinness Book of Records*, the Internet or an encyclopaedia to find out the following facts about units of measurement:

1. The highest artificial temperature on Earth was in a fusion reactor in the USA in 1994. How hot did it get?
2. How long can people go without food or water? Has anyone made it past 18 days?
3. What are the masses of the heaviest man and woman ever recorded? How many times greater than that of the lightest person are they?
4. Gold is the most ductile element known — it can be drawn into a very fine wire. How many metres of wire can be produced from 1 g of gold?

NOVEL CHALLENGE

This book is printed on paper classified as 80 gsm (80 grams per square metre). The cover is made of 249 gsm paper. What should the mass of this book be? Check it and see. What went wrong?

Derived units

New quantities can be made up of the base quantities. These are called **derived quantities**. For example, you can have combinations of the base units, such as metres per second and cubic metres or you can have derived quantities that have been given specific names, such as newton, coulomb and watt.

Table 1.6 lists some derived quantities.

Table 1.6 SOME DERIVED QUANTITIES

DERIVED QUANTITY	UNIT	SYMBOL FOR UNIT
Acceleration	metre per second^2	m s^{-2}
Angle	radian	rad
Area	metre^2	m^2
Capacitance	farad	F
Density	kilogram per metre^3	kg m^{-3}
Electric charge	coulomb	C
Energy	joule	J
Force	newton	N
Frequency	hertz	Hz
Momentum	kilogram-metre per second	kg m s^{-1}
Potential difference	volt	V
Power	watt	W
Pressure	pascal	Pa
Resistance	ohm	Ω
Velocity	metre per second	m s^{-1}
Volume	metre^3	m^3

Activity 1.3 WORKING SCIENTIFICALLY

Physicists spend their professional lives investigating relationships between physical quantities. In 1665 Robert Hooke described the relationship between the length of a spring and the stress (force) applied to it. Currently, physicists are trying to work out how the fundamental forces of nature are related to each other. In the three activities that follow, you will set up some experiments, collect data and look for relationships between some of the physical quantities mentioned earlier.

Part A: The bent ruler

1. Clamp a 30 cm steel ruler to the edge of a bench leaving most of it overhanging. Measure the distance from the floor to the tip of the ruler. (See Figure 1.2.)
2. Add a 50 g mass to the tip and record how many centimetres the ruler has bent from its unladen position. This is called its **displacement**.
3. Add another 50 g mass and record the total displacement. Then another 50 g and so on until 300 g has been added.
4. Plot a graph of displacement (y-axis) versus mass added (x-axis).
   (a) Does the graph go through the origin (0, 0)? If so, why?
   (b) How many centimetres does the ruler bend per 100 g added? Express this as cm per g.
   (c) Show how the graph would look if you: **(i)** used a thicker ruler; **(ii)** used a wider ruler; **(iii)** used a plastic ruler; **(iv)** allowed only 20 cm to overhang; **(v)** used a frozen ruler; **(vi)** used a steel ruler rapidly heated and cooled (annealed); **(vii)** used a steel ruler heated and cooled slowly. Try it! For all graphs you should provide a theoretical justification of your prediction.

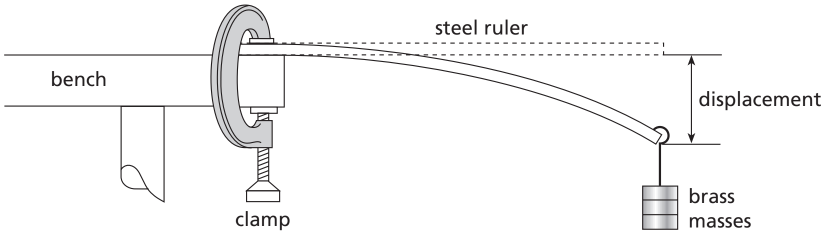


Figure 1.2
The bent ruler set-up for Part A.

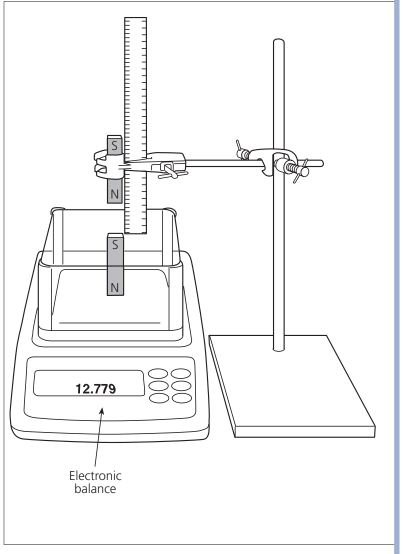
PHYSICS FACT

On his fourth voyage to the New World, Spanish explorer Christopher Columbus was marooned in Jamaica and, after a while, the local Indians refused to provide food. He knew that there would be an eclipse at noon on 29 February 1504 so he summoned the chiefs aboard and told them that unless they gave him food God would blacken the sky. When they refused, the sky went black right on time and when they relented he 'made' the sky go back to normal. Then the Spanish began the systematic plunder and destruction of an entire civilisation. Ah, no wonder science gets a bad name.

NOVEL CHALLENGE

German researcher Günther Bäumler found that people with the surname Smith (Schmidt in German) had, on average, a body mass that was 2.4 kg greater than people with the name Taylor (Schneider).
How would you test his findings in your school? What other variables would affect your results? Check our web page for why this difference occurs.

Figure 1.3
Set-up for measuring the relationship between magnetic force and distance (Activity 1.3, Part B).

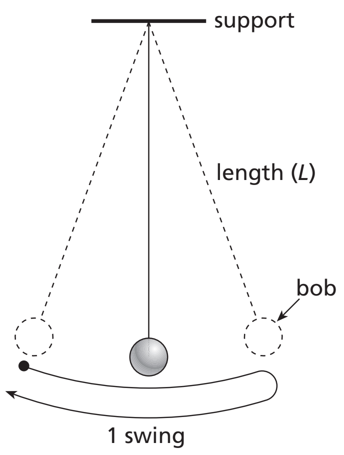


Part B: Magnetic personality

The force between two magnets varies with separation distance.

1. Place a bar magnet vertically upright on the pan of an electronic balance. (See Figure 1.3.) Zero the balance.
2. Place another magnet in a clamp above the first magnet so that the unlike poles face each other. There will be an attractive force, so the scale reading on the balance should be a negative value.
3. Start with the end of the clamped magnet 50 cm from the magnet on the balance and record a reading. If it is not zero, start with a 1 m separation (hold it in your hand).
4. Reduce the separation distance (*d*) by 5 cm at a time (or less if you like) and take balance readings.
5. Plot the data with separation distance (in cm) on the *x*-axis, and scale reading (grams) on the *y*-axis. If you are keen, convert the separation distance to metres; and convert the scale reading to force in newtons (N) by dividing it by 1000 and multiplying by 9.8.
6. Some questions:
   (a) When the distance was halved (from 50 cm to 25 cm), by what factor did the scale reading increase?
   (b) Would you get the same results if you put the magnets into repulsion?
7. Now try plotting $1/d^2$ on the *x*-axis against the scale readings. Did something magical occur?

Figure 1.4
Characteristics of a pendulum.



Part C: Let him swing!

Three variables you could change about a pendulum are the length, the mass and the distance through which it swings. (See Figure 1.4.) Using a lead fishing sinker or a brass mass tied to a metre of fishing line, construct a pendulum and measure the time for one swing at six different lengths. Keep the mass constant. Plot a graph. Keep your data for Chapter 3.

Converting units

It is often important to convert from one unit to another: for instance, from millimetres to metres or from pounds to kilograms. Two types of conversions are involved:

- From one SI unit to another.
- From a non-SI unit to an SI unit.

The first type will be needed when data are given in one particular unit but the answer has to be given in another form. This might occur when some constant is involved that is in a unit different from that of the data given. For example, if you had to calculate how far you would travel in 10 minutes at a speed of 5 metres per second then you would convert 10 minutes to seconds ($10 \times 60 = 600$) and multiply this number of seconds by the speed ($600 \times 5 = 3000$ metres).

Example

Imagine you have made measurements of a block of wood in a density experiment and need to calculate its volume in cubic metres. Length 35 cm, breadth 2.0 cm, width 1.5 cm.

Step 1: Convert the measurements to SI units (metre):

- length = 35 cm = $35 \times 1 \times 10^{-2}$ m = 0.35 m (3.5×10^{-1} m)
- breadth = 2.0 cm = $2.0 \times 1 \times 10^{-2}$ m = 2.0×10^{-2} m
- width = 1.5 cm = $1.5 \times 1 \times 10^{-2}$ m = 1.5×10^{-2} m.

Step 2: Calculate the volume:

- volume = 0.35 m $\times$ 2×10^{-2} m $\times$ 1.5×10^{-2} m
  = 1.05 m $\times$ 10^{-4} m^3.

Some other simple examples are:

- 25 000 cm = 250 m (2.5×10^2 m)
- 23 km = 23 000 m or 2.3×10^4 m
- 6 hours = 21 600 s or 2.16×10^4 s.

The other type of conversion is from a non-SI unit to an SI unit. This could occur, for instance, when data come from another source such as from some domestic measurement; from another country or from data taken in the past. The United States has yet to adopt SI units for daily use although all science, engineering and medical units throughout the world have been changed to SI. For instance, you may have to convert the mass of a person from pounds to kilograms. The conversion factor is 1 kg = 2.204 622 341 pounds. (See Table 1.7.)

Table 1.7 SOME NON-SI CONVERSION FACTORS

NON-SI UNIT	SI UNIT
Inch (in)	2.54 centimetres
Yard (yd)	0.914 401 8 metre
Gallon (gal)	4.546 litres
Pound mass (lb)	0.453 592 37 kilogram
Pound weight (lb)	4.45 newtons
Mile (mi)	1.609 kilometres
Acre (ac)	0.404 687 hectare
Pound per square inch (psi)	6896 pascals
Horsepower (Hp)	746 watts

Questions

1. From the following, select **(a)** two fundamental quantities; **(b)** two fundamental units; **(c)** two non-SI units: yard, luminous intensity, ampere, year, minute, temperature, force, second, pressure.
2. Convert the speed of light (3×10^8 m s^{-1}) to **(a)** km h^{-1}; **(b)** miles per hour.
3. Convert the following: **(a)** 10.3 m to cm; **(b)** 1.25 cm to m; **(c)** 1120 cm to m; **(d)** 143 367 mm to m; **(e)** 1.8 mm to m; **(f)** 14 cm^2 to m^2; **(g)** 4.8 cm^3 to m^3.
4. **(a)** Japanese sumo wrestlers have to be a minimum of 5 feet 8 inches tall. How many centimetres is this? (1 foot = 12 inches and 1 inch = 2.54 cm.)
   (b) The heaviest baby ever born was 23 lb 10 oz. If there are 16 ounces (oz) in 1 pound (lb) and 1 pound equals 0.454 kg, convert the baby's mass to kg.

1.3 SCIENTIFIC NOTATION

Things in the world are not always human-sized. Some are very small; some are huge. The numbers used to express these measurements can get messy. For example, the time taken for light to travel from one side of an atom to the other is about one billion billion billion billionths of a second. The mass of the Sun is two thousand billion billion billion kilograms. In his book *A Brief History of Time*, Stephen Hawking mentions that the publisher told him not to use any numerals. All numbers had to be spelt out because, it was argued, people

NOVEL CHALLENGE

In his 1997 book *Number Sense*, Stanislas Dehaene reported that his tests on brilliant scientists in France showed that it took them longer to say whether 6 was greater than 5 than it did to say whether 9 was greater than 5. Propose a testable hypothesis that could be investigated.

NOVEL CHALLENGE

People shrink in height not only as they get older, but also during each day. Some of our students shrink by $1\frac{1}{2}$ cm between first and last lesson. What is the reason for this? Can you find factual support for your suggestion? Do you think taller people shrink more than shorter ones? Does everyone shrink by a certain percentage? Do younger and older people shrink by the same percentage?

PHYSICS FACT

A very old unit of length was cubit—the length of the arm from elbow to fingertips. It comes from the Latin *cubitum*, meaning 'elbow'. The Egyptian 'royal cubit' was 542 mm long, and a master cubit of black granite was kept in a royal vault. All the cubit sticks in use in Egypt were measured at regular intervals. For example, the Great Pyramid of Giza was 280 royal cubits (RC) high. Other cubits include the biblical cubit of 457.2 mm.

NOVEL CHALLENGE

Humans have 10^{14} cells at a diameter of 0.01 mm each. If they were placed in a line, how many times around the Earth would they go? (The radius of the Earth is 6.38×10^6 m.)

NOVEL CHALLENGE

A brochure for the 1963 Ford Falcon said it averaged 26 miles per gallon of petrol. A 2002 Falcon is reported to use 12 L of petrol per 100 km.

(a) Which is the more economical?

(b) Develop a formula to convert mpg to L/100 km.

(c) In 1963, the standard Falcon engine had a capacity of 170 cubic inches, whereas the 2002 Falcon has a 4.5 litre engine. Which is the bigger?

couldn't understand exponents and wouldn't buy a book with them in. So, the speed of light appears as three hundred million metres per second.The time after the 'Big Bang' that it took for electrons to be created was a thousand billion billion billion billion billionths of a second. You probably know of a simpler way of expressing these values.

A shorthand means of expressing such numbers is called exponential notation. For example:

- 1 million (1 000 000) is written as 10^6.
- 1 billion (1 000 000 000) is written as 10^9.
- 1 millionth (1/1 000 000 or 0.000 001) is written as 10^{-6}.
- 1 billionth (1/1 000 000 000 or 0.000 000 001) is written as 10^{-9}.

Exponents tell us how many times 10 must be multiplied together and hence give the number of zeros. The expression 10^3 means 10 multiplied by itself three times ($10 \times 10 \times 10$); in other words, 1 with three zeros following it (1000).

When writing numbers using exponents, it is common practice to use scientific notation. This involves the following conventions:

- Write numbers in exponential notation with just one numeral before the decimal point, that is, the Earth–Moon distance of 382 million kilometres could be expressed as 382×10^6 km or in scientific notation as 3.82×10^8 km.
- Leave numbers between 0.1 and 100 as they are. There is no need to express 60 seconds as 6.0×10^1 s although you should be guided by your teacher on this matter.

Example

Write the following in scientific notation:

(a) The speed of light — three hundred million metres per second.
(b) The diameter of a red blood cell — 2 millionths of a metre (0.000 002 m).

Solution

(a) Three hundred million is 300×10^6 so the speed of light can be written as 3.00×10^8 m s^{-1}.
(b) 0.000 002 is written as 2.0×10^{-6}.

As you can see, with scientific notation only one numeral appears before the decimal place. The exponent has to be adjusted to allow for this. For example, when the number 300×10^6 became 3.00×10^8, the decimal point in 300 was shifted two places to the left (made smaller) to become 3.00, so to compensate, the exponent has to be increased by two units from 10^6 to 10^8 (made bigger).

Negative exponents are used to indicate numbers less than unity. For example, an electron has a mass of 0.000 549 units. To make this 5.49, we have to shift the decimal point four places to the right (made bigger by 10 000), so an exponent has to be included that compensates for this. In scientific notation it would be 5.49×10^{-4} units.

Further examples

(a) The radius of the Earth is 696 million metres or 6.96×10^8 m.
(b) The diameter of Saturn is 120 thousand kilometres or 1.20×10^5 km.
(c) The diameter of an atom is 0.000 000 000 1 m or 1.0×10^{-10} m.

Care must be taken when entering numbers in exponential notation in a calculator. On a 'scientific' calculator, to enter the number 6.96×10^8, press the buttons 6.96 EXP 8. The display should read 6.96^{08}. A common error is to enter this as 6.96×10 EXP 8. This is wrong. The display would read 6.96^{07}, which means 6.96×10^7. To enter an exponent such as 10^4 by itself you have to imagine that it means 1×10^4 and enter it as such. Remember, the EXP button symbolises the base, which is 10. It is one of the most common mistakes students make and a sure cause of lost marks in tests. When entering negative exponents, the +/– button is pressed after the exponent. A graphing calculator is different, but the problem is the same.

In some computer languages, exponents can be written in a different form. A number such as 6.96×10^7 would be written as 6.96E7 where the E stands for 'exponent'. With an exponent of 10^{-7} this would be written as 6.96E–7.

Questions

5 **(a)** Which is the larger out of **(i)** 'one hundred thousandths of a second' and **(ii)** 'one one hundred thousandth of a second'? **(b)** How can you make it clear whether you are talking about $\frac{100}{1000}$ s or $\frac{1}{100\,000}$ s when you are expressing these numbers in words? **(c)** Write both numbers in scientific notation.

6 Write the following in scientific notation: **(a)** 0.000 552; **(b)** 73 000 000; **(c)** one and a half million; **(d)** 0.000 250.

7 Work out the following on your calculator: **(a)** $1.2 \times 10^{-3} \times 2.2 \times 10^{-4}$; **(b)** $1.8 \times 10^{3} \div (6.4 \times 10^{-8})$.

8 Calculate the volume of an atom of diameter 0.000 000 001 m.

1.4 SIGNIFICANT FIGURES

When you say that it's 100 metres to the shop you are not really saying that this is the distance to the nearest metre. You are being approximate. You have not measured it — it could be 80 m or it could be 150 m. But the distance between the start and finish of a 100 m sprint race has to be 100 m and this has to be to the nearest centimetre. How would you write these distances? They are both 100 m.

It is common practice in science to record all integers that are certain and one more in which there is some uncertainty. The integers known with certainty plus the next figure are called **significant figures** (sf). Imagine you used a metre ruler marked in centimetres and measured the width of a book as 30.4 cm. This number has three significant figures. The first two integers are measured with certainty whereas the third is a mental estimate. The number could also be written as 0.304 m. It still has three significant figures — the first zero is only there to emphasise the location of the decimal point. Imagine you used the same ruler and measured the thickness of a book to be 6.3 cm. There are two significant figures — the .3 cm part is only a best guess, a mental estimate. In metres, this would be written as 0.063 m. There are still only two significant figures — the two zeros only indicate the position of the decimal point and are not significant.

Consider a ruler marked in millimetre divisions as your own ruler probably is. If you drew a line of length 10 mm, you would be drawing a line somewhere between about 9.5 mm and 10.5 mm in length — you probably can't be more accurate than that. In this case, the 1 is significant whereas the 0 is the next uncertain digit. There are two significant figures. But should it be written as 0.01 m or 0.010 m? You should write it as 0.010 m to show that the zero following the 1 is significant but the first two zeros only indicate the decimal point.

If exactly 35 000 tickets to a football grand final were sold then there are five significant figures. This is an exact figure, accurate to the last ticket, so all zeros are significant. In scientific notation it would be written as 3.5000×10^{4} tickets. If a crowd commentator estimated the crowd size as 'thirty-five thousand' then the figure is probably an estimate to the nearest thousand. It might be written in the paper as 35 000 but there are only two significant figures — the three zeros are not significant but are there to indicate where the decimal point is located. In scientific notation this would be written as 3.5×10^{4} people. Sometimes significant zeros are indicated with a small bar above the numeral. The exact figure of $35\,00\bar{0}$ people has a bar above the final zero whereas the commentator's estimate doesn't. If the crowd was estimated to the nearest hundred it could be written as $35\,\bar{0}00$, which indicates three significant figures. In scientific notation this would be written as 3.50×10^{4}. This is a better way to specify significant figures.

Rules

- All non-zero figures are significant: 3.18 has three sf.
- All zeros between non-zeros are significant: 30.08 has four sf.
- Zeros to the right of a non-zero figure but to the left of the decimal point are not significant (unless specified with a bar): 109 000 has three sf.

NOVEL CHALLENGE

Famous biologist Charles Darwin described the size of a canary finch in one of his notebooks as $3\frac{32}{64}$ inches long.

Why didn't he just write $3\frac{1}{2}$ inches? Convert the original measurement to centimetres using the correct number of significant figures.

NOVEL CHALLENGE

In the English translation of a manual on violin playing by the great Hungarian–German teacher Carl Flesch, budding violinists were told to 'lift your fingers 0.3937 inches from the fingerboard'.

Why is this funny? What do you suppose the original measurement was? Rewrite the inches measurement with the correct number of significant figures.

NOVEL CHALLENGE

In a shop in North Walsham, Norfolk, the height restriction to its carpark is written as 2300 mm.

Is there anything wrong with this? Explain!

- Zeros to the right of a decimal point but to the left of a non-zero figure are not significant: in 0.050, only the last zero is significant; the first zero merely calls attention to the decimal point.
- Zeros to the right of the decimal point and following a non-zero figure are significant: 304.50 has five sf.

Some examples of the application of these rules are given in Table 1.8.

NOVEL CHALLENGE

At a dinosaur exhibit at the Queensland Museum, the attendant said the Muttaburrasaurus was 30 million and 20 years old. 'How can you be that accurate?' asked a student. 'Well I was told it was 30 million years old when I started work here and I've been here 20 years.'
How would you explain to the attendant the folly of his statement?

Table 1.8 EXAMPLES

NUMBER	NUMBER OF SIGNIFICANT FIGURES	SCIENTIFIC NOTATION
0.003 5	2	3.5×10^{-3}
0.003 50	3	3.50×10^{-3}
0.35	2	3.5×10^{-1}
3.5	2	$3.5\ (\times 10^{0})$
3.50	3	$3.50\ (\times 10^{0})$
35	2	3.5×10^{1}
350	2	3.5×10^{2}
3500.003 5	8	$3.500\,003\,5 \times 10^{3}$

Note: normally, numbers between 0.1 and 100 are not written in exponential form but are shown here for clarity.

NOVEL CHALLENGE

The statement '19 is about 20' is reasonable.
Why then can't you say 20 is almost 19? Explain.

Multiplying and dividing

A problem arises when performing calculations using significant figures. Imagine you had to calculate the surface area of a road going through a sensitive koala habitat. The traffic engineers said the road easement would be 95.5 m wide and 26 km long. When multiplying $95.5 \times 26\,000$, the answer of 2 483 000 must show the correct number of significant figures. The rule is: when multiplying or dividing, the answer should contain only as many significant figures as that number involved in the operation that has the least number of significant figures. In this case, 95.5 m has three significant figures and 26 000 m has two. The answer should only have two, so it should be written as 2 500 000 m^2 or 2.5×10^6 m^2. That's a lot of bush.

Other examples are:

- $45.71 \times 34.1 = 1558.711$. This is rounded to 1560 or 1.56×10^3, which has three significant figures (3 sf).
- $365 \div 2.4 = 152.083\,333\,3$. This is rounded to 150 or 1.5×10^2 (2 sf).

Rounding-off If you need to round-off you can use this rule: numerals lower than 5 round-off to zero; numbers larger than 5 round-off to 10; when the number to be rounded off is 5 take it up to 10 if the number preceding is even, otherwise take it down to zero. For example: when 16.586 is rounded off to four significant figures it becomes 16.59. When 24.65 is rounded to three significant figures it becomes 24.7 as the 6 is even and hence the 5 is rounded up to 10.

NOVEL CHALLENGE

The rate at which hydrogen is consumed on the Sun is proportional to the temperature (in kelvins) raised to the power of 20 (rate $\propto T^{20}$). How much faster is the rate at 6000 K than it would be at 5000 K?

Addition and subtraction

If a 1575 g target is struck with a 2.55 g bullet, which becomes embedded in it, the mass of the target is now 1575 g + 2.55 g = 1577.55 g. Or is it? The final mass has more significant figures than either the target's mass or the bullet's mass. Intuitively this should sound wrong. The final mass should be written as 1578 g. The rule is: calculations are rounded off to the least significant decimal place value in the data.

NOVEL CHALLENGE

On the side of 'Tesco' lemonade cans it reads: 8 kJ per 100 mL, 17 kJ per 200 mL.
How can this be?

Examples

(a) $264.68 - 2.4711 = 262.2089 = 262.21$.
(b) $2.345 + 3.56 = 5.905 = 5.90$.

Questions

9 State the number of significant figures in each of the following: **(a)** 83.83; **(b)** 20.0; **(c)** 5; **(d)** 22 050; **(e)** 100; **(f)** 100.010; **(g)** 1999; **(h)** 2.222 2; **(i)** 40 000; **(j)** 0.050 70; **(k)** 0.000 000 200.

10 For the numbers in Question 9 above, write them out in scientific notation and use the correct number of significant figures.

11 How many significant figures are there in the following: **(a)** 4.6×10^3; **(b)** 1.00×10^5; **(c)** 6.07×10^{-6}; **(d)** 3.300×10^{-10}?

12 Calculate the following and express in scientific notation to the correct number of significant figures: **(a)** 12.3 m × 34.14 m; **(b)** 3.5×10^2 m × 2.18×10^4 m; **(c)** 180 cm ÷ 2.5 s; **(d)** 1.18 cm × 3.1416 cm; **(e)** 2.0×10^{-3} m × 2.0×10^{-4} m.

13 Work out the following: **(a)** 5.2 m + 16.013 m + 24.37 m; **(b)** 2.125 m + 11.473 2 m + 9.0124 m; **(c)** 3.0×10^3 m + 3.0×10^4 m; **(d)** 4.0×10^{-3} cm + 5.0×10^{-2} cm; **(e)** 1.118×10^4 m + 2.34×10^6 m; **(f)** 8.7×10^{-5} m + 3.5×10^{-2} m.

14 Calculate (2.34 kg + 1.118 kg) ÷ (1.05 cm × 22.2 cm × 0.9 cm).

15 A sheet of copper was measured as part of a density experiment. The dimensions were: length 55.5 cm, breadth 2.0 cm, thickness 0.02 cm. Calculate **(a)** the area of the largest surface; **(b)** the volume; **(c)** the perimeter of the largest face.

1.5 ORDER OF MAGNITUDE

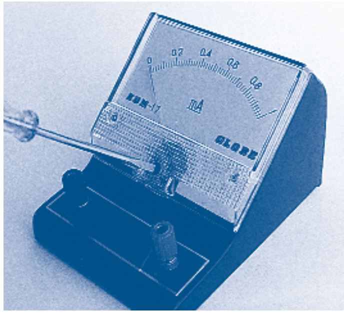
When dealing with very large or very small numbers we are often only interested in an approximate figure. For example, the remotest object known is the quasar RDJ030117 located at a distance of 2.8×10^{22} km from Earth. It is just as meaningful to say it is 10^{22} km away. This is said to be its **order of magnitude** (OM). Similarly, the mass of a hydrogen atom is 1.67×10^{-27} kg, so its order of magnitude is 10^{-27}. The order of magnitude is the power of 10 closest to the number. However, when converting a number to its nearest 10, the rule is: numerals greater than 3.16 become 10 and those below 3.16 become zero. The reason for this is that $10^{0.5} = 3.16$.

Table 1.9 ORDER OF MAGNITUDE

MEASUREMENT	DIMENSION	ORDER OF MAGNITUDE
Distance to Andromeda galaxy	1.9×10^{22} m	10^{22} m
Distance to nearest star	4.0×10^{16} m	10^{17} m
Diameter of Earth	1.3×10^7 m	10^7 m
Thickness of a credit card	5.0×10^{-4} m	10^{-3} m
Thickness of a hair	2.8×10^{-5} m	10^{-5} m

Calculations When estimating the order of magnitude of a mathematical calculation, it is convenient to convert each number to its order of magnitude first.

Photo 1.2
Correcting zero error on an ammeter.



Example

Determine the order of magnitude of this calculation: $(3.0 \times 10^{10}) \times (8.4 \times 10^6)$.

Solution

- 3.0×10^{10} has an OM of 10^{10}; 8.4×10^6 has an OM of 10^7.
- $10^{10} \times 10^7$ equals 10^{17}.

Note: the full answer is 2.52×10^{17}, which does have an OM of 10^{17}.

Ch01-Walding 4th 25/8/04 9:19 AM Page 15

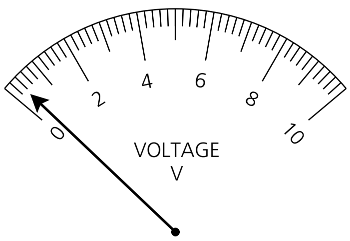
Questions

16 What is the order of magnitude of each of the following: **(a)** 1.8×10^{22}; **(b)** 3.9×10^{12}; **(c)** 2.6×10^{-10}; **(d)** 5.8×10^{-15}; **(e)** 175 000; **(f)** 66 000; **(g)** 0.000 002; **(h)** 0.000 65?

17 Estimate the order of magnitude of the answer for each of the following calculations: **(a)** $(6.2 \times 10^{20}) \times (3.8 \times 10^{-18})$; **(b)** $(600) \times (10 \times 10^{8})$; **(c)** $5.4 \times 10^{-12} \div 3.1 \times 10^{-15}$.

MAKING AND RECORDING MEASUREMENTS 1.6

Figure 1.5
Zero error. This voltmeter has a zero error of 0.4 volt. It can be zeroed by adjusting it with a screwdriver (see Photo 1.2).



If you had to count the number of desks in your classroom you would get an exact figure but if you had to measure the width of a desk with a metre ruler your measurement would be an approximation, probably to the nearest millimetre. Measurements, unlike numbers, can never be exact because they all have some amount of error or uncertainty.

You can end up with errors in a measurement because of the limitations of the measuring instrument or the conditions under which it was made. Such errors are not mistakes because they are not someone's fault. Some examples of errors include:

- zero error, for example the pointer or the end of a ruler not on the zero mark to start with (See Figure 1.5.)
- calibration error, for example a stopwatch that runs fast or slow, a thermometer badly graduated, or a metal ruler that has expanded in the heat
- parallax error, for example reading a clock at an angle so that the hand appears to be over another number, reading a thermometer at an angle
- reaction time, for example the delay in starting a stopwatch.

Figure 1.6
(a) A parallax error will occur because there is a gap between the scale and the object being measured. **(b)** There is no parallax error as the scale and object are touching.

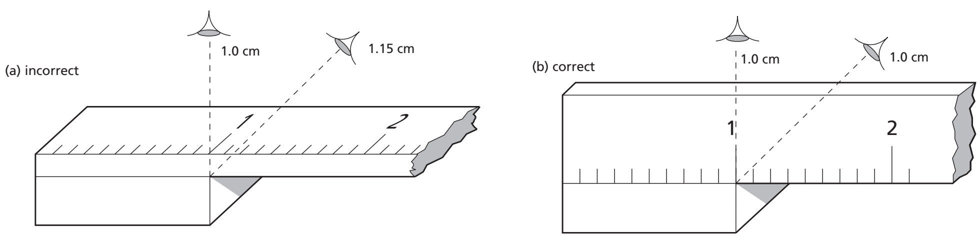
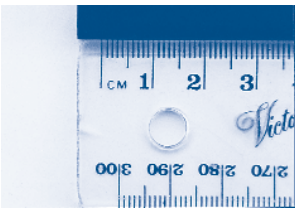


Photo 1.4
If you did this you would have a zero error of 4 mm.



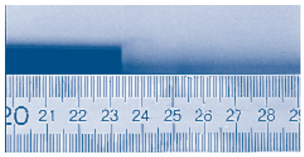


Photo 1.3a
From directly above the reading is 23.35 mm.

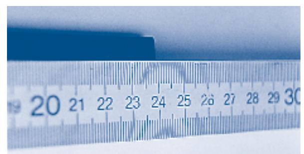
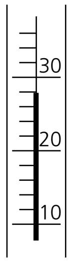


Photo 1.3b
From an angle there is parallax error and the reading is now 23.85 mm.

Figure 1.7
Scale division error on a thermometer. The reading on this thermometer is 28°C not 24°C. Each scale division is 2°C not 1°C as may be thought.



These errors can be classified into two main types:

- **systematic errors** in which all of the readings are faulty in one direction and can be usually corrected for by a simple calculation or improved experimental technique (Zero errors and calibration errors are of this type.)
- **random errors**, which are irregular errors of observation. Parallax error is an example.

Mistakes are not errors in this context. If you misread a scale (Figure 1.7) by miscalculating the value of each division, this is sometimes called a 'scale reading error' but is really just a mistake.

SR Activity 1.4 PARALLAX ERROR

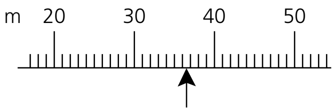
Hold your arm outstretched in front of you with your thumb pointing up. With one eye closed, line your thumb up with some mark on a wall in front of you. Close that eye and open the other and note how many centimetres your thumb has shifted to the side of the mark. Which eye was the more dominant? What are some ways of controlling parallax error?

Scale reading limitations

Students generally read scales to the nearest mark or division. For example, the reading on the ruler shown in Figure 1.8 would generally be stated as 36 mm but it really looks closer to 36.5 mm than to 36.0 mm. A better reading would be 36.5 mm.

Some people would claim to be able to read to the nearest 0.1 mm but this seems overly accurate for the type of scale used. A good rule is that scales should be read to the nearest half of a scale division. Rulers can be read to the nearest half-millimetre and laboratory thermometers to the nearest 0.5°C. An ammeter like the one shown in Figure 1.9 is best read to the nearest 0.05 A.

Figure 1.8
This ruler can be read to the nearest 0.5 mm.



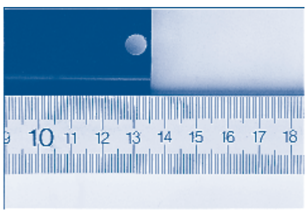


Photo 1.5a
A ruler calibrated in 1 mm divisions can be read to the nearest 0.5 mm. In this case the reading is 135.5 mm.

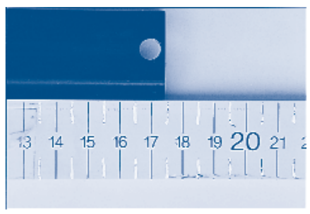


Photo 1.5b
If the ruler was calibrated in 1 cm divisions, then you could read to the nearest 0.5 cm —in this case 17.5 (175 mm).

Uncertainty

You can't measure a physical quantity exactly because all instruments have limitations. These limitations make any reading uncertain. However, some digital instruments appear to give more exact measurements than the manufacturers ever intended. For example, an ammeter with a display of 258 mA seems to be indicating that the current is exactly 258 mA, whereas it may really mean 258 ± 1 mA.

A general rule-of-thumb is that the **uncertainty** in a reading is said to be equal to a half scale division on the instrument. For a ruler marked in millimetres, the absolute uncertainty is ± 0.5 mm so the reading above could have been stated as 36.5 ± 0.5 mm. This absolute uncertainty could be also expressed as a percentage uncertainty:

$$\text{Percentage uncertainty} = \frac{\text{absolute uncertainty}}{\text{observed measurement}} \times 100\%$$

$$= \frac{0.5}{36.5} \times 100\% = 1\%.$$

The uncertainty is a way of expressing how confident you are about the readings provided by the instrument. It is a measure of the limitations of the instrument.

Figure 1.9
The ammeter scale has 0.1 A divisions, so it can be read to the nearest 0.05 A (half of 0.1).

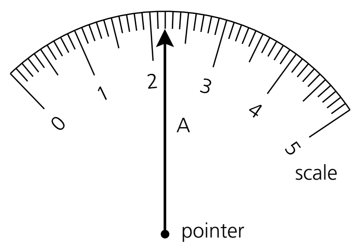
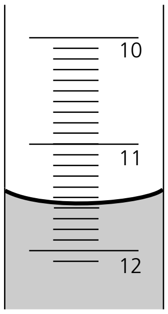


Figure 1.10
This burette shows a reading of 11.55 mL. One half-scale division equals 0.05 mL.



Uncertainty calculations

To add, subtract, multiply or divide numbers, the absolute and relative uncertainties may be required.

- For addition and subtraction, *add absolute uncertainties*.
- For multiplication and division, *add percentage uncertainties*.

Example 1

A container of water rises in temperature from 25.5 ± 0.5°C to 36.0 ± 0.5°C. Calculate the rise in temperature and its percentage uncertainty.

Solution

$$36.0 \pm 0.5°\text{C} - 25.5 \pm 0.5°\text{C} = 10.5 \pm 1.0°\text{C}$$
$$= \frac{1.0}{10.5} \times 100\% = 9.5\%.$$

Example 2

A piece of paper is measured and found to be 5.63 ± 0.05 mm wide and 64.2 ± 0.5 mm long. What is the area of the piece of paper?

Solution

Area = length × width
= (5.63 ± 0.05 mm) × (64.2 ± 0.5 mm)
= 5.63 ± 0.89% × (64.2 ± 0.78%) (convert to percentage uncertainty)
= 361.446 ± 1.67% (add percentage uncertainties)
= 361.446 ± 6.025 mm^2 (convert percentage uncertainty to absolute uncertainty)
= 361 ± 6 mm^2
(Round answer to three significant figures and round the uncertainty to one significant figure as given in the original data.)

Questions

18 A cube of brass was measured and found to have a side of length 13.0 ± 0.5 mm. Determine the volume of the cube.

19 A student made two measurements using a metre ruler calibrated in millimetres.
First measurement = 25.5 mm.
Second measurement = 174.5 mm.
(a) What are the absolute uncertainties for these measurements?
(b) Convert these absolute uncertainties to relative uncertainties.
(c) Add the two measurements and show the absolute uncertainty of the result.
(d) Multiply the two measurements and show the absolute uncertainty of the result.

20 Determine the correct value for the area of a horse paddock 645 ± 5 m long and 345 ± 5 m wide. What is the total length of fencing needed to fence this paddock?

Accuracy and precision

Students often find that despite performing an experiment as accurately as possible and reading the instruments as best as they are able, their results are different from the accepted or textbook result. This difference is called the **error**. The error is a measure of the **accuracy** of a result. Accuracy refers to the closeness of a measurement to the accepted value.

Imagine your group measured the density of water to be 1.02 g/mL when the accepted value was given as 1.00 g/mL at that temperature. Your (absolute) error would be 0.02 g/mL.

Hence:

- **The absolute error** (E_a) = | observed value – accepted value | = |O – A|.

Note: the straight lines (|) in the above equation mean the 'absolute value', that is, the sign (+/–) of the answer is ignored.

- The **relative error** (E_r) is the absolute error expressed as a percentage of the accepted value (A):

$$\text{Relative error } (E_r) = \frac{E_a}{A} \times 100\%$$

In the above example E_r, would equal $\frac{0.02}{1.00} \times 100\% = 2\%$.

This is necessary so that accuracy between different experiments can be compared. Imagine that a student measured the density of lead as 11.29 g/cm^3, while the accepted value was 11.41 g/cm^3. Which result is the more accurate — the density of water or the density of lead? In this case you need to compare relative errors: the error for water was 2% whereas that for lead was 1% and hence was more accurately measured.

Example 1

Calculate **(a)** the absolute error and **(b)** the relative error in a student's measurement of the acceleration due to gravity. They obtained 9.73 m/s^2 whereas the accepted value at their location was 9.813 m/s^2.

Solution

(a) E_a = |O – A|
= 9.73 – 9.813
= 0.08 m/s^2 (to the correct number of significant figures).

(b) $E_r = \frac{E_a}{A} \times 100\%$
$= \frac{0.08}{9.813} \times 100\%$
= 0.8%.

Example 2

When lower profile tyres are fitted to a car in place of the factory fitted ones, a speedometer reading error can occur as the new tyres have a smaller diameter. A table was compiled by a motor magazine during a road test (Table 1.10).

Table 1.10

	SPEED (km/h)			
Observed	60	80	100	110
Actual	57.0	76.0	95.0	104.5

Calculate the relative error at 80 km/h.

Solution

$$\text{Relative error} = \frac{E_a}{A} \times 100\% = \frac{4}{76} \times 100 = 5.2\%$$

Activity 1.5 ERROR IN ESTIMATING PI (π)

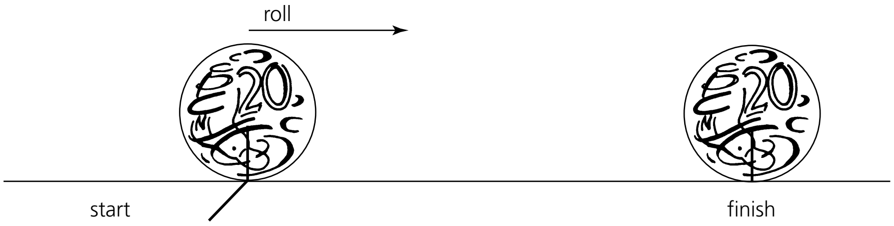


Figure 1.11
Experiment to calculate errors in estimating pi.

PHYSICS FACT

In 1783 William Shanks reported a value of pi to 707 places, beating the previous value by 200 places. In 1949 a computer was used for the first time to calculate pi mechanically and they found that Shanks made a mistake at the 528th digit and was wrong from then on. Shanks took 15 years to make his calculation—and he was wrong. What a waste!

1. Draw a line on a piece of paper and place a starting mark at one end (Figure 1.11).
2. Make a mark on the side of a 20 cent coin at the edge.
3. Line up the two marks and roll the 20 cent coin along the line until the mark on the coin touches the paper again, and then put a finish mark.
4. Measure the diameter, *d*, of the coin with whatever instrument you choose. Measure the length of the line between the start and finish marks. This is the circumference, *c*, of the coin.
5. Calculate π by using the formula $c = \pi d$ (i.e. $2\pi r$).
6. Knowing that $\pi = 3.141\ 59$, calculate the absolute and relative errors in your estimate of pi.

Summing up:

- *Uncertainty* is a measure of how confidently you can state a measurement or result and is a direct result of the limitations of an instrument. The terms absolute uncertainty and relative uncertainty are used.
- *Accuracy* is a measure of how close a measurement is to an accepted value. The terms absolute error and relative error are used.

Questions

21 Convert the following percentage errors back to absolute errors: **(a)** 27.6 ± 1.5%; **(b)** 10.35 ± 0.6%.

22 Calculate the relative error for the following speeds (as shown in Table 1.10): **(a)** 60 km/h; **(b)** 100 km/h; **(c)** 110 km/h. Does the speedo become more inaccurate at higher speeds?

23 A carbon resistor of nominal resistance 330 ohms is manufactured to a tolerance of 5%. This is, in effect, the maximum relative error. Calculate the range of resistance that this resistor could be.

MEASURING INSTRUMENTS 1.7

Just as units of measurement changed as people's needs changed, so too did the instruments they used for measuring things. Ancient societies achieved incredible accuracy with their primitive devices — rods, string and even line-of-sight. But as precision engineering became vital to industrial society, instruments were developed to achieve such precision.

In this section we will look at the:

- micrometer screw gauge
- vernier calliper
- stroboscope
- digital counter.

Ch01-Walding 4th 25/8/04 9:19 AM Page 20

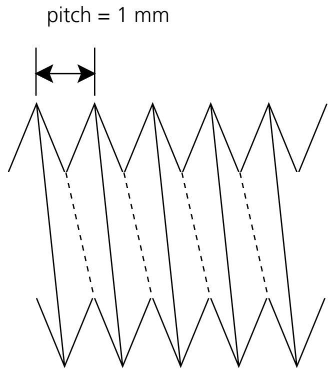
The micrometer screw gauge

To measure really tiny things a **micrometer** can be used. It can measure down to about one-hundredth of a millimetre. The principle behind the micrometer is the screw — one rotation of the screw moves it through a distance equal to the pitch (the distance from one thread to the next) as shown in Figure 1.12. If the screw is rotated only a fraction of a turn, then the screw advances that fraction of the pitch.

A common type of laboratory micrometer has a main scale marked off in half-millimetre divisions. One revolution of the thimble moves the main shaft 0.5 mm. The thimble itself is divided into 50 divisions so that 1 mm equals 100 thimble scale divisions. Hence 1 thimble scale division = 1/100 mm or 0.01 mm. The micrometer in Figure 1.13 shows a reading of 6.5 mm on the main scale and 27×0.01 (= 0.27 mm) on the thimble scale. The final reading is thus 6.77 mm.

There are many types of micrometers available. Your school's could be quite different from the one described here.

Figure 1.12
The pitch of this micrometer is 1 mm.



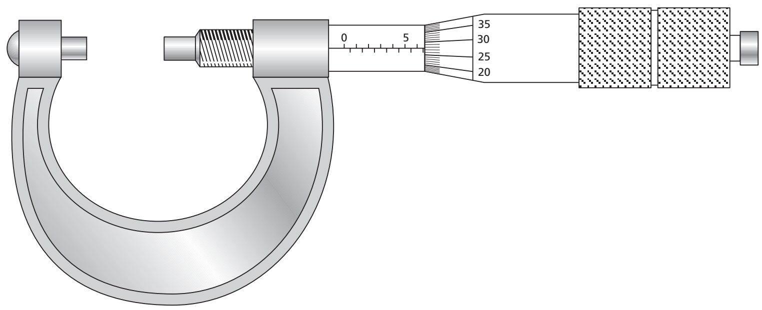


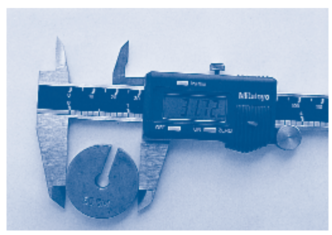
Figure 1.13
This micrometer reads 6.5 mm on the main scale and 0.27 mm on the barrel, making a total of 6.77 mm.

SR Activity 1.6 THE VERNIER CALIPER

Try the following as a good stimulus response task.

The vernier caliper has two jaws that slide together over the object being measured. The caliper was named after the French mathematician Pierre Vernier, who devised the scale. It uses an auxiliary scale (the vernier scale) in conjunction with a main scale to assist in estimating fractions of a main scale division. The main scale is graduated in millimetres (called main scale divisions or MSD) and each centimetre is numbered. The vernier scale is 9 mm long and yet is divided into ten equal divisions (called vernier scale divisions or VSD). It can be shown that the smallest possible division on the vernier scale is one-tenth of 1 mm = 0.1 mm. The procedure is: count the number of complete main scale divisions (MSD) up to the zero line on the vernier scale. Count the number of vernier scale divisions (VSD) to the point where a vernier scale mark and a main scale mark coincide. This will be in 0.1 mm units. For example, in Figure 1.14, the object is 11 mm long plus 5×0.1 mm, which equals 11.5 mm or 1.15 cm.

Photo 1.6
Digital vernier calipers are now becoming more commonplace especially as their price has come down to about $100.



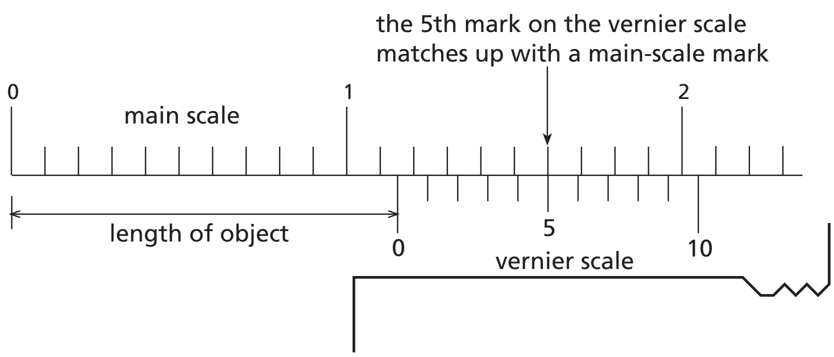


Figure 1.14
This vernier calliper reads $(1.1 + 5 \times 0.01)$ cm, i.e. 1.15 cm.

Question

24 What is the reading on the vernier calipers shown in Figure 1.15?

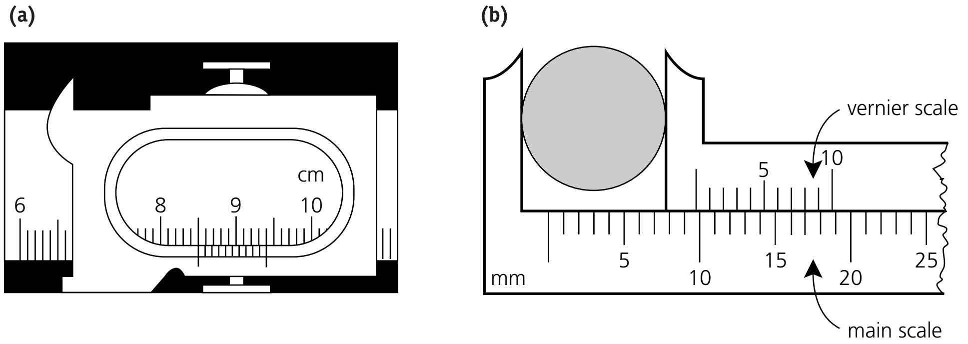


Figure 1.15
For question 24.

The stroboscope

The **stroboscope** owes its name to the Greek *strobos* meaning 'to whirl around' and *skopion* meaning 'see'. The most common form consists of a xenon flash tube similar to that found in a camera flash (Photo 1.8). It can be made to flash at a variable rate from about 1 per second to tens of thousands per second. If a rotating object is in fairly dim conditions and the light flashes when the object is in the same position every time then the object will appear stationary. However, you wouldn't know if the object rotated two or three or a hundred times between flashes so you have to make sure by starting at the lowest strobe frequency and gradually increasing it until motion 'freezes'.

One problem with strobe illumination is that by freezing a rotating object (e.g. a fan blade or a part of a lathe) onlookers may be confused into thinking it is stationary and this would of course be very dangerous. In factories, special precautions are taken with machinery that is illuminated by fluorescent lights. Fluorescent lights flicker at 100 times per second or 100 hertz — once for each crest and trough of an alternating current. Machinery operating at multiple frequencies of this could appear stationary. Such lights have different capacitors added to make them flicker out of synchronisation, which breaks up the strobe effect.

The digital counter

The term '**digital**' conjures up images of modern high-technology but in reality it just means counting in units. This could be like counting 'yes' and 'no' votes in an election; like 'present' and 'not present' when marking a class roll; 'off' and 'on' for an electrical switch or 'light' and 'dark' as cans of soft drink pass a light sensor on a packaging line.

Photo 1.7
A mechanic using a timing light.
(J.A.T. Mechanical, Brisbane)

Photo 1.8
A xenon stroboscope.

1.8 A FINAL NOTE

There is one final caution about measurement and measuring instruments that applies to all devices mentioned throughout this book. An *ideal* measuring device will have no effect on the measurement itself. For instance, when you measure the width of your desk, the desk is unaffected by the measurement. But this is not the case for all measuring devices. When you measure the pressure of a car's tyres some gas is sampled and the tyre has less gas than before you started. The loss is insignificant, however. A voltmeter or ammeter samples electrons from an electrical circuit and will affect the voltages and currents being measured. But again, if the meters are used properly the effect will be minimal. Can you think of other instruments that affect the phenomena being measured and how the effect is minimised?

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***25** The space shuttle orbits the Earth at an altitude of 300 km. How many millimetres is this?

***26** The Earth is approximately a sphere of radius 6.37×10^6 m. **(a)** What is its circumference? **(b)** What is its volume in cubic metres? **(c)** What is its volume in cubic kilometres?

***27** Submarines typically dive at a rate of 36 fathoms per second. If a fathom is 6 feet and 1 foot is 0.305 m, convert this diving speed to metres per second.

***28** Write the following in scientific notation: **(a)** 3 558.76; **(b)** 40.00; **(c)** 79 000; **(d)** 200 326; **(e)** 1994; **(f)** 20.009; **(g)** 0.050 0; **(h)** 2 500 000; **(i)** 0.000 000 8; **(j)** 5 million.

***29** Do the following calculations on your calculator, using the correct number of significant figures: **(a)** $4.2 \times 10^3 \times 8.1 \times 10^4$; **(b)** $3.7 \times 10^7 \times 4.1 \times 10^{-4}$; **(c)** $7.2 \times 10^4 \div 1.8 \times 10^6$; **(d)** $4.8 \times 10^6 \div 1.6 \times 10^{-3}$; **(e)** $\pi\,(4.1 \times 10^{-6})^2$; **(f)** $2.8 \times 10^3 \div (\frac{4}{3}\pi\,(4.7 \times 10^{-5})^3)$.

***30** Express each of the following as an order of magnitude: **(a)** 4.28×10^7; **(b)** 32 000 000; **(c)** 1.2×10^5; **(d)** 1.13×10^{-4}; **(e)** 4.5×10^{-8}; **(f)** 9 192 000; **(g)** 0.000 000 38.

***31** How many significant figures are there in each of the following: **(a)** 95.2 km; **(b)** 3.080×10^5 g; **(c)** 0.0067 L; **(d)** 0.000 670 L?

***32** Convert the following to relative errors: **(a)** 2.40 ± 0.02 V; **(b)** 3.25 ± 0.05 A; **(c)** 25.4 ± 0.4 mm; **(d)** 0.0035 ± 0.0001 T; **(e)** 325 ± 10 cm.

****33** A student is required to determine the density of a particular metal. The object is in the shape of a cylinder. She uses a micrometer calibrated in 0.01 mm (i.e. a limit of reading of 0.01 mm) to measure the diameter of the cylinder and uses a vernier calliper with a limit of reading of 0.1 mm to measure the length. Recall that the error associated with a reading is half the limit of reading. The results are shown in Table 1.11.

Table 1.11

READING	DIAMETER (mm)	LENGTH (mm)
1	16.446	28.4
2	16.444	28.3
3	16.442	28.5

(a) What absolute error is associated with each reading?
(b) Determine the average values for length and diameter.
(c) Determine the value for radius and length. Include the correct error.
(d) Determine the volume of the cylinder, including its error.
(e) If the cylinder has a mass of 56.4 ± 0.2 g, determine the density of the cylinder in g mm^{-3}.

***34** Use your ruler and calculate **(a)** the surface area of the front cover of this book; **(b)** the total external surface area; **(c)** the volume of this book; **(d)** the thickness of one page.

Extension — Complex, challenging and novel

****35** What does the prefix 'micro' signify in the words 'microwave oven'? Does it mean it is a small oven? It has been proposed that food that has been irradiated by gamma rays to lengthen its shelf life should be called 'picowaved'. What do you suppose that means?

*****36** Convert the speed of light (3.0×10^8 m s^{-1}) to furlongs per fortnight. A furlong is one-eighth of a mile; there are 5280 feet in a mile and one foot is 0.305 m.

*****37** A wire of length 756.5 ± 0.5 mm has a mass of 8.5 ± 0.5 g. Calculate the mass per millimetre.

*****38** Isaac Asimov proposed a unit of time based on the highest known speed of light and the smallest measurable distance. It is the light-fermi, the time taken by light to travel a distance of 1 fermi (= 1 femtometre = 1 fm = 10^{-15} m). How many light-fermis are there in 1 second? Recall that light travels at 3×10^8 m s^{-1}.

****39** Some of the prefixes of the SI units have crept into everyday language. What is the weekly equivalent of an annual salary of \$36 K (= 36 kilodollars)?

*****40** The hard disk of a particular computer was stated as 200 MB (= 200 megabytes). At 8 bytes per word, how many words can it store? Note that in computerese, kilo means 1024 (= 2^{10}) not 1000 and mega means 2^{20} not 1 million.

****41** When the length of a metre was defined in 1983, the speed of light was accepted as 2 997 924 58 ms^{-1}. Why was it not defined as exactly 3.000×10^8 m s^{-1} to make it simpler?

*****42** The following is an extract from *The Times* newspaper, London. Read it and answer the question below.

Time, gentlemen

The Gregorian calendar, which celebrated its quatercentenary in October 1982, is still working well. So well, in fact, that it will be some time in AD 4316 before it is a complete day out.

The trouble is that when God created, he did so without benefit of digital timekeeping, and a year is currently 365.2422 days long. This leaves a rather useless plane-shaving of time at the end of each year. Julius Caesar was without digitals, too, but his astronomer Sosigenes did a remarkably fair job in 46 BC to produce a year of 365 days and six hours — only a week out every 1000 years. This was perfectly adequate for the ancients, who rose and retired by the sun, but not for those pernickety Christians, who became deeply concerned about Easter being on the correct day.

By the time Pope Gregory XIII wrestled with the problem, the Julian calendar was 10 days adrift. So at midnight on October 5, 1582, he declared the next day to be October 15. It brought the vernal equinox in the northern hemisphere back to March 21 and the peasants never felt a thing.

Britain, having long grown wary of such popish tricks, did not deign to accept the Gregorian calendar until 1752. But it was by no means the last country to abandon old Caesar's almanac. Russia did not go Gregorian until 1918, after the Revolution, and the last country of all to abandon it seems to have been Greece, in 1923.

(a) Why did the Christians need a more accurate calendar?

(b) Russian chemist Mendeléev devised the Periodic Table of the Elements on 1 March 1901 in Moscow. What date would this have been in London?

(c) By 1989 the calendar was only out by 2 hours 49 minutes since 1582. How far out will it be in 2005?

*****43** The size of a molecule can be determined by placing a drop of oil on the surface of water and noting the maximum area of the oil slick which is assumed to be one molecule thick. We tried this and found that one drop spread to a circle with a diameter of 14 cm. We also found that there were 20 drops of oil to the milliliter. Calculate the thickness of the slick.

*****44** There are 6×10^{23} molecules of water in 18 mL of water. If the ocean has a volume of 1.3×10^{18} m^3, how many glasses of water (at 250 mL each) are there in the ocean? Comment on the assertion that 'there are more molecules of water in a glass of water than there are glasses of water in the ocean'.

*****45** Neutron stars have a radius of 20 km and a mass equal to our Sun (2×10^{30} kg). What is the mass of a cubic centimeter of neutron star?

UNIT 02

Forces & Motion

CHAPTER 02

Motion in a Straight Line

2.1 OBSERVING MOTION

People have been watching and recording things move for thousands of years. The motions of the heavens are some of the oldest recorded observations we have. Later, a need to measure the speed of advancing armies or athletes or ships required better ways of measuring distance and time. Over the centuries measurements became more accurate and now form the basis of modern physics. We can now measure distances and times to incredible accuracy.

Many types of motion are occurring around us all the time. Blood flow, moving bullets, cricket balls, athletics, cars, stars, planets, neutrinos and weaving looms are some of the areas where motion is measured. Some need to be measured carefully, others not. A car speedometer that is a few kilometres per hour over or under makes little difference but better accuracy is needed when timing a 100 metre sprint or controlling the speed of videotape through the heads of a VCR.

Sometimes the motion of objects doesn't make sense. Can you make sense of these questions?

- We live on a world that is round, yet we do not fall off. Many people used to believe the world was flat. Some still do. What evidence is there that it is round?
- Before Copernicus, most people believed that the Earth was stationary and the Sun moved around it. We now believe that the Earth is moving around the Sun but how do we know this?
- The Earth moves in a circular orbit and never slows down. Most objects in the world seem to travel in straight lines and slow down. Why is the Earth different?

The above three questions have several similarities. How many different things do they have in common?

Physics developed over the centuries as people pondered on these questions and came up with all sorts of different explanations. But people also found that knowing about the motion of everyday objects became more and more important.

It helps with your problem solving if you are familiar with some common motions and their measurements.

NEI Activity 2.1 SPEEDOMETER

Have a look at your family car's speedometer.

1. What is the maximum speed that it can record?
2. Do you know what your car's top speed is? If you don't, where would you find out? Assuming that it can't go as fast as the maximum value on the speedo, why do manufacturers use this sort in cars?
3. How many km/h are there per division?
4. The odometer (Greek *hodos* = 'a way') measures the total number of kilometres travelled by the car from when it was new. What is the maximum number of kilometres your car can travel before the odometer returns to all zeros?

Photo 2.1
A car speedometer.

Ch02-Walding 4th 25/8/04 10:14 AM Page 27

5 Does your odometer measure to the nearest kilometre or tenth of a kilometre?
6 What is the maximum distance your 'trip meter' will record?
7 Some unscrupulous people illegally 'wind back' the odometer. What is the purpose of this and how do they do it?
8 Does the odometer go backwards when your car is reversed?
9 Does the speedo of your car go lower than zero when reversed?

NEI Activity 2.2 SEWING MACHINE

Look at a sewing machine. How can you change the speed of a sewing machine motor? Is it variable? Are all electric motors controlled in the same way?

NEI Activity 2.3 VIDEO RECORDER

If you have a VCR and can find the instruction manual, find out the tape speed on standard play. Should everyone in the class get the same result? Is the speed the same in videocameras? Are speed and image quality related?

A knowledge of physics enables us to analyse all types of motion. Without accurate measurement and control, life would be difficult indeed.

DISTANCE AND DISPLACEMENT 2.2

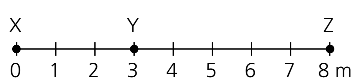
From the earliest times, being able to measure distances, angles and time was important in the daily lives of people. Often it was for religious reasons — worshipping sun gods; other times it was an attempt to plot the motion of the stars — a primitive astronomy. But sometimes it had a more practical purpose. Measuring distance, for instance, was important in the construction of houses, building canals and cultivating fields.

Plato told the story of how Posiedon (421 BC) inherited the island of Atlantis with its irrigated plain of 3000 by 2000 *stades* (a 'stade' is 185 metres, hence the word 'stadium'). Today, of course, we would be more likely to use metres or kilometres.

Whereas *length* is a measure of how long or wide an object is, we use the term *distance* to say how far the object has moved. A person travelling from one city to another may have moved a distance of 1200 km. In physics, we need to be able to measure not only distance but also 'displacement'.

Displacement is the change in position of an object in a given direction. You can think of it as the position measured relative to the origin. It is given the symbol '*s*'.

Figure 2.1

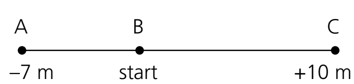


In Figure 2.1, if you started at point X and walked 8 m east to point Z and then turned around and walked 5 m west to point Y, you would have moved a distance of 13 m but would only have a displacement of 3 m east. That is, your position would only have changed by 3 m to the east. In symbols this could be written as s = 3 m E.

Displacement is called a **vector** quantity. That is, it involves both a number and a direction. Other vector quantities are velocity, acceleration and force. Quantities that do not include a direction are called **scalar** quantities. Distance, speed, mass and time are all scalar quantities. In the next chapter, vectors will be discussed in more detail.

When discussing vector quantities like displacement we use the compass points (N, E, W, S) to define directions as we did above, or alternatively, we can use a positive sign for forward motion or motion to the right and a negative sign for backward motion or motion to the left.

Figure 2.2



For example, in Figure 2.2 the displacement of C can be written as s_C = +10 m; and the displacement of A can be written as s_A = –7 m.

Either way, you'll need to be able to use both conventions. It's up to you and it is also up to you to define the positive and negative directions.

Representation of vector quantities

A vector quantity can be represented by a vector. A vector is an arrow. The length of the arrow represents the magnitude of the vector quantity, and the direction of the arrow shows the direction of the vector quantity. For example, the three vectors in Figure 2.3 represent cars travelling at 30 km/h east, 60 km/h west and 10 km/h north respectively.

Figure 2.3

When vectors do not lie along the compass points (N, E, S, W), angles need to be specified. Figure 2.4 shows how the direction is indicated.

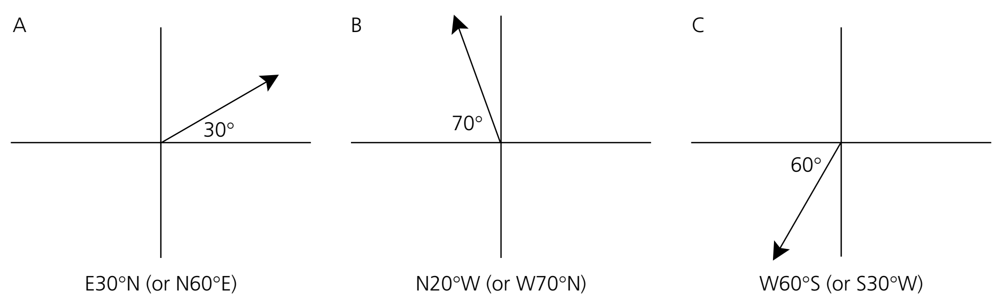


Figure 2.4

Students often find it hard to work out the directions. You can think of diagram A in Figure 2.4 as saying: going **e**ast but rotated **30°** to the **n**orth.

Example

In Figure 2.5, an orienteering competitor starts at point A and goes 2 km N, 4 km E and then 2 km S. What is the final displacement at point D?

Solution

The displacement at D is 4 km east (s_D = 4 km E).

Questions

1 In Figure 2.5:

(a) What is the displacement of the competitor at point B? (s_B = ?)
(b) What is the total distance travelled when at point D?
(c) What is the distance travelled when at point C?
(d) What is the displacement at point C? Remember to include the direction by stating the value of the angle CAD.

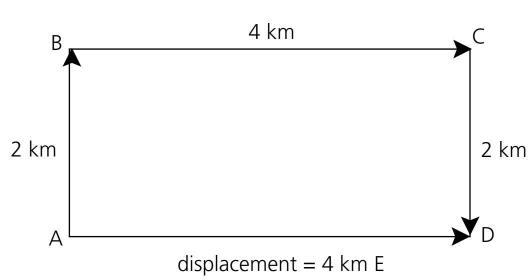


Figure 2.5
For question 1.

2 You watch your dog following a cat's scent trail. He walks 50 m north, turns and walks 60 m east and then walks 50 m south. What is his displacement?

3 A toy train is running around a circular track of diameter 120 cm. What is its distance travelled and its displacement after **(a)** one-half of a lap; **(b)** one full lap; **(c)** two laps; **(d)** one-quarter of a lap?

SPEED AND VELOCITY 2.3

Newspapers and magazines use the terms 'speed' and 'velocity' as if they mean the same thing. They do — almost. When a newspaper report mentions a high-speed car chase we know what is meant. But why do they also talk about hunting rifles being high-velocity? Newspapers say high-velocity atomic particles but they also talk of a cyclone's wind speed. Newspapers mean the same thing by speed and velocity. Why do you think they refer to some motions as speed and others as velocity?

In physics, speed and velocity are slightly different terms. Speed is a scalar quantity whereas velocity is a vector quantity. If it takes 2 hours to travel the 120 kilometres from Brisbane to Noosa then the average speed is 60 kilometres per hour. Speed is the rate at which distance is covered. Remember, the word 'rate' is a clue that something is being divided by time. Speed is always measured in terms of a unit of distance divided by a unit of time, such as metres per second.

$$\text{Average speed} = \frac{\text{total distance travelled}}{\text{time taken}}$$

This of course doesn't mean the driver sat on 60 km h^{-1} all the way. Sometimes the car would have gone at 100 km h^{-1} and at other times it would have been stationary. While the car's speedometer was reading 60 km h^{-1} then the car was actually travelling at that speed for that moment. This is called its instantaneous speed.

When we talk of a car's speed as being 60 km h^{-1} we have no idea about the direction it is travelling. Speed is a *scalar* quantity.

Velocity is defined as speed in a particular direction, for example 60 km h^{-1} north. Velocity is a *vector* quantity and the direction must be stated. In this book we represent a vector by printing its symbol in ***bold italics***.

Imagine a person running to catch a bus. Figure 2.6 shows him running north up Main Street at 5 m s^{-1}, turning east into Mary Street and continuing to run at 5 m s^{-1}. Although he was running at constant speed, his velocity changed because his direction changed.

Instantaneous velocity is similar to instantaneous speed except that a particular direction must be stated.

$$\text{Instantaneous velocity} = \frac{\text{small distance travelled in a stated direction}}{\text{time taken for this small distance}}$$

As distance moved in a stated direction is called 'displacement',

$$\text{instantaneous velocity} = \frac{\text{displacement}}{\text{time taken}}$$

$$\boldsymbol{v} = \frac{\boldsymbol{s}}{t} \text{ metres per second} = \frac{\text{metres}}{\text{seconds}}$$

NOVEL CHALLENGE

Try out these 'Fermi questions':

A How many golf balls will fit in a suitcase?

B How many hairs are there on a human head?

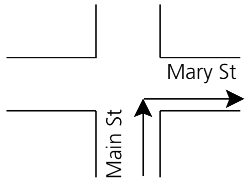
C How quickly does human hair grow (in kilometres per hour)?

D If all the people of the world were crowded together, how much area would we cover?

E What is the relative cost of fuel (per kilometre) of rickshaws and cars?

F How far does a car travel before a one-molecule layer of rubber is worn off the tyres?

Figure 2.6
Turning into Mary Street. A change of direction means a change in velocity.



As with speed, we can use the term 'average velocity' to describe the motion of an object such as a car.

$$\text{Average velocity} = \frac{\text{displacement}}{\text{time taken}} \text{ or } \boldsymbol{v}_{av} = \frac{\Delta s}{\Delta t}$$

where Δ (delta, the Greek 'D') means 'change in', that is, Δt means change in time, but usually the deltas are omitted. The formula can be rearranged like this:

$$\boldsymbol{v}_{av} = \frac{s}{t} \qquad s = \boldsymbol{v}_{av}\, t \qquad t = \frac{s}{\boldsymbol{v}_{av}}$$

Table 2.1 COMPARISON OF SOME COMMON SPEEDS

MOVEMENT	m/s	km/h
Worm	0.005	0.01
Walking	1.4	5
Jogger	2.8	10
Cheetah	28	100
Sound in air	330	1200
Light	3×10^8	1 billion

Example 1
The trip meter on a car's speedo was set at zero and after a journey lasting half an hour the reading was 35 km. What was the average speed?

Solution

$$\text{Average speed} = \frac{\text{distance}}{\text{time}} = \frac{35\text{ km}}{0.5\text{ h}} = 70\text{ km/h or } 70\text{ km h}^{-1}$$

Example 2
A person rides a bicycle 5 km east and then 5 km north (Figure 2.7). The trip takes 1.5 hours. Find **(a)** the total distance travelled; **(b)** the average speed; **(c)** the displacement; **(d)** the average velocity.

Solution

(a) Total distance = 5 km + 5 km = 10 km.

(b) $\text{Average speed} = \frac{\text{distance}}{\text{time}} = \frac{10\text{ km}}{1.5\text{ h}} = 6.7\text{ km h}^{-1}$.

(c) Displacement = $\sqrt{5^2 + 5^2}$ = 7 km in a NE direction (s = 7 km NE or 7 km N45°E).

(d) $\text{Average velocity} = \frac{\text{displacement}}{\text{time}} = \frac{7\text{ km}}{1.5\text{ h}} = 4.7\text{ km h}^{-1}$ NE (N45°E).

NOVEL CHALLENGE

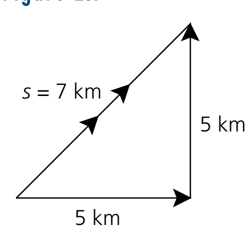
A lizard runs 30 m west, rests and heads 40 m north where it meets the base of a tree. It scampers 5 m straight up the tree. What is the magnitude of its displacement? How are you going to indicate the angle?

NOVEL CHALLENGE

The Greek symbol for 'D' is delta, Δ. In science, we use Δ to represent 'difference' because this also starts with 'D'. A delta is a triangular piece of flood plain where a river meets the sea, as in the Nile delta.

Was the symbol Δ called delta because it looked like the delta of a river, or was the flood plain called a delta because it looked like the Greek symbol Δ? The Greeks got the word delta from the inventors of the alphabet — the Phoenicians — who used it to mean 'door'.

Figure 2.7



Questions

4 To help you rearrange equations and substitute numbers, do the simple calculations shown in Table 2.2. Do not write in this book.

NOVEL CHALLENGE

Confirm or refute the following statement: 'When an object is moved, its displacement can be smaller than the distance travelled, but the distance travelled can never be smaller than the displacement.'

Table 2.2

DISPLACEMENT	TIME	VELOCITY
(a) 200 m	10 s	
(b) 50 km	1.5 h	
(c)	30 s	140 m/s
(d)	3 h	220 km/h
(e) 300 m		15 m/s
(f) 130 km		65 km/h

5 An archer can fire an arrow at 390 m/s. What time would an arrow take to hit a target 100 m away?

6 The highest speed on land in a car is 1190.4 km/h recorded by Stan Barrett (USA) in 1979 in his rocket-engined three-wheeled car at Edwards Airforce Base. What time would it have taken him to cover the 1.6 km test distance?

7 A person rides a bicycle to a shop by travelling 300 m north along a straight road and then travels west for another 400 m. If the trip takes 3 minutes, find **(a)** the average speed and **(b)** the average velocity.

8 A Ferrari Testarossa when driven by an experienced racing driver can cover 400 m from a standing start in 14.2 s. If it crosses the 400 m line at a speed of 203 km/h, what is its average speed?

DISPLACEMENT—TIME GRAPHS 2.4

It is often useful to show records of motion in the form of a graph. These can be in the form of a distance–time graph or as a displacement–time graph. In the graph shown in Figure 2.8 the position of a quarter-horse is shown, as recorded at six different times.

Photo 2.2
The Texas TI-83 graphing calculator and ranger has become a popular way of collecting and displaying data on the motion of objects, particularly in real-time.

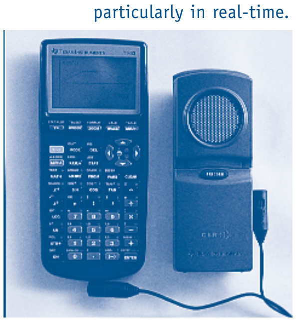


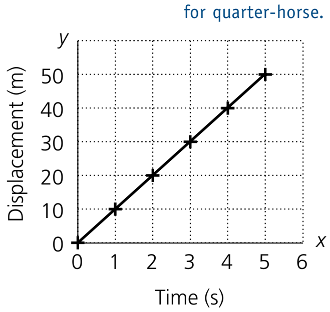
Table 2.3 DISPLACEMENT AND TIME MEASUREMENTS FOR A QUARTER-HORSE

Time elapsed (s)	0.0	1.0	2.0	3.0	4.0	5.0
Displacement (m)	0.0	10.0	20.0	30.0	40.0	50.0

In drawing the graph, it is usual to show the time elapsed on the *x*-axis and the displacement on the *y*-axis as in Figure 2.8.

Note that the six plotted points are the six observations of the quarter-horse. When we draw a line between these points we are assuming that the motion was uniform. This is called **interpolation** (Latin *inter* = 'between', *polire* = 'polish'; that is, to polish-up your data by supplying in-between points). When a line is extended past the first or last data points, this is called **extrapolation** (Latin *extra* = 'beyond').

Figure 2.8
Displacement–time graph for quarter-horse.



Questions

9 Table 2.4 records the motion of a dog chasing a ball. **(a)** Draw a displacement–time graph of the motion and describe it in words.

Table 2.4

Time elapsed (s)	0	1	2	3	4	5	6	7	8	9
Displacement (m)	0	2	4	4	4	6	6	4	2	0

(b) When was the dog stationary?
(c) When was its displacement increasing?
(d) When was the dog moving with the greatest speed?

2.5 SLOPE AND VELOCITY

Figure 2.9 is the displacement–time graph of a sprinter who runs 100 m in 10 s, rests for 20 s and then sprints back to the starting point in the next 30 s.

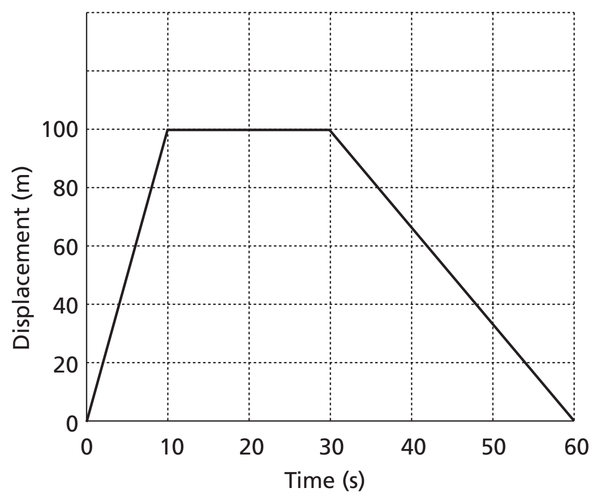


Figure 2.9
Displacement–time graph of sprinter.

The sprinter's average velocity in the first 10 seconds is calculated by dividing the displacement by the time taken. This is the same as calculating the slope of the line. The slope of any line is given by change in y divided by change in x ('rise over run'):

$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

From the above displacement–time graph, the slope is given by:

$$\boldsymbol{v}_{av} = \text{slope} = \frac{\Delta y}{\Delta x} = \frac{\text{change in position}}{\text{time taken}} = \frac{100\text{ m} - 0\text{ m}}{10\text{ s} - 0\text{ s}} = 10\text{ m s}^{-1}$$

The slope of the line is constant for the first 10 seconds, indicating that the velocity was also constant.

From $t = 30$ s to $t = 60$ s the average velocity can be calculated:

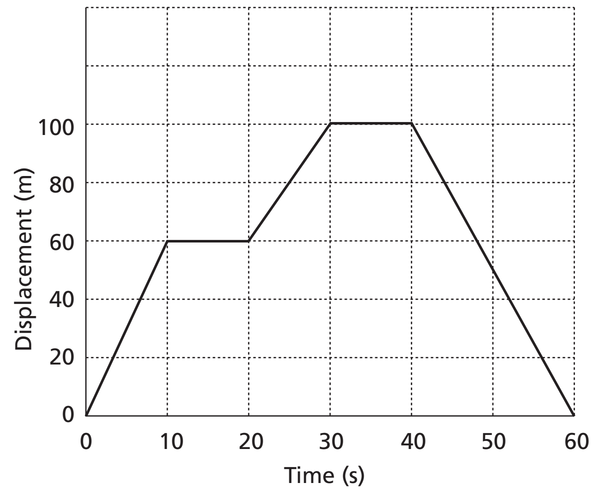
$$\boldsymbol{v}_{av} = \frac{0 - 100}{60 - 30} = 3.3\text{ m s}^{-1}.$$

Note that when the slope of the line is positive the velocity is in the positive direction. When the slope is negative the velocity is negative; this simply means that the direction of motion has reversed.

Questions

10 For the graph of a rollerskater shown in Figure 2.10:
(a) calculate his average velocity for each of the five sections of the graph;
(b) calculate his average velocity for the whole journey;
(c) calculate his average speed for the whole journey.

Figure 2.10
Displacement–time graph of rollerskater.



ACCELERATION

2.6

The velocity of a car increases when it starts moving from rest and decreases when the brakes are applied and it slows down. Cars can thus accelerate (speed up) or decelerate (slow down). The rate at which the velocity changes is called its **acceleration**. Consider the measurements of a car taking off from the traffic lights, shown in Table 2.5.

NOVEL CHALLENGE

The 08.00 express from Cleveland to Brisbane arrives at 9.00, and the 08.30 from Brisbane to Cleveland arrives at 9.30. Assuming both trains travel at constant speed, at what time should they pass each other?

Table 2.5

TIME ELASPSED (s)	DISPLACEMENT (m)	VELOCITY (m s^{-1})
0	0	0
1	1	2
2	4	4
3	9	6
4	16	8
5	25	10

The car's velocity is changing by 2 m s^{-1} every second. Its acceleration is said to be 2 m s^{-1} per second or 2 m s^{-2}. The formula for acceleration is then:

$$\text{Acceleration} = \frac{\text{change in velocity}}{\text{time taken}} = \frac{\Delta \boldsymbol{v}}{\Delta t}$$

$$= \frac{\text{final velocity} - \text{initial velocity}}{\text{time taken}}$$

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

where $\boldsymbol{v}$ is the final velocity and $\boldsymbol{u}$ is the initial velocity.

Note that the *change* of displacement is increasing for every second elapsed. In the 1st second, the displacement changes by 1 m, whereas in the 2nd second the displacement changes by 3 m. The above data are plotted on the three graphs shown in Figure 2.11. Graphs of uniformly accelerated motion are related as shown in the figure.

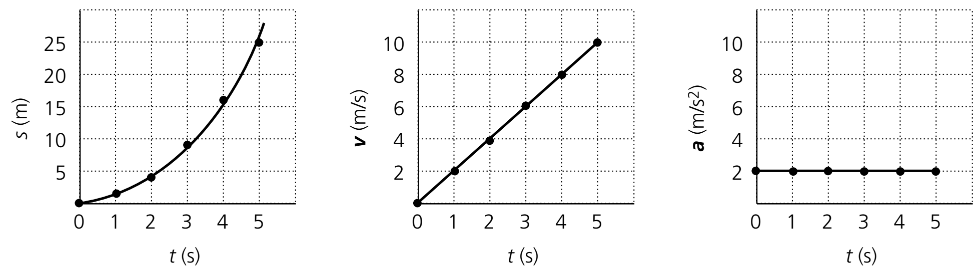


Figure 2.11

Questions

11 Plot an **s**–*t* graph and a **v**–*t* graph of the data of a ball rolling down an incline, listed in Table 2.6. Don't write in this book.

Table 2.6

TIME ELASPSED (s)	DISPLACEMENT (m)	VELOCITY (m s^{-1})
0	0	0
1	3	6
2	12	12
3	27	18
4	48	24
5	75	30

By inspection of the data, state the acceleration of the rolling ball.

Example

American experiments reveal that the beak of the red-headed woodpecker hits the bark of a tree at an impact velocity of 5.8 m s^{-1} and comes to rest in 0.059 s. Calculate the deceleration of the bird's head.

Solution

$$a = \frac{v - u}{t} = \frac{0 - 5.8}{0.059} = -98 \text{ m s}^{-2} \ (-9.8 \times 10^1 \text{ m s}^{-2})$$

The negative sign indicates that the bird slowed down.

Questions

12 Complete Table 2.7. This will give you practice at manipulating the equation for acceleration. Don't write in this book.

Table 2.7

	v (m s^{-1})	*u* (m s^{-1})	Δv (m s^{-1})	*t* (s)	*a* (m s^{-1})
(a)	18	10		2.0	
(b)	42		4		4.0
(c)		20		10	−2.0
(d)	18	25		3.5	
(e)		−5		1.3	−0.5

NOVEL CHALLENGE

Here's an interesting theory that could be investigated experimentally. R. McNeill Alexander from Leeds University, England, measured the speed at which animals switched from walking to running. For humans, the speed is about 8 km h^{-1}. He developed a rule which, stated mathematically, is: $v^2 = \frac{1}{2} g d_H$, where **v** is the speed at which an animal switches, d_H is the distance from the hip to the ground, and ***g*** is the acceleration due to gravity. His rule applies to animals from insects to humans. Can you confirm this rule by experiment?

13 The highest road-tested acceleration reported for a standard production car is 0 to 96.5 km/h (26.8 m s^{-1}) in 3.98 s for a Ferrari F40 driven by Mark Hales of *Fast Lane Magazine* in the UK on 9 February 1989. Calculate the acceleration of the car.

14 The highest speed by a rocket-engined wheeled land vehicle was 1046 km h^{-1} (290 m s^{-1}) recorded by Gary Gabelich in The Blue Flame on the Bonneville Salt Flats in 1970. His acceleration was measured as 4.2 m s^{-2} in getting to this speed from rest. How many seconds would he have taken to reach this speed?

15 The head of a rattlesnake can accelerate at 50 m s^{-2} when striking a victim. If a car could do as well, how long would it take for it to reach a speed of 27 m s^{-1} (100 km h^{-1}) from rest?

16 A muon (an elementary particle) enters an electric field with a speed of 5.00×10^6 m s^{-1}, whereupon the field causes it to decelerate at 1.25×10^{14} m s^{-2}. How much time elapses before it stops?

2.7 INSTANTANEOUS VELOCITY

When you read a car's speedo you are seeing the instantaneous speed of the car. If it reads 60 km h^{-1}, then it means that at the current speed you would cover 60 km in 1 hour. But you could be accelerating and the speedo might be gradually changing from 50 km h^{-1} to 100 km h^{-1}. When it read 60 km h^{-1} this was its **instantaneous speed**. If a direction is also specified, then you would be talking about its **instantaneous velocity**.

Consider the case of an accelerating car. In this case the velocity is getting faster as time goes by so the *s*–*t* graph is a curve as shown in Figure 2.12.

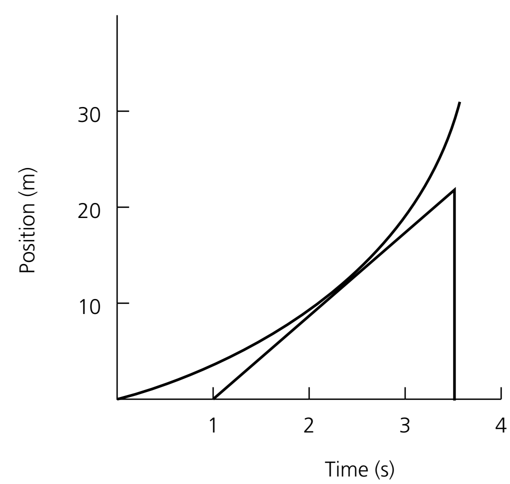


Figure 2.12(a) The instantaneous velocity at time t = 2.5 s is given by the slope of the tangent to the curve at that point.

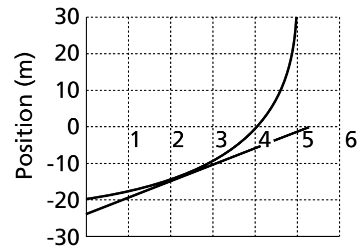
To calculate the instantaneous velocity at 2.5 s in Figure 2.12**(a)**, a tangent is drawn to the curve at the 2.5 s mark. The tangent is a line that just touches the curve at that point. The slope of the tangent can be calculated:

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{20 - 0}{3.5 - 1} = 8 \text{ m s}^{-1}$$

A more difficult case is shown in Figure 2.12**(b)**. To calculate the instantaneous velocity at 2 seconds in Figure 2.12**(b)**, a tangent to the curve has been drawn and the slope of the tangent can be calculated:

$$\text{Slope} = \frac{0 - -23}{5.2} = 4.4 \text{ m s}^{-1}$$

Figure 2.12(b)



Question

17 From Figure 2.12b: **(a)** Calculate the instantaneous velocity at 4 s. **(b)** Calculate the average velocity over the whole 5 s. Do not draw in this book. Use your ruler.

2.8 VELOCITY–TIME GRAPHS

Graphs can also be used to show the changes in velocity of an object with time. The graph in Figure 2.13 represents a car being accelerated from rest to 20 m s^{-1} in 10 s and being held at that speed for 10 s before the driver slows down to a stop.

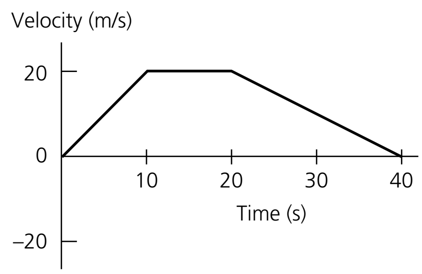


Figure 2.13

A straight line sloping upward indicates constant acceleration, whereas a straight line sloping down indicates deceleration or negative acceleration. A horizontal line indicates zero acceleration, that is, constant velocity. As the formula for acceleration $\boldsymbol{a} = \dfrac{\boldsymbol{v} - \boldsymbol{u}}{t}$ is equivalent to $\dfrac{\Delta y}{\Delta x}$ then acceleration is equal to the slope of a *v*–*t* graph.

The displacement can be calculated by finding the area under the line. For instance, in the case above, the car has travelled at an average speed of 10 m s^{-1} for the first 10 s. Hence the displacement must be 10 m s^{-1} × 10 s = 100 m. The area under the line for the first 10 s is (20 × 10)/2, that is, (base × height)/2, which equals 100 m also. **The area under a *v*–*t* graph equals displacement.**

Example

Using the graph shown in Figure 2.13:

(a) Calculate the acceleration of the car at **(i)** 5 s, **(ii)** 15 s and **(iii)** 30 s.
(b) Calculate the displacement after 40 s.
(c) Calculate the average velocity.
(b) Sketch an acceleration–time graph.

Solution

(a) **(i)** The acceleration at 5 s is equal to the slope at 5 s:

$$\boldsymbol{a} = \text{slope} = \frac{20 - 0}{10} = 2 \text{ m s}^{-2}$$

(ii) Slope equals zero, therefore acceleration is zero.

(iii) Slope $= \dfrac{0 - 20}{40 - 20} = -1 \text{ m s}^{-2}$.

NOVEL CHALLENGE

You have learnt that the rate of change of position with respect to time is **velocity**, and the rate of change of velocity is **acceleration.** Did you know that the rate of change of acceleration is known as **jerk** (symbol ***j***)? Jerk is important when evaluating the destructive effect of motion on a mechanism, or the discomfort caused to passengers in a vehicle. The movement of delicate instruments needs to be kept within specified limits of jerk as well as acceleration to avoid damage. When designing a train the engineers will typically be required to keep the jerk less than 2 m s^{-3} for passenger comfort. In the aerospace industry they even have such a thing as a jerkmeter — an instrument for measuring jerk.
In the case of the Hubble space telescope, the engineers specified limits on the magnitude of the rate of change of jerk. There is no universally accepted name for this fourth derivative.
Is the slope of, or area under, an *a*–*t* graph related to jerk? Does the slope of, or area under, a jerk–time graph mean anything?

Ch02-Walding 4th 25/8/04 10:14 AM Page 37

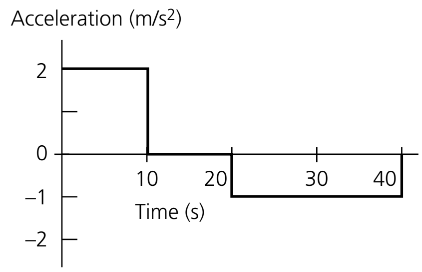
(b) Total area = $\frac{20 \times 10}{2} + 20 \times 10 + \frac{20 \times 20}{2} = 500$ m (or 5×10^2 m).

(c) Average velocity = displacement ÷ time:

$$v_{av} = \frac{s}{t} = \frac{500}{40} = 12.5 \text{ m s}^{-1} \text{ (or 10 m s}^{-1} \text{ to one significant figure)}$$

(d) See Figure 2.14.

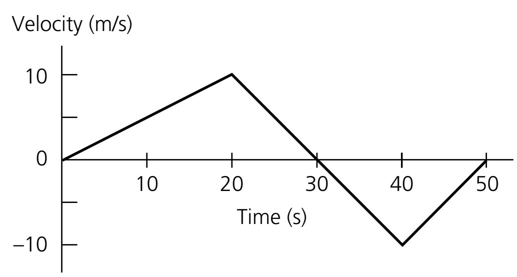
Figure 2.14



For cases where the velocity becomes negative, the area beneath the *x*-axis is also negative and this must be taken into account when calculating displacement.

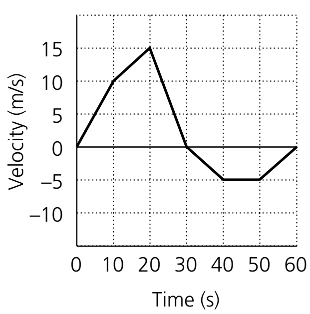
For example, imagine the motion of a bungee jumper, jumping off a tower (Figure 2.15).

Figure 2.15



The **displacement** after 50 s is $\frac{30 \times 10}{2} + \frac{20 \times -10}{2} = 50$ m. The **distance** travelled, however, is not a vector quantity and the area underneath the *x*-axis is not considered to be negative. The distance travelled is 250 m (150 m down plus 100 m back up).

Figure 2.16
For question 19.



Questions

18 For the motion of the bungee jumper shown in Figure 2.15 above:
- **(a)** calculate the displacement and distance travelled after 40 s;
- **(b)** calculate the acceleration at 10 s, 30 s and 45 s;
- **(c)** sketch an acceleration–time graph.
- **(d)** When was he stationary?
- **(e)** When was his acceleration constant but not zero?
- **(f)** When was his velocity constant but not zero?

19 The graph shown in Figure 2.16 illustrates the motion of a skateboard rider.
- **(a)** Calculate his displacement after 1 minute.
- **(b)** Calculate how far he travelled in the minute.

(c) At what stage was the magnitude of his acceleration the greatest?
(d) When was he stationary?
(e) When was his acceleration negative and constant?
(f) When was his velocity constant but not zero?

20 Olympic equestrian 'Three-day eventing' is held over 4 days. The first 2 days consist of dressage while the 4th day is for show-jumping. The 3rd day is the speed and endurance section. At the 1996 Atlanta Olympics, the gold medallist achieved these results for Day 3:
Stage 1 (The Trot) was at 13 km h^{-1} for 10 minutes followed by Stage 2 (The Fast Steeplechase), which took 5 minutes at 41 km h^{-1}. Stage 3 was another trot the same as Stage 1. Before Stage 4 there was a compulsory 10 minute rest. Stage 4 was a testing 14-minute cross-country gallop at 34 km h^{-1}.
(a) Draw a $\boldsymbol{v}$–t graph of the motion.
(b) Calculate the total distance travelled in this event.

2.9 EQUATIONS OF MOTION

The equations used so far can be combined to provide other useful ways of calculating and describing the motion of objects.

In real life we encounter several main kinds of motion:

- Constant velocity (zero acceleration).
- Regularly changing velocity (constant acceleration).

Case 1: Constant velocity

The simplest kind of motion we can study is that in which the object moves with constant velocity and hence zero acceleration. The graphs for this type of motion are illustrated in Figure 2.17. Some examples drawn from everyday life are:

- a car being driven at 60 km h^{-1}
- ball bearings being rolled on a very smooth horizontal surface
- a person jogging
- water flowing in a pipe.

Accelerated motion is also easy to find. Examples are:

- objects falling freely under gravity
- a car moving away from the traffic lights
- an aircraft being catapulted by a steam catapult from an aircraft carrier.

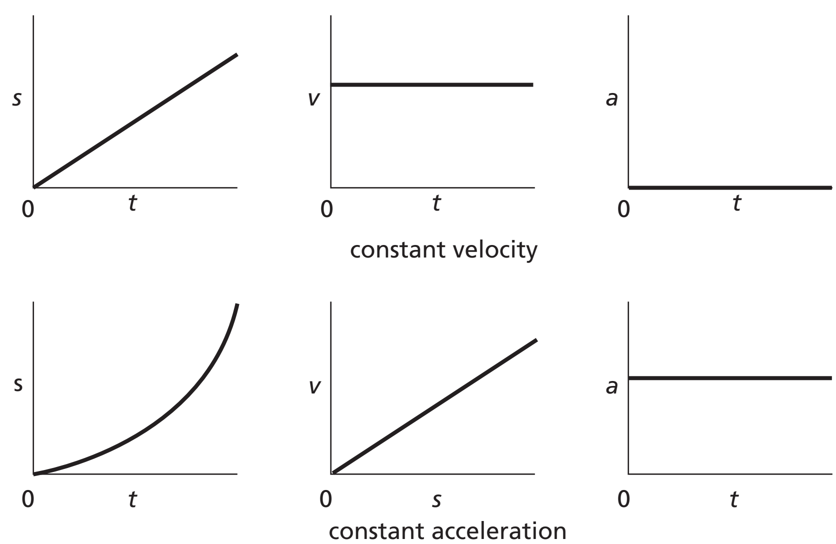
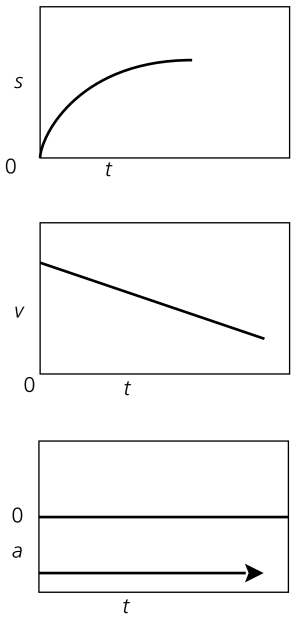


Figure 2.17
Motion graphs showing corresponding displacement–time, velocity–time and acceleration–time relations for situations of constant velocity and constant acceleration.

Figure 2.18



Case 2: Constant (uniform) acceleration

Graphs representing this type of motion are also shown in Figure 2.17. Objects falling freely under gravity are the most common examples of this.

Another case of constant acceleration is for an object slowing down (decelerating or negative acceleration). Figure 2.18 shows graphs of this motion.

The quantities displacement, time, velocity and acceleration are all related to each other. In this book the symbols shown in Table 2.8 will be used.

Table 2.8

QUANTITY	QUANTITY SYMBOL	UNIT	UNIT SYMBOL
Displacement	s	metres	m
Initial velocity	u	metres per second	m s^{-1}
Final velocity	v	metres per second	m s^{-1}
Acceleration	a	metres per second per second	m s^{-2}
Time (elapsed)	t	second	s

Development of formulas

1 Acceleration $= \dfrac{\text{final velocity} - \text{initial velocity}}{\text{time}}$:

$$a = \frac{v - u}{t} \quad \text{or} \quad v = u + at \qquad (1)$$

2 Average velocity $= \dfrac{s}{t}$ and also equals $\dfrac{u + v}{2}$

$$\frac{u + v}{2} = \frac{s}{t} \quad \text{or} \quad s = \frac{(u + v)t}{2} \qquad (2)$$

3 If we substitute equation (1) into (2) we get:

$$s = \frac{(u + (u + at))t}{2} \quad \text{or} \quad s = ut + \tfrac{1}{2}at^2 \qquad (3)$$

4 From equation (1) we get $t = \dfrac{v - u}{a}$. Substituting this into equation (2), we get:

$$s = \frac{u + v}{2} \times \frac{v - u}{a} \quad \text{or} \quad 2as = (u + v)(v - u)$$

$$2as = v^2 - u^2$$

$$v^2 = u^2 + 2as \qquad (4)$$

Note: these formulas only apply when the acceleration is constant and the motion is in a straight line. Velocity, acceleration and displacement are vector quantities and therefore may be positive or negative.

We can finally summarise the equations of motion as listed in Table 2.9.

NOVEL CHALLENGE

A car travels from A to B at an average speed of 100 km/h and returns at 60 km/h. *What is the average speed for the journey?*

Table 2.9 SUMMARY OF FORMULAS

MOTION WITH UNIFORM VELOCITY	MOTION WITH UNIFORM ACCELERATION
$\boldsymbol{v}$ = constant $\boldsymbol{v}_{av} = \frac{\boldsymbol{s}}{t}$ $\boldsymbol{s} = \boldsymbol{v}t$	$\boldsymbol{a}$ = constant $\boldsymbol{v}_{av} = \frac{\boldsymbol{v} + \boldsymbol{u}}{2}$ $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}t$ $\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$ $\boldsymbol{v}^2 = \boldsymbol{u}^2 + 2\boldsymbol{as}$ $\boldsymbol{s} = \frac{(\boldsymbol{u} + \boldsymbol{v})t}{2}$

Example 1

A car starts from rest and reaches a velocity of 60 km h^{-1} (16.67 m s^{-1}) in 8 seconds. Assuming the acceleration to be constant, calculate **(a)** the acceleration and **(b)** the displacement in this time interval.

Solution

Data: $\boldsymbol{u} = 0$, $\boldsymbol{v} = 16.67$ m s^{-1}, $t = 8$ s, $\boldsymbol{a} = ?$, $\boldsymbol{s} = ?$

(a) $$\boldsymbol{a} = \frac{\boldsymbol{v} - \boldsymbol{u}}{t} = \frac{16.67 - 0}{8} = 2.08 \text{ m s}^{-2}.$$

(b) $$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2 = 0 + \frac{1}{2} \times 2.08 \times 8^2 = 66.6 \text{ m}.$$

Example 2

A train starting from rest travels 30 m in 6 s. Find **(a)** its acceleration and **(b)** its velocity after the 6 s.

Solution

Data: $\boldsymbol{u} = 0$, $\boldsymbol{s} = 30$ m, $t = 6$ s, $\boldsymbol{a} = ?$, $\boldsymbol{v} = ?$

(a)
$$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$30 = 0 + \frac{1}{2}\boldsymbol{a}6^2$$
$$30 = 18\boldsymbol{a}$$
$$\boldsymbol{a} = 1.67 \text{ m s}^{-2}.$$

(b)
$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}t$$
$$= 0 + 1.67 \times 6$$
$$= 10 \text{ m s}^{-1}.$$

Questions

21 Table 2.10 will give you practice in selecting equations of motion and substituting values into them. Complete the table but do not write in this book.

NOVEL CHALLENGE

A column of troops 3 km long is marching along a road. An officer rides from the rear to the head of the column and back once, and he reaches the rear of the column just as an advance of 4 km has been made from where he first left.
How far did he ride?

NOVEL CHALLENGE

A man goes from A to B at 30 km/h.
How fast must he return to average 60 km/h for the whole trip?

NOVEL CHALLENGE

A boy is carried up an escalator in 1 minute. He can walk up a stationary escalator in 3 minutes. *How long will it take him to walk up a moving escalator?*

Table 2.10

QUESTION	s (m)	u (m s^{-1})	v (m s^{-1})	a (m s^{-1})	t (s)
(a)		0		2.5	3
(b)	100	0			2.4
(c)		10	25	2	
(d)	300	9		1.5	
(e)	40	2		4	
(f)		10	5		2.5
(g)	160	50			8

22 A cyclist starts from rest and attains a velocity of 21 m s^{-1} in 3.5 seconds. Calculate **(a)** the acceleration, assumed constant; **(b)** the displacement.

23 A bus travelling in a straight line accelerates from 60 km h^{-1} to 100 km h^{-1} in 1 minute. Calculate the acceleration in m s^{-2}.

24 The click beetle (*Athous haemorrhoidalis*) experiences an acceleration of 24 000 m s^{-2} over a distance of 5 mm when it jack-knifes into the air to avoid predators. For what time duration does this acceleration occur?

ACCELERATION DUE TO GRAVITY 2.10

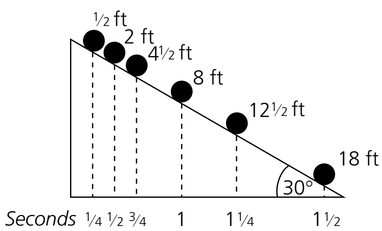


Figure 2.19 Galileo's data from his inclined plane experiments.

One of the most common examples of motion in a straight line with uniform acceleration is that of an object that falls freely due to gravity. Until Galileo (1564–1642), people thought that heavy objects fell faster than light objects. They saw no need for experiments that may have confirmed or refuted these beliefs. They relied on the theories of Aristotle, who believed that objects fell at speeds that depended on their weight. Galileo performed some of the earliest experiments, which showed that both heavy and light objects in the absence of air and other resistance fell with constant acceleration. Thus, two objects of different masses, dropped from the same height at the same time, should strike the ground simultaneously.

Motion due to gravity can take two main forms. The first is vertical motion, where the object moves in one dimension only, that is, up and down. The second is projectile motion, where the object moves horizontally as well as vertically, for example a stone thrown off a cliff. Only vertical motion will be dealt with in this chapter. Projectile motion will be discussed in Chapter 5.

Types of free-fall motion

Free-fall motion can be grouped into two classes:

1 The object is being dropped or thrown down.
2 The object is being thrown upward.

Positive and negative convention When dealing with calculations involving acceleration due to gravity we need to assign a positive and negative direction of motion. In this chapter we will use the convention in which **up is positive**. Throughout the world, this is the most common. It is a matter of your choice, however, but you may find it simplest to stay with the one convention.

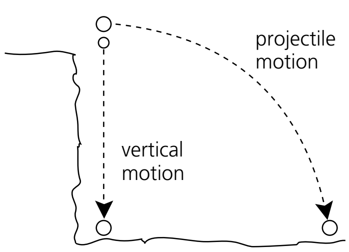


Figure 2.20 Vertical and projectile motion.

Acceleration due to gravity is constant at 10 m s^{-2} in the negative direction (down), hence $a = -10$ m s^{-2}. This means that an object will increase its velocity in the negative direction by 10 m s^{-1} every second, or by 10 metres per second *per second*. Students often think that a negative acceleration means *de*celeration or slowing down but this is not always so. If an object is moving in the negative direction (down) and has negative acceleration then it will get faster in that negative direction. If it is moving in the positive direction (upward) and has a negative acceleration then it is slowing down in that positive direction.

Case 1: Dropped or thrown down

Most typically, these situations involve dropping a rock off a cliff or throwing something vertically downward. In both cases the velocity increases. The only difference is the initial velocity. When dropped, the initial velocity is zero but when thrown down the velocity begins at some negative value. Either way, the velocity increases in the negative direction.

Example 1

A spanner is dropped from a sixth-floor window and takes 2.2 s to hit the ground. Calculate **(a)** the height from which it was dropped and **(b)** its impact velocity.

Solution

Take the downward direction as negative.
Data: $\boldsymbol{u} = 0$ m s^{-1}; $\boldsymbol{a} = -10$ m s^{-2}; $t = 2.2$ s; $\mathbf{s} = ?$; $\boldsymbol{v} = ?$

(a)

$$\mathbf{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$= 0 + \frac{1}{2} \times -10 \times 2.2^2$$
$$= -24.2 \text{ m}$$

(b)

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}t$$
$$= 0 + -10 \times 2.2$$
$$= -22 \text{ m s}^{-1}$$

Example 2

The Zero Gravity Research Facility at the NASA Research Centre includes a 150 m drop tower. This is an evacuated vertical tower through which a 1 m diameter sphere can be dropped. If this sphere is projected downward at an initial speed of 5 m s^{-1}, how long would it take to reach the bottom?

Solution

Data: $\mathbf{s} = -150$ m; $\boldsymbol{u} = -5$ m s^{-1}; $\boldsymbol{a} = -10$ m s^{-1}; $t = ?$

$$\mathbf{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$-150 = -5 \times t + \frac{1}{2} \times -10 \times t^2$$
$$5t^2 + 5t - 150 = 0$$
$$t^2 + t - 30 = 0$$
$$(t - 5)\ (t + 6) = 0$$

Hence $t = -6$ s or $t = +5$ s. The answer must be 5 s as the negative time is not meaningful here.

Note: in cases where the quadratic equation doesn't factorise simply as shown above, the quadratic formula will be needed:

$$\text{Quadratic formula: } x = \frac{-b \pm \sqrt{b^2 - 4\boldsymbol{a}c}}{2\boldsymbol{a}}$$

Without the quadratic formula you would first need to calculate $\boldsymbol{v}$.

NOVEL CHALLENGE

If you put a row of coins on a 1 metre ruler that has one end on the ground and let the other end fall, which coins will stay on the ruler and which ones will be left behind?

PHYSICS FACT
THE POWER OF MINUS

(a) In 1962, the Mariner I mission launched towards Venus but the rocket separated from the boosters too soon and plunged into the ocean 4 minutes after take-off. Some klutz left a negative (–) sign out of the computer program.

(b) The old equation for the energy of a photon was $^1/_2\ mv^2 = hf$. Einstein added $-W$ and got a Nobel Prize.

Example 3

A person aboard a balloon moving downward at 30 m s^{-1} drops a sandbag at an elevation of 500 m. **(a)** What time will it take for the sandbag to hit the ground? **(b)** What will be the speed of the bag on impact?

Solution

Data: $\boldsymbol{s} = -500$ m; $\boldsymbol{u} = -30$ m s^{-1}; $\boldsymbol{a} = -10$ m s^{-2}; $t = ?$

NOVEL CHALLENGE

A flea (*Pulex irritans*) can jump about 4 m high. If the flea was a big as a person, how high would it be able to jump (proportionally)?

(a)

$$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$

$$-500 = -30t + \frac{1}{2} \times -10 \times t^2$$

$$t^2 + 6t - 100 = 0$$

$$t = \frac{-b \pm \sqrt{b^2 - 4\boldsymbol{a}c}}{2\boldsymbol{a}} = \frac{-6 \pm \sqrt{6^2 - 4 \times 1 \times -100}}{2 \times 1}$$

$$= \frac{-6 \pm 20.9}{2}$$

$$= -13.4 \text{ s or } +7.4 \text{ s}$$

The negative time has no real meaning in this case, so the answer is 7.4 s.

(b)

$$\begin{aligned} \boldsymbol{v} &= \boldsymbol{u} + \boldsymbol{a}t \\ &= -30 + -10 \times 7.4 \\ &= -30 + -74 \\ &= -104 \text{ m s}^{-1} \end{aligned}$$

Remember, the magnitude of the acceleration due to gravity ($\boldsymbol{g}$) is about 9.8 m s^{-2}. This means that an object falling freely under gravity increases its speed by about 10 m s^{-1} every second. It is given the negative sign because we have adopted the convention that upward is positive and downward is negative.

SR Activity 2.4 VERTICAL MOTION ON THE SPREADSHEET

If you have access to a computer and are familiar with spreadsheeting, set up a spreadsheet with the following headings (Table 2.11):

Table 2.11 SPREADSHEET

	A	B	C
1	t (s)	s (m)	**v** (m s^{-1})
2	0	0	0
3	1		
4	2		

1. The formula for cell B2 would be = (0.5 * 9.8*A2*A2) for example.
2. Extend Column A to 20 seconds and compute the value for all cells.
3. Use the graph commands to draw s–t and $\boldsymbol{v}$–t graphs. Are they what you would expect?
4. Discuss your output.

Case 2: Throwing an object upward

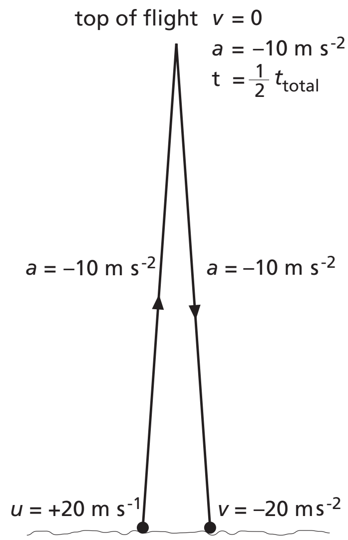
When a ball is thrown vertically upward, it starts at a high initial velocity in the positive direction, gradually slows to a halt at the top of its flight and gradually increases velocity in the negative direction until it returns to the ground.

Figure 2.21 shows the flight of the ball; although its downward path is exactly the same as the upward path, it is drawn slightly to the right for clarity.

Note: it can be shown that:

- velocity equals zero at the top of flight
- time of flight up equals time down
- acceleration is constant even at the top of flight when velocity is zero
- initial speed equals final speed
- final velocity equals the negative of the initial velocity
- air resistance is negligible and can be neglected.

Figure 2.21
Trajectory of an object thrown vertically.



Example

A ball is thrown vertically upward at 20 m s^{-1}. Ignoring air resistance and taking $g = -10$ m s^{-2}, calculate **(a)** how high it goes; **(b)** the time taken to reach this height; **(c)** the time taken to reach the ground from the highest point; **(d)** the final velocity; **(e)** time of flight.

Solution

Data: Taking down as negative: $\boldsymbol{u} = +20$ m s^{-1}, $\boldsymbol{a} = -10$ m s^{-2}, $\mathbf{s} = 0$ m.

(a) At the top of flight $\boldsymbol{v} = 0$ m s^{-1}:

$$\boldsymbol{v}^2 = \boldsymbol{u}^2 + 2\boldsymbol{as}$$
$$0 = (+20)^2 + 2 \times -10 \times \boldsymbol{s}$$
$$20\boldsymbol{s} = 400$$
$$\boldsymbol{s} = 20 \text{ m (i.e. 20 m up in the air).}$$

(b)

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}t$$
$$0 = +20 + -10t$$
$$t = 2 \text{ s}$$

(c) The ground is 20 m in the negative direction from the top of flight. Hence $s = -20$ m:

$$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$-20 = 0 + -5t^2$$
$$t = 2 \text{ s}$$

(d)

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}t$$
$$= 0 + -10 \times 2$$
$$= -20 \text{ m s}^{-1}$$

(e) Time up = 2 s; time down = 2 s. Hence total time of flight equals 4 s. Using the equations of motion it can be shown that when the displacement is zero, the time for this to occur is zero seconds (the start) and 4 seconds (the finish):

$$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$0 = +20t + \frac{1}{2} \times -10t^2$$
$$5t^2 = 20t$$
$$t = 4 \text{ s}$$

Questions

25 A rock is dropped off a cliff and it takes 4 s to reach the base below. How high is the cliff?

26 A pot-plant falls 25 m from rest to the ground below.
(a) What is its impact velocity?
(b) What time did it take to fall?

27 A rock is launched vertically upward from the ground at a starting speed of 35 m s^{-1}.
(a) What is the maximum height reached?
(b) What time does it take to reach this maximum height?
(c) What time does it take to fall back to the ground again?

28 A person in a balloon moving vertically upward at a constant speed of 4.9 m s^{-1} drops a sandbag at an elevation of 98 m.
(a) What time will it take until the sandbag hits the ground?
(b) What will be the velocity of the sandbag on impact?

29 A startled armadillo leaps upward and rises 54.4 cm in 0.20 s and keeps rising.
(a) What was its initial speed? **(b)** What is its speed at this height?
(c) How much higher does it go?

GRAPHS OF FREE-FALL MOTION 2.11

The two most common types of free-fall motion mentioned in the previous section can be examined graphically. **Case 1** is that of an object dropped off a cliff. Figure 2.22 shows the relationship between a velocity–time graph **(b)** and its corresponding acceleration–time graph **(a)** for this type of free-fall motion. The downward direction is negative.

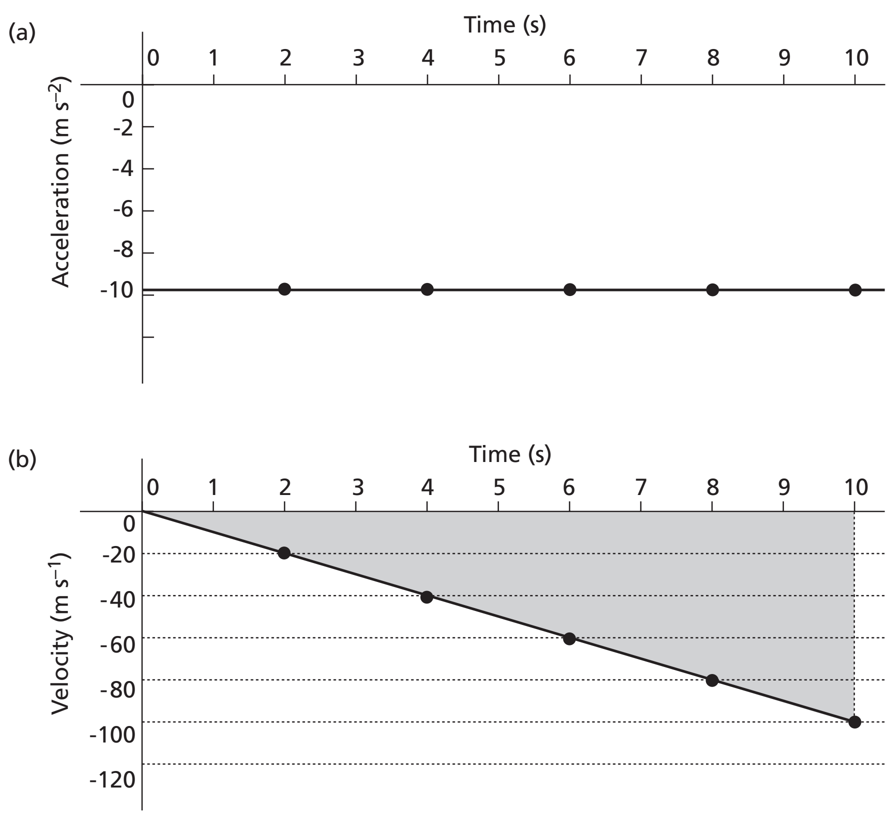


Figure 2.22
(a) An acceleration–time graph;
(b) a velocity–time graph.
(The shaded area indicates the displacement.)

Ch02-Walding 4th 25/8/04 10:14 AM Page 46

Case 2 is that of an object thrown upward into the air and allowed to return to its starting place. Figure 2.23 shows the graphs of motion of a ball thrown in this manner. Note that the acceleration is constant, even at the top of flight when the ball is stationary. Again, down is negative.

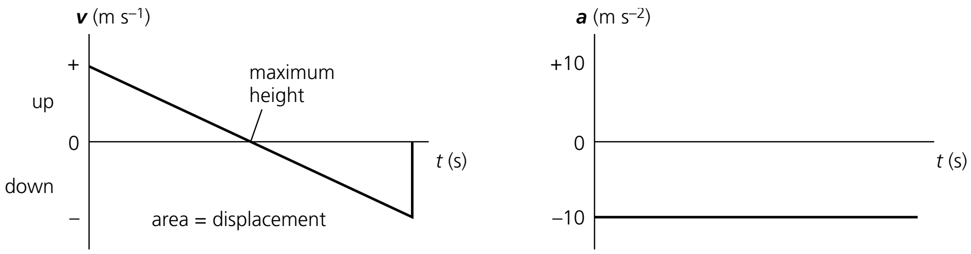


Figure 2.23

Questions

30 What does the slope of the line in the ***v***–*t* graph of Case 1 (Figure 2.22) represent?

31 Which one of the graphs in Figure 2.24 is the displacement–time graph of the rock's motion in Case 2 (Figure 2.23)?

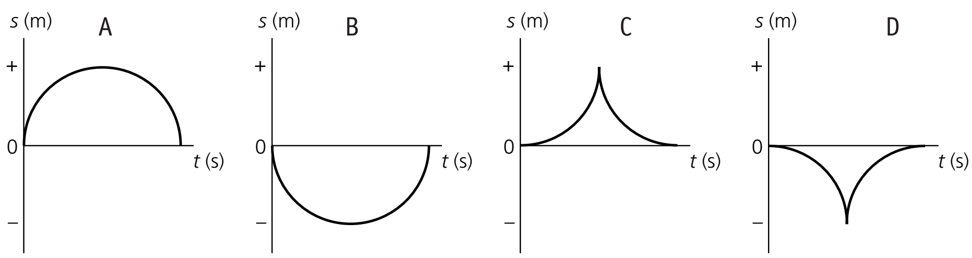


Figure 2.24

32 Draw a displacement–time graph of the motion of the ball as described in Case 1 (Figure 2.22).

2.12 MEASURING MOTION

Motion of an object can be recorded by using a **ticker timer** as shown in Photo 2.3. It has been specifically designed for physics experiments and has little other use outside the physics laboratory. It consists of a pointed hammer, which vibrates up and down 50 times per second. When a paper tape is pulled through the timer, a piece of carbon paper allows an imprint of the hammer to be made on the paper. The distance between successive dots can be used to calculate the velocity of the moving object as the time interval is a constant $\frac{1}{50}$ of a second (0.02 second).

Consider a section of tape as shown in Figure 2.25. Table 2.12 lists the data from the tape.

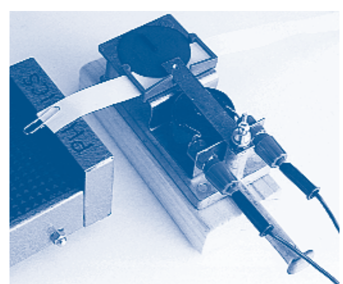


Photo 2.3
A ticker timer.

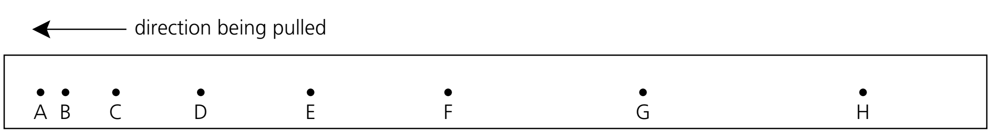


Figure 2.25
A segment of ticker timer tape.

Table 2.12

DOT	A	B	C	D	E	F	G	H
t (seconds)	0	0.02	0.04	0.06	0.08	0.10	0.12	0.14
s (cm)	0	0.3	1.1	2.6	4.6	7.1	10.3	14.0
v (cm/s)	0	27.5	57.5	87.5	112.5	142.5	172.5	–

The average velocity can be determined by dividing the total displacement (14.0 cm) by the time elapsed (0.14 s) to give 100 cm s^{-1}. The instantaneous velocity at each dot can be calculated by measuring the distance travelled between dots either side of the one being considered. For example, to calculate the velocity at dot D, the distance between dots C and E is measured (3.5 cm — see Figure 2.26) and this is divided by the time interval (2×0.02 s). The velocity at D is thus 87.5 cm s^{-1}. The velocity of the other dots can also be calculated.

Figure 2.26
Tape.

If acceleration is constant, a graph of velocity vs time should be linear and the slope of this line will equal the acceleration.

To calculate the acceleration at a point, the velocity at the dot before this point and at the dot after the point should be determined. The difference ($\boldsymbol{v} - \boldsymbol{u}$), when divided by the time elapsed, will equal the acceleration.

For example, the acceleration at point D can be calculated by subtracting the velocity at C from the velocity at E and dividing by 0.04 seconds: $\boldsymbol{v}_C = 57.5$ cm s^{-1}, $\boldsymbol{v}_E = 112.5$ cm s^{-1}, hence $\Delta\boldsymbol{v} = \boldsymbol{v}_E - \boldsymbol{v}_C = 55$ cm s^{-1}. The result: $\boldsymbol{a}_D = \dfrac{\Delta \boldsymbol{v}}{t} = \dfrac{55}{0.04} = 1375$ cm s^{-2} is the acceleration at D. The acceleration at E likewise is 1375 cm s^{-2}. You should check this for yourself.

Questions

33 The following questions refer to the section of tape described above. (Figure 2.26)

(a) Plot the displacement vs time graph of the data above.
(b) Calculate the slope of the graph at points C and F. How do these slopes compare with the calculated velocity at these points?
(c) Plot the velocity vs time graph of the above data.
(d) Calculate the area under the line to point G. How does it compare with the displacement at G (10.3 cm)?
(e) Calculate the slope of the velocity vs time graph. How does it compare with the calculated acceleration (1375 cm s^{-2})?

34 The following questions refer to the ticker tape shown in Figure 2.27.

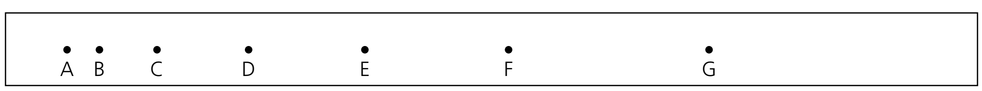


Figure 2.27
For question 34.

(a) Draw up a data table similar to Table 2.10 and measure the displacements using your ruler. Enter the displacements in your data table. Do not write in this book.
(b) Plot a displacement–time graph.
(c) Calculate the slope of the tangent at point E.
(d) Calculate the velocity at each dot and add to the data table.
(e) How does the velocity at E compare with the slope at E on the *s*–*t* graph?
(f) Plot a velocity–time graph and draw a line of best fit.
(g) Calculate the slope of the line.
(h) Calculate the acceleration at points B, C, D and E and add these to the data table.
(i) How does this compare with the slope of the $\boldsymbol{v}$–t graph?
(j) Calculate the displacement at point F by measuring the area under the $\boldsymbol{v}$–t graph. How does it compare with the actual displacement at F as measured on the tape?

Ch02-Walding 4th 25/8/04 10:14 AM Page 48

2.13 ELECTRONIC RECORDING AND COMPUTER INTERFACING

There are several devices that enable motion to be recorded electronically. Data-loggers are used extensively in research and industry to monitor the performance of various devices under test. The data-logging system comes with a package including an interface system and various sensors to pick up data from the environment such as motion, temperature, voltage, sound and light. The sensor converts physical or chemical changes into electrical signals; these **analog** signals are carried to the interface system where the signals are digitised. Such **digital** signals can be analysed and displayed on the computer monitor and calculations can be performed. The graphical display function that accompanies data-logging programs transforms data into graphs, which help show trends and anomalies.

Data-loggers are used for measuring not only motion but also an enormous range of other data. You will have heard of the heart monitors in hospitals and 'black box' flight recorders in planes. But they are also used for purposes as diverse as designing and producing torpedos, counting biscuits on a production line, measuring causes of stress on individual soldiers in combat situations, and monitoring the drying of paint and curing in industrial ovens. Data-logging equipment is in use at smelters, refineries, tailings dams, mines, landfills, construction sites, manufacturing and processing plants, and industrial and hazardous waste sites; and meteorological conditions can be monitored to yield data for determining air stability or for use in air quality and dispersion modelling.

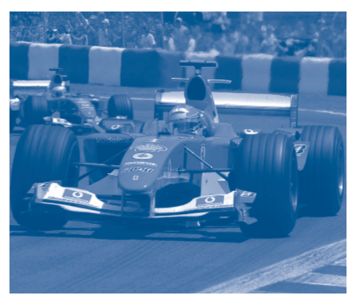
SR NEI Activity 2.5 DATA-LOGGER IN MOTORSPORT

Try the following as a good stimulus response task, or it could be the start of a non-experimental investigation.

Racing cars

An interesting use of data-logging is in the racing car industry. Car manufacturers need to run their cars at high speeds for predetermined times as part of their endurance testing program, so they are packed with temperature and pressure sensors that feed data into computers.

Photo 2.4
Formula One racing cars have a huge number of transducers being monitored by data-loggers to give them a winning edge. Shown here is world champion Michael Schumacher in his Ferrari F2002, winning the French Grand Prix.

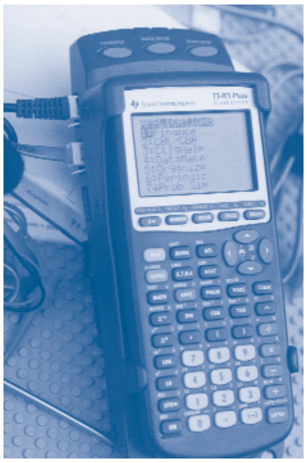


The success of a racing car depends on hundreds of components working together at peak performance under the most extreme conditions. Components such as displacement sensors are designed to control and monitor a growing number of vital functions on racing cars and supply information to engineers, who can then help trim precious seconds off the car's lap times. Although most categories of motor racing do not allow the performance of the suspension to be modified during a race, the use of computerised data-logging in testing and practice allows race engineers to tune the suspension to match the particular conditions and type of circuit.

Monitoring the movement of the suspension with displacement sensors allows electrical signals to feed back to the logging/telemetry system and then display a graphical representation of the car's performance around a track. Using the data, engineers can easily recognise areas where improvements can be made, and fine-tune the car by adjusting ride heights and stiffness to suit a particular track and driver. Movement of the suspension can usually be sensed by a linear displacement sensor, but some need rotary sensors.

Throttle controls have a rotary motion, so a rotary displacement transducer (sensor) can be attached to the linkage. The position of the throttle mechanism is usually in a very hostile environment such as the top of the engine or underneath air intake ducts, so either device must be extremely rugged and able to withstand high levels of shock, vibration and high temperatures.

Photo 2.5
Many calculator manufacturers - including Texas Instruments and Casio - make attachable data loggers. In this photo a TI-CBL2 computer based laboratory (data logger) is connected to a TI-83 graphing calculator. The CBL/CBR program shown on the displat provides the interface for this to work. A huge range of probes are available to connect to these devices.



Special 'paddles' on the driver's steering wheel electronically control the clutch actuating mechanism on today's high-performance racing cars, overcoming the need for the driver to use the feet to engage or disengage the clutch. This arrangement allows faster up-changing and down-changing of the gears during acceleration and braking.

When it comes to braking, recent developments in GT and Formula One brake caliper design have enabled systems to be fitted to monitor the wear of the brake pads and discs during a race. Advising the driver to back off by one second a lap can make a significant difference to brake wear. The movement of the brake caliper piston is sensed by a very small sensor embedded in the caliper body, which has been specially designed to withstand extremes of shock and vibration from the track as well as the high temperatures from the brake discs. The back of the brake pads can reach temperatures as high as 400°C, while the caliper body can reach 150–200°C. On Formula One cars up to eight sensors per car are fitted. The signals from the sensor are fed to the car's data-acquisition system and can tell race engineers the condition of the brake pad and disc wear characteristics.

Question: As a work experience student you have been asked to prepare a leaflet for some Year 8 students who will be visiting the racing car development laboratories of the Ford Motor Company. In 200 words what would you say?

Questions

35 A car moving at 30 m s^{-1} decelerates at a uniform rate of 1.5 m s^{-2}. How many seconds will it take to stop and how far will it travel in this time?

36 Analysis of traffic camera data shows that a car 4 m long takes 1.2 seconds to cross an intersection 16 m wide. The time taken is from the moment the car's headlights enter the intersection to the moment the tail-lights depart. Was the car exceeding the speed limit of 60 km h^{-1}?

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***37** To help you rearrange equations and substitute numbers, do these simple calculations (Table 2.13):

Table 2.13

DISPLACEMENT		TIME	VELOCITY
(a)	300 m	6 s	
(b)	150 km	4 h 30 min	
(c)		30 s	340 m/s
(d)		3 h 15 min	220 km/h
(e)	300 m		15 m/s
(f)	3.5×10^6 km		65 km/h

***38** The best time by an Australian in the 40 km marathon is that of Robert de Castella, who ran the Boston Marathon in 1986 in 2 h 7 min 51 s. Calculate: **(a)** his average speed; **(b)** the time it would take Michael Johnson if he ran the distance at 10.15 m s^{-1}.

***39** The fastest lap of the British Motorcycle Grand Prix at Donnington Park was in 1993 by Luca Cadalora on a 500 cc Yamaha in 1 min 34.716 s, averaging 152.908 km/h. **(a)** How long is the track? **(b)** If the race was 30 laps and he took 47 min 45.630 s, what was his average speed for the race?

***40** Australian fast bowler Brett Lee was electronically timed to deliver a cricket ball at 157.4 km h^{-1} in the second test against South Africa in 2002. How many seconds would it take for the ball to travel the 20 m length of the cricket pitch?

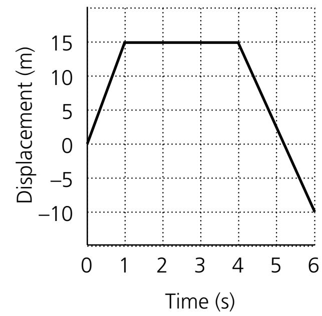
***41** A person runs in a straight line 84 m south in 9.0 s and then 160 m north in 18.0 s. What is his **(a)** displacement; **(b)** average speed; **(c)** average velocity?

***42** A car travels on a straight road for 50 km at 30 km h^{-1}. It then continues in the same direction for another 20 km at 60 km/h. What is the average velocity of the car during this trip?

***43** The graph in Figure 2.28 shows the displacement of a radio-controlled car being driven in a straight line:

(a) What is its displacement after 3 s?
(b) Calculate how far it travelled in the 6 s.
(c) At what stage was its velocity the greatest?
(d) When was it stationary?
(e) When was its velocity constant but not zero?
(f) What was its velocity at 5 s?

Figure 2.28
For question 43.



***44** Practise applying the acceleration formula by completing Table 2.14. Do not write in this book.

Table 2.14

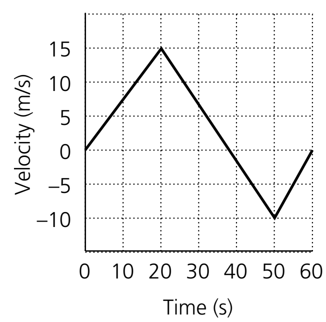
	v (m s^{-1})	u (m s^{-1})	t (s)	a (m s^{-2})
(a)	100	40	3.5	?
(b)	60	130	0.85	?
(c)	250	?	1.5	4.0
(d)	16.7	27.0	?	−1.5
(e)	0.55	0.15	?	60
(f)	?	0	1.2×10^{-3}	2.0

***45** A car with an initial velocity of 3.0 m s^{-1} has a velocity of 34 m s^{-1} after 3.0 s. Calculate **(a)** its acceleration; **(b)** its average velocity; **(c)** how far it moved in its third second of motion; **(d)** its speed after travelling 20 m.

***46** The graph in Figure 2.29 shows the motion of a girl on rollerblades as a function of time.

(a) Calculate her displacement after 50 seconds.
(b) Calculate the distance she travelled in the minute.
(c) At what stage was her acceleration the greatest?
(d) When was she stationary?
(e) When was her velocity constant but not zero?

Figure 2.29
For question 46.



***47** Table 2.15 will give you practice in selecting equations of motion and substituting values into them. Complete the table but do not write in this book.

Table 2.15

QUESTION	s (m)	u (m s^{-1})	v (m s^{-1})	a (m s^{-2})	t (s)
(a)		0		3	1.5
(b)	200	0			1.4
(c)		20	65	2.6	
(d)	315	7.5		2.5	
(e)	400	25		− 0.4	
(f)		30	8.7		2.3
(g)	1550	80			800

****48** A cyclist is travelling at a constant 10 m s^{-1} when he begins to coast up a hill. Assuming that he decelerates uniformly at 1.8 m s^{-2}, calculate **(a)** how far he will travel before coming to rest; **(b)** how long this will take.

****49** A pedestrian steps on to the road while an approaching car is travelling at 30 km h^{-1}. If the driver's reaction time is 0.3 s and the braking deceleration is 4.5 m s^{-1}, calculate **(a)** the stopping distance; **(b)** the stopping time.

****50** A car travelling at 100 km h^{-1} takes 65 m to stop after the driver sees a child run on to the road chasing a ball. If the driver's reaction time is 0.25 s, calculate the deceleration of the car.

***51** In the 1993 British Motorcycle Grand Prix, Kevin Schwantes was eliminated after a crash. *Australian Motorcycle News* described the crash: 'Schwantes was the first to crash after asking too much of a cold rear tyre. He hit the grass at 290 km/h and slid to a halt in a set of sand traps 50 m down the track.' Calculate Schwantes' deceleration in this accident.

***52** The results of experiments published in 1966 show that nerve impulses can travel at 288 km/h in the human body. How many seconds would elapse if they travelled at this speed from your toe to your brain (say 170 cm)? Assume the speed is constant.

***53** The Lee Enfield Rifle (.303) was used by Commonwealth Forces during the Second World War. Its projectiles had a muzzle velocity of 745 m s^{-1} and came to rest at a range of 700 m. Calculate **(a)** the deceleration (assumed uniform); **(b)** the time of flight.

***54** Suppose a rocketship in deep space moves with a constant acceleration of 9.8 m s^{-2}, which will give the illusion of normal gravity during the flight. **(a)** If it starts from rest, what time will it take to reach a speed one-tenth that of the speed of light (3×10^8 m s^{-1})? **(b)** How far will it travel in doing so?

****55** Consider a case where air resistance is taken into account. A tennis ball was dropped from a 120 m high cliff and accelerated uniformly to a terminal speed of 20 m s^{-1} after 5 s. From then on to the ground it travelled at this speed. Calculate **(a)** its acceleration over the first 5 s; **(b)** how far it travelled before it reached terminal speed; **(c)** its total time of flight; **(d)** its impact velocity; **(e)** its average velocity for the entire flight.

****56** Consider cases where an object is thrown vertically into the air. In these cases upward is still the positive direction. 'Time of flight' means the total time from launch to impact. Complete Table 2.16:

Table 2.16

	INITIAL VELOCITY u (m/s)	TIME OF FLIGHT t (s)	MAXIMUM HEIGHT s (m)
(a)	10		
(b)			100
(c)		5.5	

****57** The single cable supporting a construction elevator breaks when the elevator passes the sixth floor (25 m) on its way up while at a speed of 3.0 m s^{-1}. **(a)** Calculate velocity on impact. **(b)** How much time will elapse before the elevator strikes the ground?

****58** In an experiment to investigate the relationship between time, displacement, velocity and acceleration, a trolley was allowed to run down an inclined plane with its motion being recorded by a ticker timer. Figure 2.30 shows a 1 m length of the tape cut into five continuous segments so that it can be displayed in this textbook. Note the time interval between successive dots is 0.02 second.

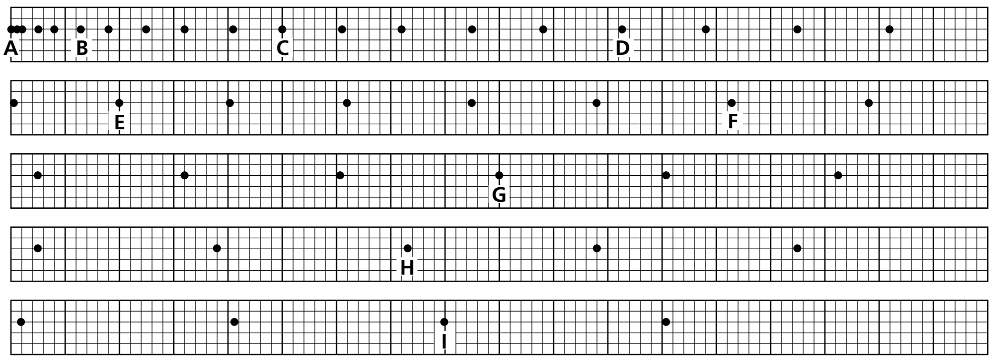


Figure 2.30
A ticker timer tape cut into five segments to fit the page.

Part A Displacement

(a) Every fifth dot was marked alphabetically. Check that this is correct.
(b) Calculate the time elapsed for each lettered dot and measure the displacement of each dot from the start. Add this to the data table (Table 2.17). Do not write in this book. Redraw the data tables or photocopy them.

Table 2.17

DOT	A	B	C	D	E	F	G	H
t (seconds)	0	0.1	0.2	0.3				
s (cm)								

(c) Plot a graph of *t* (*x*-axis) versus *s* (*y*-axis).
(d) What is the displacement of the last lettered dot (I)?

Part B Velocity

(e) Draw tangents at each of the lettered dots C, E and G and calculate their slope. Add to Table 2.18 in the second row ('slope').

Table 2.18

DOT	A	B	C	D	E	F	G	H
t (seconds)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Slope (cm/s)								
v (cm/s)	0							

(f) Calculate the velocity of each lettered dot by measuring the distance between dots either side of each lettered dot and dividing by the time interval (2×0.02 s). Add these data to Table 2.19. in the ***v*** row.
(g) How does the average velocity for each lettered dot compare with the instantaneous velocity as calculated from the slope of the *s*–*t* graph in Question **(c)**?
(h) For the average velocity as calculated in question **(g)** plot velocity vs time (*x*-axis) and draw a line of best fit.
(i) Determine the area under the graph up to the last lettered dot (I). How does this compare with the measured displacement of dot I?

Part C Acceleration

(j) Calculate acceleration by measuring the slope of the graph of ***v***–*t*.
(k) Calculate acceleration from the tape for dots C, E and G by subtracting the velocity five dots before from the velocity five dots after and dividing by the time interval over the ten dots. For example, to calculate the velocity at C, subtract the velocity at B from the velocity at D and divide by ten dot intervals of time. Add this to Table 2.19.

Table 2.19

DOT	A	B	C	D	E	F	G	H	I
t (seconds)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$\boldsymbol{v}_1$ (cm s^{-1})									
$\boldsymbol{v}_2$ (cm s^{-1})									
$\boldsymbol{a}$ (cm s^{-2})									

(l) Plot a graph of acceleration vs time and draw a line of best fit.
(m) Calculate the average acceleration by averaging the acceleration at the lettered dots C, E and G.
(n) How does the value of average acceleration compare with the slope of the ***v***–*t* graph?

Note: if you have access to a computer and spreadsheet, you may like to set up the spreadsheet to make the various calculations.

****59** Table 2.20 is taken from a *Wheels Magazine* comparison of some popular four-cylinder cars.

Table 2.20

	MAZDA 626	SUBARU LIBERTY	TOYOTA CAMRY
Engine capacity (litres)	1.991	2.212	2.164
Engine — Max. power (kW)	85	100	95
Top speed (km/h):			
• First gear	63	62	63
• Second gear	113	118	113
• Third gear	174	175	176
• Fourth gear	190	195	185
Acceleration (seconds):			
• 0–60 km/h	5.7	5.2	5.8
• 0–80 km/h	9.3	8.3	9.3
• 0–100 km/h	13.6	12.0	13.6
• 0–120 km/h	20.1	17.5	20.1
Standing 400 m (km/h):	19.1 (117)	18.3 (123)	19.1 (118)
• 40–70 km/h	4.1	3.5	3.9
• 60–90 km/h	5.5	4.9	5.5
• 80–100 km/h	7.2	6.0	7.1
• 100–130 km/h	10.6	9.0	11.0

(a) Which car has the best overall acceleration? Justify your choice.
(b) Which, if any, of the cars reaches its maximum speed in less than 400 m?
(c) Does a car's ability to accelerate get progressively less at higher speeds? Justify your answer.
(d) Calculate the distance over which the 40–70 km h^{-1} acceleration test would have occurred for the Toyota.
(e) List five other criteria that would be important to include in a car comparison.
(f) 'The greater the engine power, the greater the acceleration.' Comment critically on this claim with reference to the above data.
(g) 'The greater the engine capacity the greater the acceleration.' Comment critically.

Extension — complex, challenging and novel

*****60** Chris beats Sandy by 10 m in a 100 m sprint. Chris, wanting to give Sandy an equal chance, agrees to race her again but to begin 10 m behind the starting line. Does this really give Sandy an equal chance?

*****61** The General Dynamics F-111 jet has been in service with the RAAF since 1963. Its maximum speed above 50 000 feet is 825 m s^{-1} (Mach 2.5) but this drops to 396 m s^{-1} (Mach 1.2) at sea level because of air resistance. Calculate the deceleration as an F-111 drops and decreases speed as stated in 30 s.
Note: Mach numbers are the number of times the speed is greater than the speed of sound at that place (330 m s^{-1}).

*****62** Two bus stops are 1200 m apart. A bus accelerates at 0.95 m s^{-2} from rest through the first quarter of the distance and then travels at constant speed for the next two quarters and decelerates to rest over the final quarter.
(a) What was the maximum speed? **(b)** What was the total time taken for the journey? **(c)** Draw a $\boldsymbol{v}$–t graph of the journey.

*****63** A basketball player, standing near the basket to grab a rebound, jumps 76.0 cm vertically. On his way up, how much time does he spend **(a)** in the bottom 15 cm of his jump; **(b)** in the top 15 cm of his jump? Does this help to explain why such players seem to hang in the air at the tops of their jumps?

*****64** A juggler tosses balls vertically into the air. How much higher must they be tossed if they are to spend twice as much time in the air?

*****65** A stone is dropped off a bridge 50 m above the water. Exactly 1 s later another stone is thrown down and both stones strike the water together. **(a)** What must the initial speed of the second stone have been? **(b)** Plot a $\boldsymbol{v}$–t graph of both stones on the one graph.

*****66** Who would have the more thrilling ride: Kitty O'Neil in her dragster, which reached 628 km/h in 3.72 s or Eli Beeding who reached 116 km/h in 0.04 s on a rocket sled? Justify your choice by commenting on what determines how thrilling a ride might be — the speed, the time, the acceleration or something else.

*****67** A person standing on the edge of a cliff throws a ball straight up with speed '$\boldsymbol{u}$', allowing it to crash on to the rocks below. He later throws a ball with the same speed '$\boldsymbol{u}$' straight down. Which ball has the higher speed when it hits the rocks? Neglect air resistance.

*****68** A ball is dropped down an elevator shaft and then 1 s later a second ball is dropped. **(a)** How does the distance between the two balls vary as time passes? **(b)** How does the ratio $\boldsymbol{v}_1 : \boldsymbol{v}_2$ vary with time?

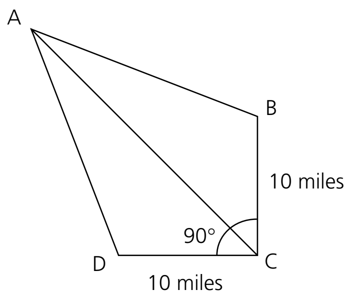
*****69** The Australia vs USA Nitro-Harley Challenge is the world's richest motorcycle drag race meeting. One of the most successful riders, Phil Hill (USA), is 61 years old. With a 103 cubic inch nitromethane injected 350 horsepower engine he can cover the standing quarter mile (400 m) in 7.22 seconds with a final speed of 305 km h^{-1}. The acceleration required to cover 400 m from a standing start in 7.22 s is more than the acceleration needed to reach 305 km h^{-1} from a standing start in the same time. How can you explain this apparent discrepancy in the calculations?

*****70** A rule-of-thumb in motorcycle drag racing is that 'sixty pounds is three-tenths of a second'. This is meant to show how a rider's weight affects the time to cover 400 m from a standing start. Australian national record holder Bill Curry has a best time of 6.92 s. Calculate how much his average acceleration would be if he was 20 kg heavier.
Note: 1 kg equals 2.2 pounds.

***71 The speed of non-land-based vehicles such as ships and planes is usually measured in 'knots'. A knot is one nautical mile (6080 feet) per hour. To measure the speed of a ship, a line with knots at set intervals was attached to a log that was thrown overboard from the stern of a ship. As the log drifted away from the ship a sailor would count how many knots passed through his fingers while the sandglass emptied. Usually the 'log line' had knots every 100 feet and the sandglass emptied in 1 minute. **(a)** Prove that a speed of 30 knots equals 30 nautical miles per hour. **(b)** How many km h^{-1} is 30 knots if 1 foot equals 0.305 metres?

***72 If you have access to a computer, set up a spreadsheet to compute the distance an object falls, as a function of time of falling, near the surface of the Earth ($\boldsymbol{g}$ = 9.8 m s^{-2}); our Moon ($\boldsymbol{g}$ = 1.6 m s^{-2}); Mars ($\boldsymbol{g}$ = 3.8 m s^{-2}) and the Sun ($\boldsymbol{g}$ = 270 m s^{-2}). Compute the distance of fall for each fifth of a second from 0 to 2 seconds.

Figure 2.31
The 37-year puzzle.



***73 The Incredible Tale of the 37-year Puzzle. This puzzle remained unsolved for 37 years until *Popular Science Magazine* published it again in October 1976. Two thousand responses were sent in and five different solutions appeared. The problem (Figure 2.31): A man always drives at the same speed. He makes it from A direct to C in 30 minutes; from A through B to C in 35 minutes; and from A through D to C in 40 minutes. How fast does he drive?

***74 Galileo's first attempt at producing a law of falling bodies was limited by his lack of mathematical means of describing continuously varying motion. In a letter to a friend Paolo Scarpi in 1604 he wrote: 'Spaces traversed in natural motion are in squared proportion of the times, and consequently the spaces traversed in equal times are as the odd numbers beginning with unity. And the principal in this, that the naturally moving body increases its velocity in the proportion that it is distant from the origin of the motion.' Can you convert this to mathematical statements and then comment on whether Galileo was correct with these early theories?

***75 The Sukhoi Su-29 is a Russian built two-seat aerobatic competition aircraft becoming popular in Australian competitions. If one was flying at its cruising speed of 160 knots (298 km/h) and an altitude of 1000 m and suddenly encountered terrain sloping upward at 4.3°, an amount difficult to detect, how much time would the pilot have to make a correction if he is to avoid flying into the ground?

***76 An article in the newspaper quoting a safety expert said that: 'An unrestrained child in a 50 km/h car crash suffered the same effects as being dropped on to concrete from a building's second floor. It said some parents still held the belief that merely placing children in the back seat would protect them in a crash.' Confirm or refute these comments made by the paper, making whatever approximations are required.

***77 A car has an oil leak from the sump and a drop falls every 2 seconds. Draw a diagram of how the spots would appear on a 64 m driveway as the car accelerates up it from rest at 2 m s^{-2}. Assume the first drop falls at the instant the car moves.

✔ TEST YOUR UNDERSTANDING

(Answer true or false)

- Two objects side by side must have the same speed.
- Acceleration is in the same direction as velocity.
- Velocity is a force.
- Heavier objects fall just a bit faster than light ones.
- If velocity is zero, acceleration is zero.
- In the absence of gravity all things move with equal ease.
- At the top of its flight a vertically thrown object has zero acceleration.

CHAPTER 03

Vectors and Graphing

3.1 VECTORS AND SCALARS

To completely specify some physical quantity it is not sufficient to just state its magnitude. In the previous chapter you saw that to describe the motion of an object you needed to state its speed and the direction it was heading. In other words, its velocity. Displacement and acceleration needed a direction too. They were all called **vector** quantities. The word 'vector' comes from the Latin *vectus* meaning 'to carry' — a word implying direction. In biology a vector is an organism that carries disease from one place to another; for example, a mosquito is the vector for malaria. In physics it means a quantity that needs both magnitude and direction to specify it fully. This chapter continues the discussion about the nature of vectors as used in physics. Later in the chapter there is a discussion on graphing and how graphs can be used to solve problems.

Some of the questions that puzzle students about vectors and graphs are:

- Can two vectors having different magnitudes be combined to give a zero result? Can three?
- Can a vector have zero magnitude if one of its components is not zero?
- If time has magnitude and a direction (past → present → future), is it a vector?
- How can a statistician look at an unemployment graph and say the unemployment rate is increasing whereas another statistician can say the rate is decreasing?

Table 3.1 Some scalar and vector quantities

SCALAR	VECTOR
Length	displacement
Speed	velocity
Time	acceleration
Volume	force
Mass	weight
Energy	momentum
Frequency	torque
Pressure	moment
Power	electric current
Temperature	electric field
Charge	magnetic flux density

Photo 3.1
XYZ Plotter. A computer controls the X, Y and Z coordinates of the cutting head in a tool-maker's workshop. A steel mould is cut and used to make plastic parts by injection moulding.

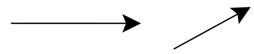
Representation of vectors

A vector quantity can be represented by an arrow called a vector. The length of the arrow represents the magnitude of the vector quantity, and the direction of the arrow shows the

Ch03-Walding 4th 25/8/04 10:23 AM Page 57

direction of the vector quantity. For example, Figure 3.1 shows two vectors representing the vector quantities velocity and force:

Figure 3.1



$\boldsymbol{v}$ = 5 m/s E F = 8 N, E30°N

NOVEL CHALLENGE

A ranger at Mt Mungo National Park published an booklet entitled Twenty Family Walks. In the introduction he wrote, 'The walks are short, ranging from a kilometre and a half to five kilometres; the average is two and a half kilometres.'

A What is the total length of all twenty walks?

B What is the greatest possible number of walks more than 4 km long?

C If there are three walks of 5 km each, what is the greatest possible number of walks shorter than 5 km but longer than $2\frac{1}{2}$ km?

Note: **wind directions** are confusing. A wind direction is where the wind is coming *from*. For instance, a south-easterly breeze comes *from* the south-east (S45°E or E45°S) but is *heading* north-east. Be careful to draw your diagrams carefully when wind directions are mentioned.

Scalar quantities require no statement about direction. For example, time = 3.5 s, mass = 25.5 g and current = 2.0 A are scalar quantity measurements — no direction has to be specified. The word 'scalar' comes from the Latin *scalaris* meaning 'pertaining to a ladder'. This refers to the stepwise change in the size of something without any reference to direction. *Note:* in maths class you may be taught how to work with 'unit vectors' using the symbols **i**, **j** and **k**. You could still use this system in physics if you like but it won't be discussed further in this book.

SR Activity 3.1 Finding your way home

Photocopy a map of your local area (e.g. a street directory). Draw in the route you normally take to school and estimate the distance travelled. Draw a straight line from your home to the school and determine your displacement or distance 'as the crow flies'. Include the direction.

VECTOR MANIPULATION 3.2

There are many cases in the world around us where more than one vector quantity is involved. When this is the case, we need rules to perform some form of arithmetic. We apply normal rules to scalar quantities — rules for addition, subtraction, multiplication and division. In the world of vector arithmetic, these rules must also take into account direction of the vector quantities. If you go 3 m N and then 4 m E your displacement is not 7 m. You'll see why below.

In this book we will represent a vector by printing it in ***bold italics***. For example, vector *A* will be represented as ***A*** and vector *v* as ***v***. Some books and teachers may prefer to underline the vector with a tilde (~), e.g. $\underset{\sim}{A}$, $\underset{\sim}{a}$, $\underset{\sim}{v}$ instead of using bold.

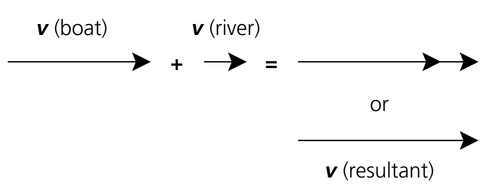
Vector addition

Two or more vector quantities can be combined to produce a single **resultant** vector.

Case 1

Consider rowing a boat at 5 m s^{-1} E in water that is also moving E at 1 m s^{-1}. Your actual velocity is 6 m s^{-1} E and is found by placing the two vector arrows head-to-tail. The resultant is a line drawn from the tail of the first arrow to the head of the second arrow (Figure 3.2).

Figure 3.2



Note: when *adding* vectors they should be placed *head-to-tail* and the resultant will always start at the tail of the first arrow and end at the head of the second arrow.

Case 2

Consider the same boat being rowed *against* the current. In this case the velocity of the river is 1 m s^{-1} W and is in the direction opposite to that of the boat and hence will slow the boat down:

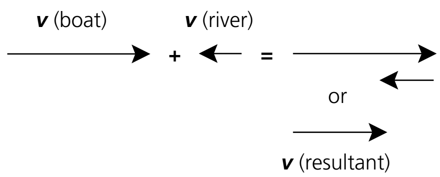


Figure 3.3

The resultant velocity is 4 m s^{-1} E. Note that when two vectors in the same line are added, the resultant has a direction the same as the larger vector.

Case 3: Vectors not in a line

Imagine you are rowing north at 3 m s^{-1} across a river but the river current is flowing east at 4 m s^{-1}. You would be dragged off-course by the current and your resultant velocity would be 5 m s^{-1} (Figure 3.4). Note again that the two vectors are added head-to-tail. The resultant is a line drawn from the tail of the first arrow to the head of the second arrow. This resultant can also be drawn as the diagonal of the parallelogram constructed by using the two given vectors as sides. Either method is acceptable.

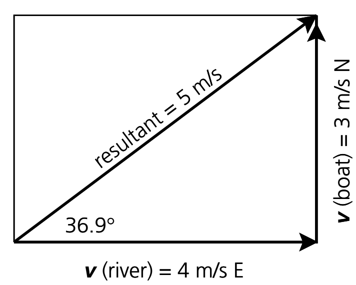
The solution to this problem is in two parts — a magnitude component (5 m s^{-1}) and a direction component (E36.8°N). This is achieved in the following manner:

1 *Magnitude* Because the starting vectors for the boat and river are at right angles (N and E), Pythagoras's theorem can be used. Resultant = $\sqrt{4^2 + 3^2}$. If a scale diagram was used, the resultant could be measured with a ruler.

2 *Direction* Because the two vectors and the resultant form a right-angled triangle, trigonometry can be used: i.e.

$$\tan\theta = \frac{\text{opposite side}}{\text{adjacent side}} = \frac{3}{4} = 0.75\text{, hence } \theta = 36.9°$$

Figure 3.4

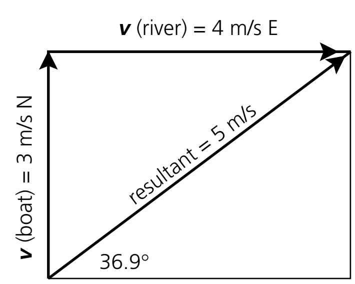


Note that the order of addition is not important. Figure 3.4 could also be drawn as shown in Figure 3.5. The resultant is still the same.

Refresher The trigonometric ratios for the right-angled triangle shown in Figure 3.6 are given below.

$$\sin\theta = \frac{\text{opposite side length}}{\text{hypotenuse}}\text{; } \cos\theta = \frac{\text{adjacent side length}}{\text{hypotenuse}}\text{; } \tan\theta = \frac{\text{opposite side length}}{\text{adjacent}}$$

Figure 3.5



Hint: when the calculator displays 0.75, usually you need to press 'shift' then either sin, cos or tan to convert this to the angle. *Note:* Latin *sinus* = 'curve'.

Questions

1 Calculate the values of θ in the following right-angled triangles (do not write in this book):

Figure 3.6

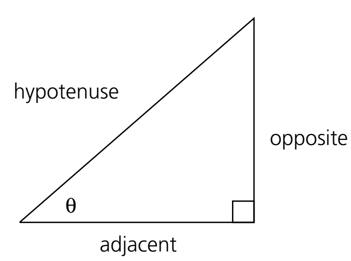


Table 3.2

	ADJACENT	OPPOSITE	HYPOTENUSE	RATIO	θ
(a)	10	7		tan θ =	
(b)	8		13	cos θ =	
(c)		9	20	sin θ =	
(d)	200	50		tan θ =	

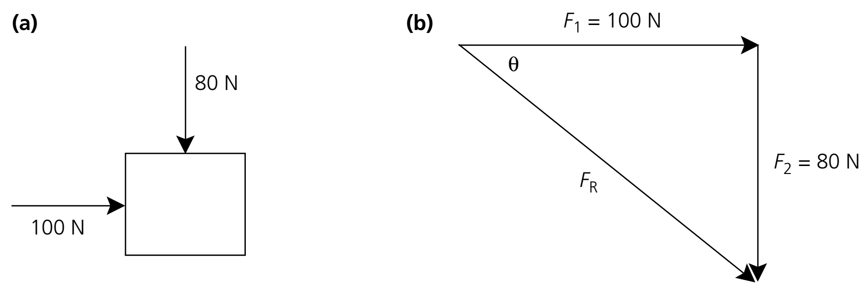
Note that the ratios for sin and cos are always 1.0 or less; only tan can go beyond ±1.0. If you disagree with the values of θ shown in the back of this book, check that your calculator is in degrees (shown by a small DEG in the display). A common mistake occurs when the calculator is put into radians (RAD). Ask someone near you or your teacher if you get stuck.

Example 1

Two forces act on a crate as shown in Figure 3.7**(a)**. Calculate the resultant force.

Solution (See Figure 3.7**(b)**)
Resultant ($\boldsymbol{F}_R$) = $\sqrt{100^2 + 80^2}$ = 128 N. The angle θ is found by $\tan\theta = \frac{80}{100}$, so θ = 38.7°.

Figure 3.7



Case 4: Vectors not at right angles

In such cases, Pythagoras's theorem and the three trigonometric formulas do not apply. However, there are several other solutions you may use. Check with your teacher whether you are expected to use these additional formulas:

Method 1 If you have the value of one angle and its opposite side use the sine rule:

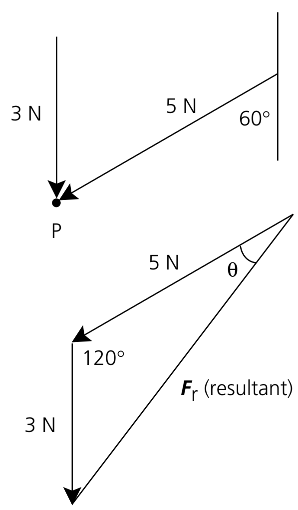
$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

Method 2 If the sine rule won't work and you have two sides and one angle you may be able to use the cosine rule if you are familiar with it:

$$c^2 = a^2 + b^2 - (2ab\cos C)$$

Method 3 By dropping a perpendicular from one apex to the opposite side. This produces two right-angled triangles, which may possibly be solved using Pythagoras's theorem or trigonometry.

Figure 3.8



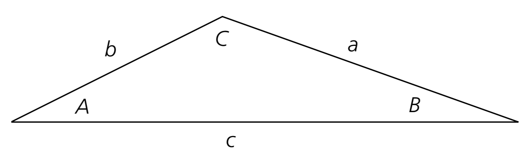
Example 2

A force of 3.0 N south and 5.0 N S60°W act on the same point (Figure 3.8). Determine the total force that must be acting on that point.

Solution (See Figure 3.9)
The cosine rule must be used to determine the magnitude of the resultant force:

$$\begin{aligned} \boldsymbol{F}_r^2 &= 3^2 + 5^2 - (2 \times 3 \times 5 \times \cos 120°) \\ \boldsymbol{F}_r &= \sqrt{49} \\ \boldsymbol{F}_r &= 7\ \text{N} \end{aligned}$$

Figure 3.9



Use the sine rule to determine the direction:

$$\frac{\sin\theta}{3} = \frac{\sin 120°}{7}$$

$$\therefore \sin\theta = \frac{3}{7} \times \sin 120°$$

$$\theta = 22°$$

Total force = 7 N, S38.2°W.

Questions

2 Find the magnitude and direction of the resultant vector obtained by adding **(a)** displacements of 30 m E and 20 m N; **(b)** velocities of 16 m s^{-1} W and 30 m s^{-1} S; **(c)** forces of 20 N NW and 10 N NE.

3 Calculate the resultant force when the forces listed below act together on a wooden log: **(a)** 18 N horizontal and 24 N vertical; **(b)** $2.5 \times 10{-}3$ N east and $1.8 \times 10{-}3$ N north; **(c)** 300 N horizontal (pulling) and 150 N at an angle of 25° to the horizontal (pulling).

4 A red cricket ball of mass 150 g is thrown upwards with a velocity of 20 m s^{-1}. Gravitational acceleration allows the ball to reach a maximum height of 20 m in a time of 2.0 s, after which it falls to the ground and strikes it 2.0 s later. From this statement, name three scalar quantities and three vector quantities.

5 A force of 18.0 N south and 14.0 N S50°W acts on the same point P. Determine the total force acting on that point.

NOVEL CHALLENGE

Is it best to run or walk in the rain if you want to stay as dry as possible? Imagine you had to get from one building to another and the rain was falling vertically (for simplicity). *What are the benefits and disadvantages of running versus walking? So what should you do?*

Vector subtraction

If your mass was 65 kg and after the Christmas holidays you had gained 5 kg, your mass would of course be 70 kg. You could say that your **change in mass** was +5 kg. If, instead, you dieted over Christmas and lost 5 kg your final mass would be 60 kg. Your change in mass would be –5 kg. You could have worked this out in Grade 8. In physics we need to be very particular in the way we talk about 'change' in a measurement, particularly vectors. In physics 'change' means difference, but with this convention:

Change in a measurement (Δ) = final measurement – initial measurement

After dieting: change in mass (Δm) = final mass – initial mass
= 60 kg – 65 kg
= –5 kg

This is simple for scalar quantities like mass, temperature and bank balances. But in physics it is also necessary to subtract vectors. In maths, you should have learnt that subtraction is the same as adding a negative, that is, 10 – 8 is equivalent to 10 + –8 and the answer is +2 either way. When subtracting vector ***B*** from vector ***A***, the direction of vector ***B*** is changed to its opposite and then added to vector ***A*** (head to tail).

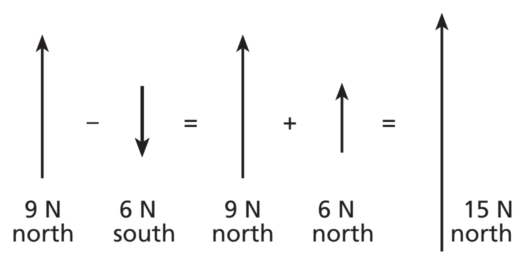


Figure 3.10
Subtraction of vectors.

Ch03-Walding 4th 25/8/04 10:23 AM Page 61

Example 1

A ball falling vertically strikes the ground at a velocity of 10 m s^{-1} (downward) and rebounds vertically upward at a velocity of 8 m s^{-1}. Calculate the change in velocity. Assume the downward direction is negative.

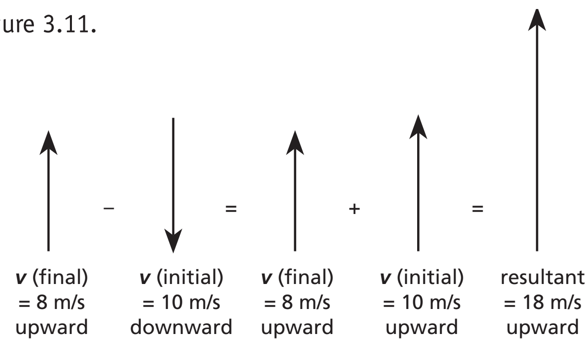
Solution

'Change' means difference, hence:

$$\text{Change in velocity } (\Delta \boldsymbol{v}) = \text{final velocity} - \text{initial velocity}$$
$$\text{Final velocity} = +8 \text{ m s}^{-1}$$
$$\text{Initial velocity} = -10 \text{ m s}^{-1}$$
$$\text{Change in velocity} = +8 \text{ m s}^{-1} - (-10 \text{ m s}^{-1}) = +18 \text{ m s}^{-1}$$
(the positive means upward).

This is shown in Figure 3.11.

Figure 3.11

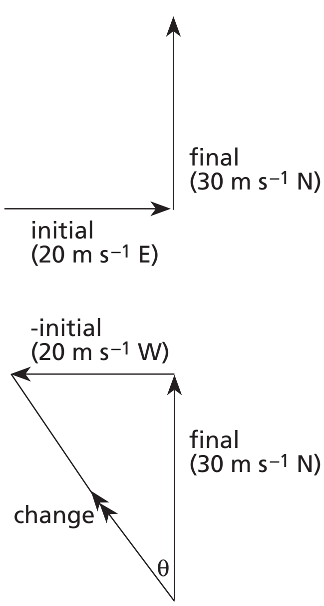


Example 2

A car travelling east at 20 m s^{-1} turns and accelerates to 30 m s^{-1} north (Figure 3.12). Calculate the change in velocity.

Solution

Figure 3.12



$$\text{Magnitude of change in velocity} = \sqrt{20^2 + 30^2}$$
$$= 36.1 \text{ m s}^{-1}$$
$$\text{Direction of change in velocity} = \tan^{-1}\frac{20}{30} = 33.7°$$
The change in velocity is 36.1 m s^{-1} in a direction N33.7°W

Questions

6 Calculate the change in velocity for each of these cases:

	Initial velocity	*Final velocity*
(a)	20 m s^{-1} south	30 m s^{-1} north
(b)	50 m s^{-1} west	10 m s^{-1} east
(c)	25 m s^{-1} north	35 m s^{-1} east
(d)	50 m s^{-1} south	20 m s^{-1} west

7 Determine the change in velocity of:

(a) a basketball with an initial velocity of 18 m s^{-1} down and a final velocity of 10 m s^{-1} up;

(b) a cricket ball travelling south at 30 m s^{-1} that struck a bat and was deflected to square leg (90°E of original path) at 25 m s^{-1};

(c) a bus travelling north at 20 km h^{-1}, which makes a 90° turn to the right without changing its speed.

8 A cricket ball delivered at 40 m s^{-1} strikes the pitch at 30° to the surface and bounces off the pitch at the same angle and speed. Calculate the change in velocity of the ball.

3.3 RESOLVING VECTORS INTO COMPONENTS

So far we have seen how two vectors can be added together to give a resultant third vector. The reverse process is called **resolution** (Latin *re* = 'back').

Why bother? In many cases it is convenient to 'break up' a vector into two **components** at right angles, for example vertically and horizontally. It is then sometimes easier to apply the laws of physics to the components.

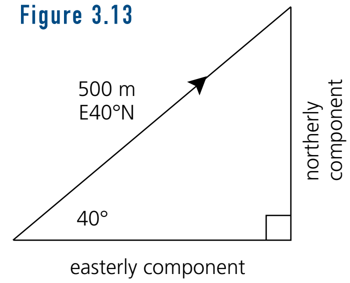
Imagine that a person walked 500 m in a direction 40° to the north of east. This could be resolved into a northerly component and an easterly component:

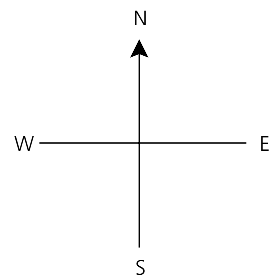
$$\sin 40° = \frac{\text{northerly component}}{500 \text{ m}}$$

$$\text{Northerly component} = \sin 40° \times 500 \text{ m} = 321 \text{ m}$$

$$\text{Easterly component} = \cos 40° \times 500 \text{ m} = 383 \text{ m}$$

Figure 3.13





Example

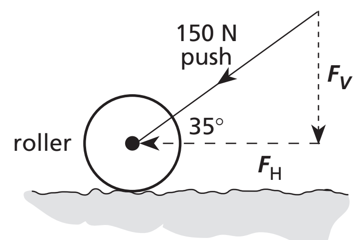
A roller is pushed along the ground with the handle at an angle of 35° to the horizontal. (See Figure 3.14.) Calculate **(a)** the horizontal component of the force pushing the roller over the ground; **(b)** the vertical component of the force pushing the roller into the ground. The push on the handle is 150 N.

Solution

$$\sin 35° = \frac{F_V}{150 \text{ N}} \text{ or } F_V = 150 \sin 35° = 150 \times 0.57 = 86 \text{ N}$$

$$\cos 35° = \frac{F_H}{150 \text{ N}} \text{ or } F_H = 150 \cos 35° = 150 \times 0.82 = 123 \text{ N}$$

Figure 3.14



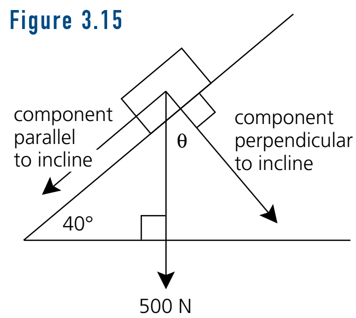
Questions

9 A girl pushes a shopping trolley along a horizontal path with a force of 100 N on the handle. If the angle between the handle and the ground is 30°, calculate the horizontal and vertical components of the pushing force.

10 A box of paper is being dragged along a vinyl floor by means of a rope angled at 20° to the floor. If an 80 N force is applied to the rope, calculate **(a)** the component of the force moving the box along the floor; **(b)** the component of the force tending to lift the box off the floor.

11 A 50 kg crate of rotting tomatoes rests on a 40° incline (Figure 3.15). If the force of gravity acting on the crate is 500 N vertically down toward the ground, calculate the components of the force **(a)** down the incline; **(b)** perpendicular to the incline. *Hint:* work out the value of θ first.

Figure 3.15



3.4 RELATIVE MOTION

When several motions are combined into one, some intriguing questions arise:

- In cricket, why does the bowler run to the wicket before delivering the ball?
- Why do you go slower when you row a boat upstream compared with rowing downstream?
- Why are jet planes launched off aircraft carriers into the wind?
- Why is it more dangerous to get out of a moving car than a stationary one?
- Why is a head-on collision worse than one with another car travelling in the same direction?

NOVEL CHALLENGE

A passenger on a train travelling at 60 km/h observes that it requires 4 s for another train 100 m long to pass her. *What is the speed of the second train?*

- When a person drives past you in a car you say they are moving. Couldn't they say you are moving and they are stationary? Who is right? Is this what physicists call **relative motion**?

The answers are very obvious, even without an understanding of vectors. But use of vectors can help us make predictions about the likely outcomes of such incidents.

Consider the first question about the cricketer. Imagine he could deliver a ball (B) from a standing position at 25 m s^{-1} relative to himself ($\boldsymbol{v}_{BC}$ = 25 m s^{-1}). If the cricketer runs at 5 m s^{-1} relative to the ground ($\boldsymbol{v}_{CG}$ = 5 m s^{-1}) while delivering it, then the ball would travel up the pitch at 30 m s^{-1}. This is a typical fast bowler's delivery speed. These motions can be expressed in equation form:

Velocity of ball relative to ground = velocity of ball relative to cricketer + velocity of cricketer relative to ground.

$$\boldsymbol{v}_{BG} = \boldsymbol{v}_{BC} + \boldsymbol{v}_{CG}$$
$$30 = 25 + 5$$

Note the order of symbols used in each case. When you write $\boldsymbol{v}_{BG}$, the first subscript (B) refers to a body in motion and the second letter (G) refers to whatever is being used for comparison. The second letter indicates the **frame of reference**, in this case the **G**round.

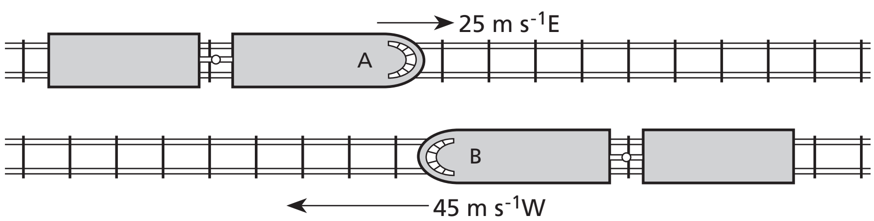
Note also that in the equation above the two inside subscripts on the right are the same letter (C). If this convention is always used then the two outside subscripts (B/G) are in the order that they should appear on the left.

$$\boldsymbol{v}_{BG} = \underbrace{\boldsymbol{v}_{BC} + \boldsymbol{v}_{CG}}_{\text{inside letter the same}}$$

Case 1: Parallel and toward each other

To illustrate relative motion further, consider two trains A and B moving toward each other on adjacent sets of tracks (Figure 3.16).

Figure 3.16



If train A is moving at 25 m s^{-1} to the right (relative to the ground) and train B is moving at 45 m s^{-1} to the left (relative to the ground) we can write their velocities as:

$\boldsymbol{v}_{AG}$ = +25 m s^{-1} $\quad$ $\boldsymbol{v}_{BG}$ = −45 m s^{-1} $\quad$ ('G' is the ground, the reference frame.)

The velocity of A relative to B is then:

$$\boldsymbol{v}_{AB} = \boldsymbol{v}_{AG} + \boldsymbol{v}_{GB} \quad \text{(using the conventions developed earlier)}$$

We don't have a value for $\boldsymbol{v}_{GB}$ but we do have a value for $\boldsymbol{v}_{BG}$. We can use the conversion:

$$\boldsymbol{v}_{GB} = -\boldsymbol{v}_{BG}$$

Hence

$$\begin{aligned}\boldsymbol{v}_{AB} &= \boldsymbol{v}_{AG} + \boldsymbol{v}_{GB}\\ &= \boldsymbol{v}_{AG} + -\boldsymbol{v}_{BG}\\ &= +25 \text{ m s}^{-1} + -(-45 \text{ m s}^{-1})\\ &= +70 \text{ m s}^{-1}\end{aligned}$$

Ch03-Walding 4th 25/8/04 10:23 AM Page 64

You can think of this as being the same as train B being stationary and train A coming towards it at 70 m s^{-1}.

Note: you may prefer to assume all velocities are relative to the earth or ground (G). In this case the G term is often omitted and the relationship becomes $\boldsymbol{v}_{AB} = \boldsymbol{v}_A - \boldsymbol{v}_B$.

Case 2: Parallel and moving in the same direction

The example of the cricketer belongs to this type. Consider another example: Car X is moving along a highway at 100 km h^{-1} relative to the ground and passes car Y, which is travelling at 85 km h^{-1} relative to the ground. You should be able to find the speed of X relative to Y.

$$\boldsymbol{v}_{XG} = +100 \text{ km h}^{-1} \qquad \boldsymbol{v}_{YG} = +85 \text{ km h}^{-1}$$
$$\boldsymbol{v}_{XY} = \boldsymbol{v}_{XG} + \boldsymbol{v}_{GY} = \boldsymbol{v}_{XG} + -\boldsymbol{v}_{YG} = +100 \text{ km h}^{-1} + -(+85 \text{ km h}^{-1}) = +15 \text{ km h}^{-1}$$

Questions

12 A girl is jogging at 4 m s^{-1} E and passes a boy who is jogging at 4 m s^{-1} W. What is the velocity of **(a)** the boy relative to the girl; **(b)** the girl relative to the boy; **(c)** the ground relative to the girl?

13 Why is it advantageous for planes to take-off into the wind?

14 A jet aircraft has an air speed of 720 km h^{-1} and is travelling the 4800 km west from Sydney to Perth. If the average wind speed is 40 km h^{-1} from the west, how long will it take the plane to reach Perth?

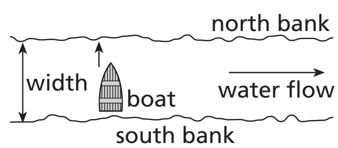
15 A boat that is capable of travelling at 4.5 km h^{-1} in still water is travelling on a river whose current is 1.5 km h^{-1} E. Find **(a)** the time it takes for the boat to travel 3.0 km downstream relative to the shore; **(b)** the time the boat takes to travel 3.0 km upstream against the current.

Case 3: Motion at right angles

The analysis of relative motion is fairly straightforward when the objects are moving parallel to each other. When other angles are involved, the situation becomes more complex.

Imagine a case where a person is rowing a boat *across* a river as shown in Figure 3.17. In this case the motion of the boat relative to the water is north and the motion of the water relative to the ground is west. The two motions are at right angles.

Figure 3.17



Example 1

A man can row a boat in still water at 30 m min^{-1}. He starts from the south shore of a river 600 m wide and aims due north. If the river is flowing west at 10 m min^{-1}, calculate:

(a) the velocity of the boat relative to the ground;
(b) the time taken to cross the river;
(c) the boat's landing position on the north shore.

Solution

(a) $\boldsymbol{v}_{BG} = \boldsymbol{v}_{BW} + \boldsymbol{v}_{WG}$

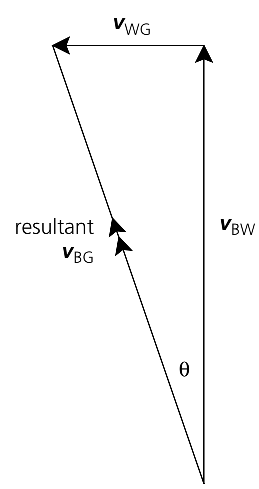
$= 30 \text{ m min}^{-1} \text{ N} + 10 \text{ m min}^{-1} \text{ W}$

$\boldsymbol{v}_{BG} = \sqrt{(\boldsymbol{v}_{BW})^2 + (\boldsymbol{v}_{WG})^2} = 32 \text{ m min}^{-1}$; $\theta = 18°$ (N18°W).

(b) The boat is moving at 30 m min^{-1} towards the opposite bank, which is 600 m away. It doesn't matter that the current is dragging the boat sideways at the same time. The crossing time is independent of the river current.

$$t = \frac{s_1}{v_1} = \frac{600 \text{ m north}}{30 \text{ m min}^{-1} \text{ north}} = 20 \text{ min}$$

Figure 3.18



(c) The boat is carried downstream by the current. If the boat journey took 20 minutes and the current was flowing at 10 m min^{-1}, then the distance moved downstream is:

$$\boldsymbol{s}_2 = \boldsymbol{v}_2 t = 10 \text{ m min}^{-1} \text{ west} \times 20 \text{ min} = 200 \text{ m west}$$

Example 2

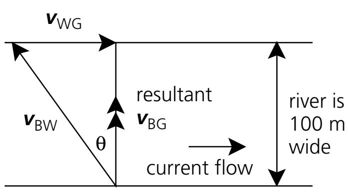
A person wishes to cross a 100 m wide river that is flowing east at 5 m s^{-1}. If they can row a boat in still water at 8 m s^{-1}, at what angle upstream should they head to end up on a point on the bank directly opposite? What is their crossing time?

Solution

In this case the resultant is not the hypotenuse of the vector triangle as is more common — it is the side adjacent to the angle θ.

Note that the convention about subscripts still holds for this diagram. The inside subscripts are the same (W):

Figure 3.19



$$\boldsymbol{v}_{BG} = \boldsymbol{v}_{BW} + \boldsymbol{v}_{WG}$$

$$\text{Using Pythagoras's theorem: } \boldsymbol{v}_{BG} = \sqrt{8^2 - 5^2} = 6.2 \text{ m s}^{-1}$$

$$\theta = \tan^{-1} 5/8 = 32°$$

$$\text{Crossing time: } t = \frac{s}{v} = \frac{100}{6.2} = 16 \text{ s}$$

Questions

16 A boat is driven with a velocity of 4.5 m s^{-1} relative to the water in a direction north straight across a river that is flowing at 2.0 m s^{-1} east.

(a) What is the boat's speed relative to the ground?

(b) If the river is 100 m wide, how far downstream will the boat reach the other side?

17 An airship is fitted with motors that can propel it at 6 km h^{-1} in still air. The captain wishes to travel 600 km due W but there is a wind blowing at 12 km h^{-1} from the east. Determine: **(a)** in what direction he must head; **(b)** what time his journey will take; **(c)** his ground speed.

If the wind was blowing at the same speed as before but from a direction of E30°S, determine: **(d)** in what direction he would now have to head; **(e)** what time his journey would now take; **(f)** his ground speed.

DRAWING GRAPHS 3.5

In the previous chapter you saw how graphing data was an important way of showing how motion varied as time passed. **Graphs** are a useful way of showing how one quantity depends on another.

On a graph the horizontal or *x*-axis is where the **independent variable** or cause is plotted. This also includes variables that progress regardless of the experiment. A good example is 'time elapsed', for time marches on whether any experiment is being carried out or not. The effect of that cause is plotted on the vertical or *y*-axis. This is called the **dependent variable**.

Linear relationships: direct proportion

Suppose, for example, that an experiment is performed to determine how much a certain rubber band stretches when masses are hung vertically from it. Data recorded from it are shown below:

mass (g)	0	20	40	60	80
stretch (mm)	0	9	21	30	42

The relationship becomes obvious when the points are plotted (Figure 3.20). Note that each point is plotted as a dot. You should put a circle around your dots. You should do this or use a cross rather than just a dot by itself because if the line joining the dots goes over the dots themselves they disappear. In this case a **line of best fit** is drawn. This has as many points on the line as possible. There are usually some points that aren't on the line and the line should be drawn so that there is an equal number below the line as above it. Scientists and engineers use a complex mathematical procedure (the method of least squares) to determine where the line of best fit should be. In some cases the line may not pass through any of the points. Any point that is a long way out of place is called an **outlier** and can be said to be **spurious**. It should be noted and the reasons for its existence be discussed but it should be left off the line.

If the line is straight, as in the graph in Figure 3.20, it takes the general form of:

$$y \propto x$$

The proportional sign ($\propto$) can be replaced by an equals sign and a constant (m).

$$y = mx \quad \text{or} \quad y = mx + c$$

where x and y are the **variables**, m the **gradient** or **slope** of the line and c the **intercept** or point where the line cuts the y-axis.

In the graph of mass versus stretch, the intercept c is zero. The slope is found by dividing the change in y value by the change in the x value for the same section of the line. This can be written as:

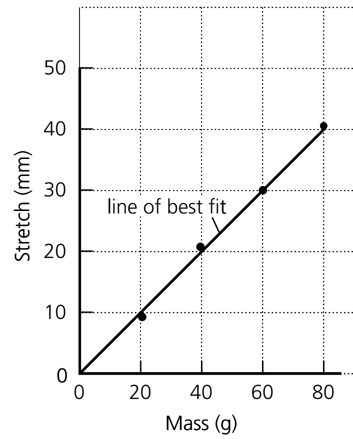
$$\text{Slope } (m) = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Students often find it easier to remember this as 'rise over run' where *rise* refers to the y-axis and *run* refers to the x-axis.

The slope of the graph is given by: $m = \dfrac{40 - 0}{80 - 0} = 0.5$ and the intercept c is zero. This means that for every 1 g change in mass (x-axis), the rubber band changes by 0.5 mm in length (y-axis). Be careful with significant figures!

A straight line graph is said to be **directly proportional**. If it also passes through the **origin** (0,0) it is also said to be **linear**.

Figure 3.20
Graph of linear relationship ($y \propto x$).



Example

The position of a car on a road is noted every 5 seconds and the following data obtained:

t = time elapsed (seconds)	0.0	5.0	10.0	15.0	20.0
s = displacement (metres)	16.0	23.5	34.0	41.0	50.0

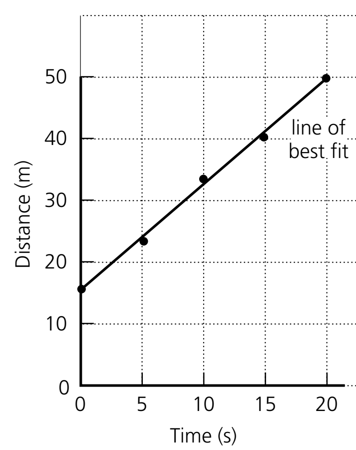
(a) Plot the data.
(b) Calculate the average velocity (slope).
(c) State the intercept.
(d) State the equation for the line.
(e) Predict the position at 25.0 s.
(f) State the position at 15.0 s.

Solution

(a) See Figure 3.21.

(b) Slope $= \dfrac{50.0 - 16.0}{20.0 - 0.0} = 1.7 \text{ m s}^{-1}$.

Figure 3.21
Graph of direct proportion ($y = mx + c$)



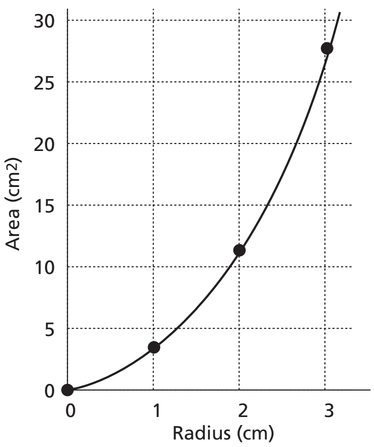
(c) Intercept = 16.0 m.
(d) $y = mx + c$, hence $s = 1.7t + 16$.
(e) The graph has to be extended to determine the position at 25.0 s. This is called **extrapolation**. Its value is approximately 58 m. Alternatively, the value $t = 25$ s could be substituted into the equation $s = 1.7t + 16$ to give the answer of 58.5 m.
(f) A value that is between two measured points is determined by **interpolation**. Its value is 37 m.

Non-linear relationships

The most common non-linear relationships you will meet in physics are:

- parabolic
- inverse
- inverse square
- exponential
- logarithmic.

Figure 3.22
Parabolic relationship ($y \propto x^2$).



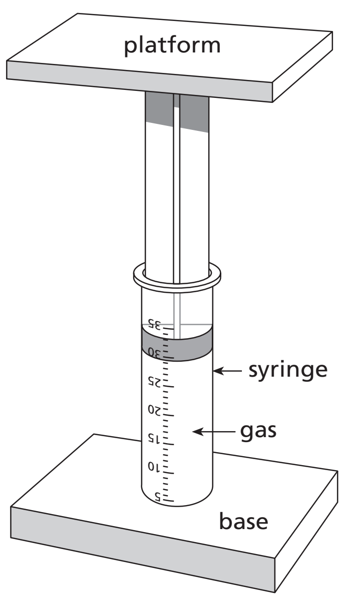
Parabolic relationships

Another direct relationship you will encounter is the parabolic relationship. The relationship between the area of a circle (A) and its radius (r) is a good example. The data below show these variables and Figure 3.22 shows a graph of the area as a function of the radius.

r = radius (cm)	0.0	1.0	2.0	3.0
A = area (cm^2)	0.0	3.1	12.6	28.3

Other phenomena which exhibit parabolic relationships are the paths of comets (except Halley's which is elliptical), curved mirrors in telescopes and projectiles (arrows in flight).

Figure 3.23
Boyle's law apparatus (see 'Inverse proportion').



Inverse proportion

Consider a case in which the volume of gas in a syringe is measured as the pressure on the syringe is increased (Figure 3.23):

P = pressure (kPa)	81	159	397	792
V = volume (mL)	10	5	2	1

When the data are plotted a curve like that shown in Figure 3.24 is obtained.
Note: 'data' is plural and should be followed by 'are'; a single data point is called a 'datum', which should be followed by 'is'. There is a tendency lately to use 'data' for both singular and multiple points. To be precise you should talk about 'these data' and not 'this data' but no one seems to care. It may not be important here but if your professional career involves writing technical reports, you may find your work is judged on simple grammar as much as anything else.

In the case above, the dependent variable (y-axis) decreases as the independent variable (x-axis) increases. In this case $y \propto \frac{1}{x}$ or $P \propto \frac{1}{V}$ and this is said to be **inversely proportional**. This relationship is commonly known as Boyle's law.

The proportional sign ($\propto$) can be replaced by an equal sign and a constant (k) so that the equation becomes $y = \frac{k}{x}$ or $P = \frac{k}{V}$. This implies that $P \times V$ is a constant. Examination of the above data will show that $P \times V$ is a constant and equals about 800. The values are slightly above and below 800 but these are within possible errors of an experiment.

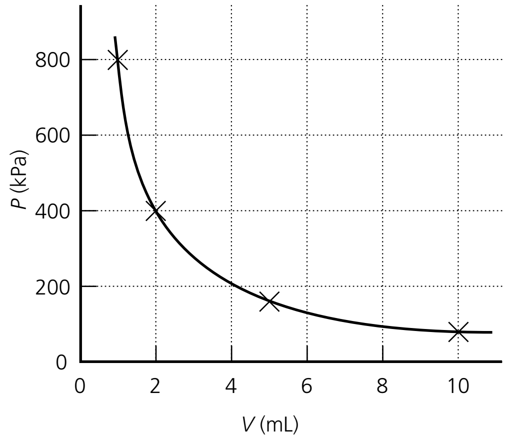


Figure 3.24
Graph of an inverse relationship ($y \propto 1/x$).

Other phenomena that exhibit parabolic relationships are the paths of comets (except Halley's, which is elliptical), curved mirrors in telescopes, and projectiles (arrows in flight).

Inverse square ($y = 1/x^2$)

This relationship looks similar to the inverse but has a much sharper bend. This type of relationship is very common in physics. For example, the variation in gravitational force with distance is given by $F \propto 1/d^2$. Other examples you will meet that vary in an inverse square relationship with distance are light intensity, centripetal force and electric field strength. Figure 3.25 shows how the magnetic force between two magnets varies with separation distance.

Exponential ($y = a^x$)

You will eventually meet some quantities in physics that are related exponentially. For example, the breakdown (decay) of a radioactive substance is given by: activity $\propto e^{-kt}$, where e and k are constants and t = time elapsed. (See Figure 3.26.)

Logarithmic ($y = \log_a x$)

A good example of this relationship is the response of the human ear (y) to sound energy (x). (See Figure 3.27.)

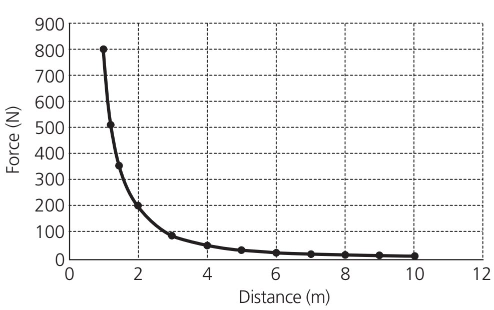


Figure 3.25
Graph of inverse square relationship ($y \propto 1/x^2$).

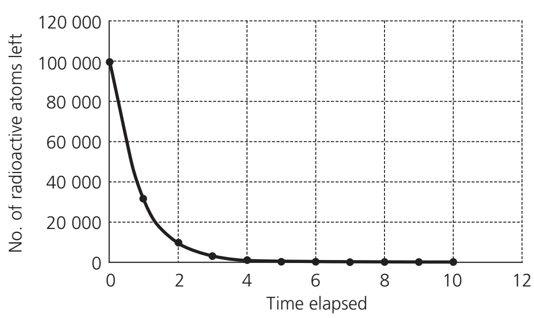


Figure 3.26
Graph of an exponential relationship ($y = a^x$).

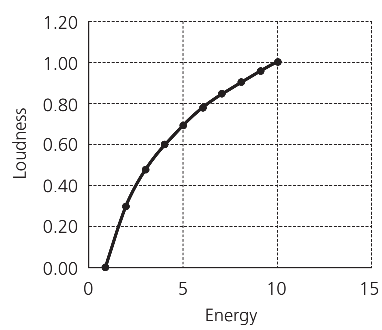


Figure 3.27
Graph of a logarithmic relationship ($y = \log_a x$).

Activity 3.2 MAKING A CUP OF TEA

1 If you have an electric jug at home, determine the time it takes to boil 2 cups (500 mL) of tap water.
2 Look underneath the jug to see what power rating the jug is. It should be somewhere in the range of 1000 watts to 2000 watts.
3 Make a summary of the results of other people in your class and plot a graph of boiling time (y-axis) against power rating.
4 What is the relationship between the two measurements — are they inverse?
5 How could this experiment be improved to collect more reliable data?

Proving a relationship

You have seen that if $y \propto x$, then a graph of y versus x is a straight line. Similarly, if $y \propto \frac{1}{x}$, then a graph of y versus $\frac{1}{x}$ will also be a straight line.

Example 1

For the pressure and volume data on the previous pages, plot a graph of P vs $\frac{1}{V}$ to demonstrate that it is a straight line.

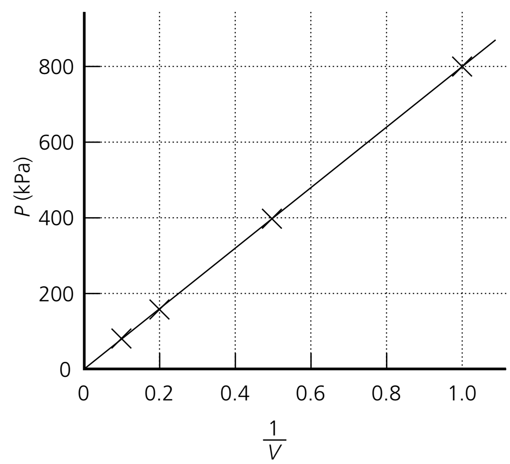
Solution

See Table 3.3 and Figure 3.28.

Table 3.3

P	81	159	397	792
V	10	5	2	1
$\frac{1}{V}$	0.1	0.2	0.5	1.0

Figure 3.28
Graph of P vs $\frac{1}{V}$ for a sample of gas.



Example 2

The displacement (s) in metres travelled by a car at various times t (in seconds) is shown below:

t	0	2	4	6	8
s	0	8	32	72	128

Draw a graph of **(a)** s vs t; **(b)** s vs t^2. What can you conclude?

Solution

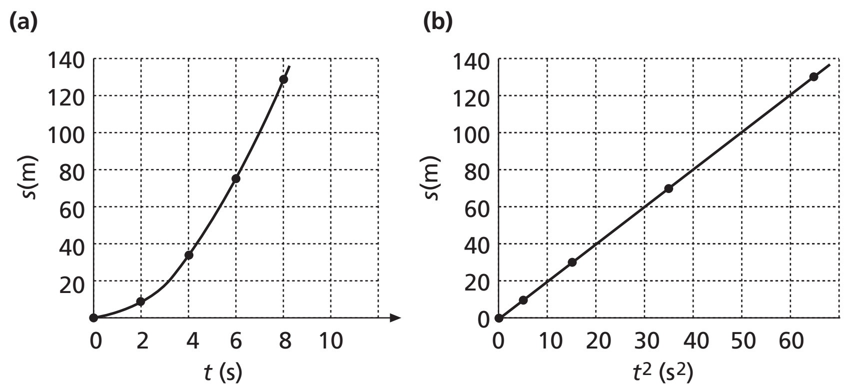


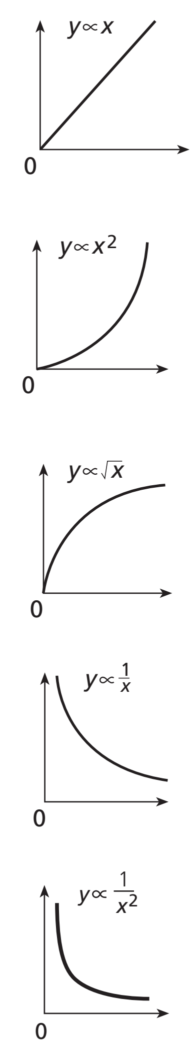
Figure 3.29

Conclude that s is proportional to t^2 as s vs t^2 is a straight line through the origin.

Working out relationships

The shapes of graphs often give a clue to the relationship between variables. Figure 3.30 shows five graphs and the relationship each suggests.

Figure 3.30 Five common relationships found in physics.



Activity 3.3 THE PLOT THICKENS

Part 1

If you have access to a computer with a spreadsheet you could construct a table that shows these relationships. You should fill the table with at least ten rows for each column and use a value of your choice for the gradient (m) — start with 2, and experiment.

For example, the formula for Cell D2 would be A2*A2 or A2^2.

Then use the plot, graph or chart function to see the shape of each relationship. You may need help to produce this spreadsheet.

	A	B	C	D	E	F
1	x-value	$y = mx$	$y = mx + c$	$y = x^2$	$y = 1/x$	$y = \sqrt{x}$
2	1					
3	2					
4	3					
5						

Figure 3.31 Computer spreadsheet.

You may even have access to a computer with graphing programs such as Sage or Omnigram.

Part 2

Try plotting sin, cos and tan of x for values of 0 to 360 degrees.

Part 3

What are the names of some computer applications that can be used to analyse data and come up with a mathematical relationship between variables? You may have to consult some computer magazines or professional journals such as *The Australian Physicist*, the *Journal of the Australian Institution of Engineers or Chemistry in Australia*.

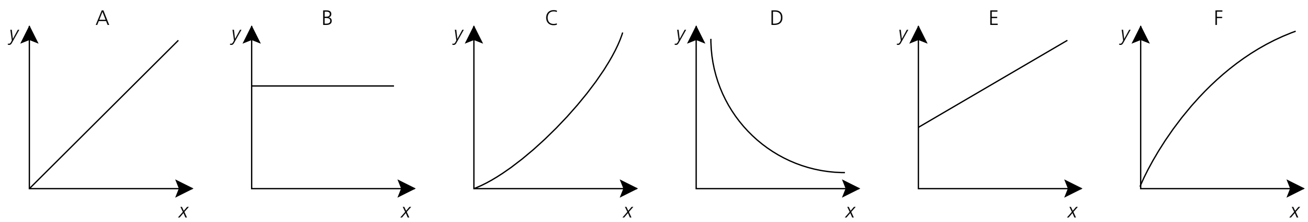
Ch03-Walding 4th 25/8/04 10:23 AM Page 71

Questions

18 If $W = kV$, then what is the effect on W of **(a)** tripling V; **(b)** halving V? What does a graph of W as a function of V look like?

19 For the graphs shown in Figure 3.32, select the graph that best represents: **(a)** y is proportional to x; **(b)** y is inversely proportional to x; **(c)** y is independent of x; **(d)** y is proportional to x^2.

Figure 3.32
For question 19.



20 Plot a graph of each of the sets of data given below. In each case draw the line of best fit. (*Note:* the independent variable is listed first.)

(a)

Diameter of circle (cm)	0.0	4.0	8.0	12.0
Circumference of circle (cm)	0.0	12.5	25.4	37.3

(b)

Time (years)	0.0	1.0	2.0	3.0	4.0
Height of tree (m)	0.0	0.32	0.66	1.00	1.30

(c)

Time (s)	0.0	2.0	4.0	6.0
Distance (m)	0.0	12	23	37

21 For each of the lines plotted in the previous question, **(a)** calculate the slope; **(b)** extrapolate to 14 cm, 5.0 years and 8.0 seconds respectively; **(c)** interpolate for 6.0 cm, 2.5 y and 3.0 s respectively.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***22** Find the magnitude and direction of the resultant vector obtained by adding **(a)** displacements of 4.0 m E to 6.0 m W; **(b)** velocities of 5.0 m s^{-1} E and 5.0 m s^{-1} S; **(c)** accelerations of 3.2 m s^{-2} N and 4.8 m s^{-2} W; **(d)** displacements of 5.0 m W and 5.0 m NW.

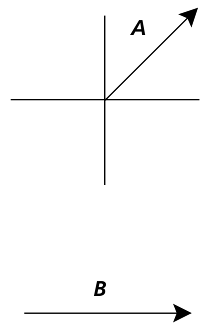
***23** Calculate the following: **(a)** Add 15 m N, 23 m W and 20 m S. **(b)** Add 20 m E, 15 m N and 25 m N30°E. **(c)** Add 5 km NE, 20 km S and 15 m E. **(d)** Calculate 20 m N minus 18 m N.

****24** What is the change in velocity when:
(a) a tennis ball travelling at 80 km h^{-1} is hit directly back at a speed of 95 km h^{-1};
(b) a car travelling N at 45 km h^{-1} turns west and travels at 60 km h^{-1};
(c) a cricket ball strikes the pitch at a speed of 30 m s^{-1} and an angle to the ground of 28° and bounces up at an angle of 35° and a speed of 25 m s^{-1};
(d) a 4.5 g bullet travelling west at 700 m s^{-1} strikes a piece of armour plate and is deflected by 40° off course with a 20% loss in speed?

***25** What are the N and E components of **(a)** 100 km North; **(b)** 50 m s^{-1} N30°E; **(c)** 25 newton N40°E?

****26** A man can row a boat in a northerly direction at 5 m s^{-1} (relative to the water) across a river 300 m wide. A current is flowing due east at 12 m s^{-1}.
(a) What is the velocity of the boat relative to the ground?
(b) What time would it take to cross the river?
(c) How far downstream would the man land on the opposite bank?

Figure 3.33
For question 27.



27** For Figure 3.33, show by means of a sketch how you would **(a)** add the two vector quantities ***A and ***B***; **(b)** subtract vector ***A*** from vector ***B***; **(c)** multiply ***B*** by a factor of 3.

****28** Plot a graph to show the relationship between heat (***H***) in joule, developed in a heater in 10 minutes by electric currents of I ampere:

Current (I)	0.5	0.8	1.0	1.6	2.4	3.0
Heat (H)	375	960	1500	3840	8640	13 500

Plot a further graph to find the relationship between H and I.

****29** The table below shows the height and mass of the world's tallest man, Robert Wadlow, from birth to death at age 22 years. Plot these data on one graph and answer the questions that follow.

Table 3.4 ROBERT WADLOW'S GROWTH CHART

AGE (YEARS)	HEIGHT (cm)	MASS (kg)
0	45	3.85
5	163	48
8	183	77
9	189	82
10	196	95
11	200	—
12	210	—
13	218	116
14	226	137
15	234	161
16	240	170
17	245	143*
18	253	195
19	258	218
20	261	220
21	265	223
22.4**	272	199

* Following influenza.
** Died 15 July 1940 from a septic blister on his ankle.
(Source: *The Guinness Book of Records*)

(a) Calculate his fastest height growth rate. Include the units.
(b) What was his fastest mass growth rate?
(c) What was the cause of the negative slope in his mass growth rate? State this rate numerically.
(d) What were the average growth rates for height and mass over his lifetime?
(e) What would his height have been at age 16.5 years?
(f) If he had lived to age 23, predict his height.
(g) Why doesn't this graph pass through the origin?

Extension —complex, challenging and novel

*****30** A student whirls a red-brown rubber stopper of mass 50 g on the end of a nylon string in a horizontal clockwise circle of diameter 1.2 m (as seen from above) at a constant speed of 8 m s^{-1}. From an instant when the stopper is moving in a northerly direction, find its change in velocity after moving round **(a)** one-half of a revolution; **(b)** one-quarter of a revolution; **(c)** one-tenth of a revolution.

*****31** A 50 kg crate of winter clothing is pulled along a horizontal polished vinyl floor by means of a rope making an angle of 30° with the floor. If the pull in the rope is 100 N, calculate **(a)** the effective component of the force pulling the crate along the floor; **(b)** the component tending to lift the crate off the floor.

***32 A 55 year old pilot wishes to fly a 15 t Lockheed SR-71 jet plane to a place 250 km due east in 30 minutes. Find his air speed and course if there is a southerly wind blowing at 50 km h^{-1}.

***33 Two solid ball bearings P and Q are made of the same vanadium alloy with a density of 11.5 g cm^{-3}. The diameter of P is four times the diameter of Q. Write the mathematical relationship between **(a)** the surface area of P and the surface area of Q; **(b)** the volume of P and the volume of Q; **(c)** the mass of P and the mass of Q; **(d)** the density of P and the density of Q.

***34 The volume of one plastic sphere is 35 times the volume of a second sphere. **(a)** Write an equation showing the relationship between the radii of sphere 1 and the radii of sphere 2. **(b)** If the radius of the first sphere is 50 cm, find the radius of the second.

***35 A wooden ramp of mass 50 kg rises vertically 3.0 m for every 5.0 m of its length. A crate of salmon of weight 1000 N is placed on the ramp 2.0 m from the lower end. Find the component of the weight **(a)** parallel to the ramp; **(b)** perpendicular to the ramp.

***36 Water in a river 1.6 km wide flows at a speed of 6.0 km h^{-1}. A captain attempts to cross the river in his ferry at right angles to the bank but by the time it has reached the opposite bank the captain awakes and notices that it is 1.0 km downstream. If the captain wishes to take his boat directly across, what angle upstream must he point the boat assuming the boat speed remains the same?

***37 A coil of wire, which has a resistance of R and an inductance of L, has an impedance Z given by the relationship: $Z = \sqrt{R^2 + 4\pi^2 L^2 f^2}$ where f is the frequency of the AC electric current flowing through the coil of wire.

In an experiment to determine Z at a variety of frequencies, f, a researcher recorded the following data:

Frequency (Hz)	0	100	200	300
Impedance (ohms)	100	121	162	210

By drawing the appropriate graph or other means, determine the values of R and L for this component.

***38 German physicist Arnd Leike, from the University of Munich, found that the decay of foam height in beer with time was exponential: $y \propto 1/x^n$, where y = height of the foam and x = time. He was awarded an 'Ig Nobel' Prize by the science humour magazine *Annals of Improbable Research* for one of the world's most useless pieces of research. Using an Excel spreadsheet or your graphing calculator, describe the difference between the graphs when $n = 3$ (Leike's result) and $n = 2$ (inverse square).

CHAPTER 04

Forces in Action

It seems incredible that a man could start railway wagons moving just by pulling on a rope with his teeth. Robert Galstyan of Armenia did just that in 1992: he set a world record by pulling two carriages a distance of 7 m with his teeth. Question: how could a 100 kg man accelerate 220 t of railway carriages from rest? Had he studied Newton's laws of motion?

4.1 SOME WRONG IDEAS ABOUT FORCES

People have often been baffled by other questions about forces:

- After you shoot an arrow does it keep going until the force runs out?
- If it takes a force to keep a thing moving, why doesn't the Moon crash into the Earth?
- Why do racing cars have 'spoilers' to increase wind resistance when really they want to go faster?
- Are there any forces acting on you if you're weightless?
- Cream seems more dense than milk so how come it floats on top of the milk?
- Cork is very lightweight — but could I lift a 1 metre diameter ball of it?
- Can rockets take off faster if they have a concrete launch pad?
- Which weighs more — a tonne of feathers or a tonne of lead?

Every one of these statements is based on a misconception about forces. Many of them go back 2000 years to Aristotle's idea that a moving thing had an internal source of 'impetus', which it was given when first thrown or moved. Such an idea acted as an obstacle to the understanding of motion for 1500 years and it still persists in students and others even today. Other *wrong* ideas are:

- If a body is not moving there is no force on it.
- The speed of an object depends on the amount of force on it.
- When the force stops, motion stops.

It's hard to convince people that these are wrong because they do sound 'right' — they seem to agree with what we see. But two cases should help to clear up misunderstandings.

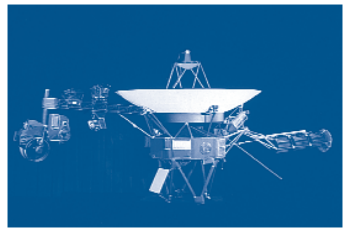
NOVEL CHALLENGE

The average mass of Sumo wrestlers in 1974 was 126 kg. In 2003 their average mass had risen to 157 kg. If this trend continues, when will they no longer be able to stand up? (The maximum mass that two legs can carry is 180 kg.)

Case 1: Space travel

Objects travelling in space keep going at constant velocity when there is no external force acting on them. The *Voyager* spacecraft left our solar system several years ago and is travelling on long after the jets ran out of fuel. On the other hand, a hockey ball rolls to a stop because frictional forces act on it and slow it down.

Photo 4.1
Voyager spacecraft.



Case 2: Ice skater

An ice skater will continue on at constant velocity until she tries to turn. The turning is a change in direction and hence a change in velocity. She will slow down unless she pushes off again.

Italian scientist **Galileo** (1564–1642) used the same logic to conclude that it is unbalanced forces that cause objects to slow down and stop. We call this force **friction**, a force that resists motion between two surfaces in contact. He took the word from the Latin *frictaire*

Ch04-Walding 4th 25/8/04 10:30 AM Page 75

Photo 4.2
Spring balances.

meaning 'to rub'. Galileo's ideas were very bold for his time because he was not able to verify them experimentally. He ended up in hot water with the Church when he asserted that other planets were much the same as Earth and revolved around the Sun, whereas the Church taught that the Sun revolved around the Earth.

MEASURING FORCE 4.2

Simply stated, a force is a push or a pull and it is fitting that the unit of force be named after one of the world's greatest physicists, Isaac Newton.

The **newton** (N) is commonly measured in the laboratory with a device called a spring balance. This has a spring that extends when masses are hung on it or when other forces are applied. The scale is calibrated in grams for mass or in newtons for force. Because the direction of the force is important, **force is a vector quantity**.

EI Activity 4.1 FEELING A NEWTON

The size of 1 newton is not familiar to most people. The 'feel' of a newton helps you in your problem solving.

1. Obtain a spring balance calibrated in newtons and check that it reads zero when held vertically. Adjust it if it doesn't. Pull gently to feel forces of 1 N, 2 N, 3 N etc.
2. Hang masses of 100 g, 200 g etc. on the hook to see what force is needed to hold them up.
3. Hold a 100 g mass stationary in your hand. This requires a force of about 1 N.
4. When you sit on a bicycle, what force does your total mass exert on the bike?
5. Use bathroom scales under the front and rear wheel of your bike to see how this force is distributed.

BALANCED AND UNBALANCED FORCES 4.3

To study the effect of forces acting on an object we need to distinguish between **balanced** and **unbalanced** forces. When spring balances are hooked onto either end of a cart and given equal pulls in opposite directions (Figure 4.1), the carts remain at rest because the forces are balanced — they are **equal and opposite**.

Figure 4.1
The forces on the cart are balanced — they are equal and opposite in direction.

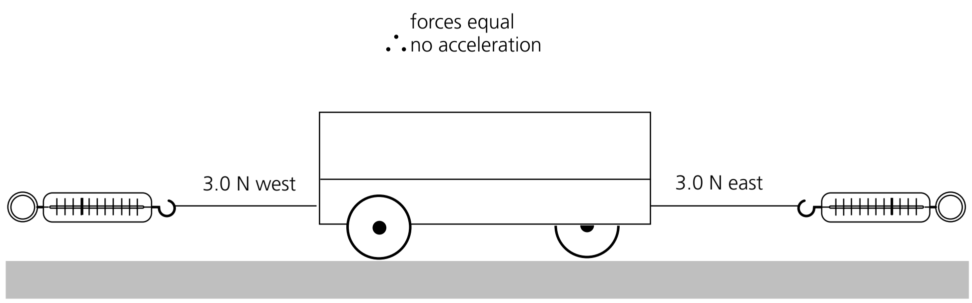
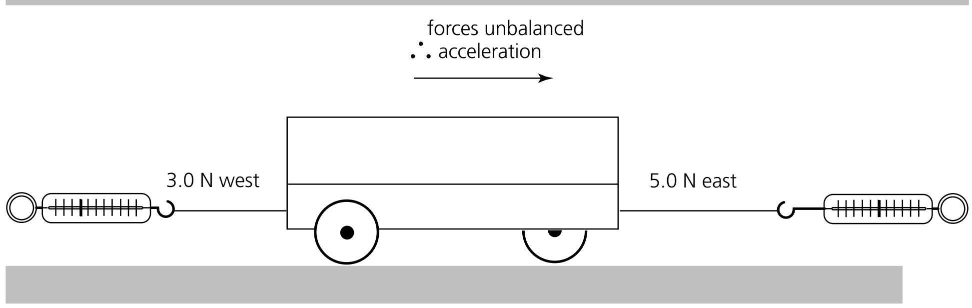


Figure 4.2
The force on the cart is unbalanced — it is greater to the right than to the left.



If the pull on the balance to the right was increased to 5 N (Figure 4.2), then the forces would become unbalanced and the cart would move off in the direction of the larger force.

Example 1

The resultant force acting on the cart in Figure 4.2 can be calculated. It seems obvious but it is important to get the setting-out correct.

Solution

Finding the resultant force is a vector addition thus:

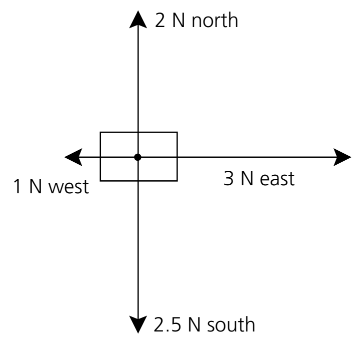
$$\begin{aligned} \boldsymbol{F}_R &= \boldsymbol{F}_1 + \boldsymbol{F}_2 \\ &= 5.0 \text{ N east} + 3.0 \text{ N west} \\ &= 5.0 \text{ N east} + (-3.0 \text{ N east}) \\ &= 2.0 \text{ N east} \end{aligned}$$

If there are several forces acting on an object you should try to reduce them to a simpler case by adding pairs in opposite directions first before combining with forces at angles.

Example 2

Calculate the resultant force acting on the object shown in Figure 4.3.

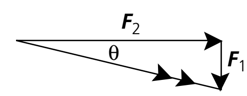
Figure 4.3



Solution

$$\begin{aligned} \boldsymbol{F}_1 &= \boldsymbol{F}_S + \boldsymbol{F}_N \\ &= 2.5 \text{ N south} + (2.0 \text{ N north}) \\ &= 2.5 \text{ N south} + (-2.0 \text{ N south}) \\ &= 0.5 \text{ N south} \\ \boldsymbol{F}_2 &= \boldsymbol{F}_E + \boldsymbol{F}_W \\ &= 3.0 \text{ N east} + (-1.0 \text{ N east}) \\ &= 2.0 \text{ N east} \\ \boldsymbol{F}_R &= 0.5 \text{ N south} + 2.0 \text{ N east (see Figure 4.4)} \\ &= \sqrt{0.5^2 + 2.0^2} \\ &= 2.1 \text{ N} \\ \theta &= 14° \text{ so the direction is E14°S} \end{aligned}$$

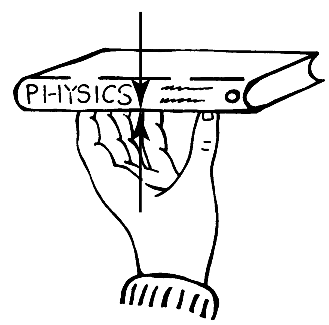
Figure 4.4



Questions

1. Calculate the resultant force when the following forces act on the same object: **(a)** 2.4 N north, 1.8 N south, 1.9 N north; **(b)** 65 N down, 92 N up and 74 N up; **(c)** 50 N north, 30 N west, 60 N south; **(d)** 26 N west, 20 N east, 30 N north, 15 N south.
2. Figure 4.5 shows a physics book held at rest in a person's hand. Two forces are shown in the diagram. One is the weight of the book pushing down and the other is the force of the hand pushing up.
   (a) Are the forces balanced? Explain.
   (b) Assume the hand was suddenly removed. Are the forces now balanced? What would you observe?

Figure 4.5
'Are the forces balanced?'



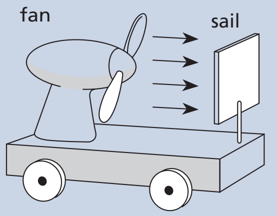
4.4 NEWTON'S FIRST LAW OF MOTION

Sir Isaac Newton was the first scientist to put Galileo's ideas into the form of a universal physical law, that is, one obeyed throughout the universe. In 1688, Newton proposed the first law of motion:

> An object maintains its state of rest or constant velocity motion unless it is acted on by an external unbalanced force.

NOVEL CHALLENGE

A fan blows a cart with a sail attached.



If the fan and the sail are on the same cart, what happens? Explain why.

PHYSICS FACT

At 5.00 pm Houston time on 17 July 1969, the *Apollo 11* spacecraft was 50 000 km from the Moon on its return journey to Earth with its engines off. Its speed was 4740 km/h at 4.30 pm and at 6.00 pm its speed was still the same. With no net force its velocity remained constant.

TEST YOUR UNDERSTANDING

Why does the agitator in a washing machine go back and forth instead of going steadily in one direction? Explain in terms of Newton's first law.

The following four examples show Newton's law applied to real life.

At rest and stays at rest

Some magicians can jerk a tablecloth out from under a dinner setting of glasses and cutlery, leaving them at rest on the table.

In motion and stays in motion

In a head-on car crash, the occupants tend to continue in their state of motion and move forward towards the dashboard. It is usually the seat belts that restrain them.

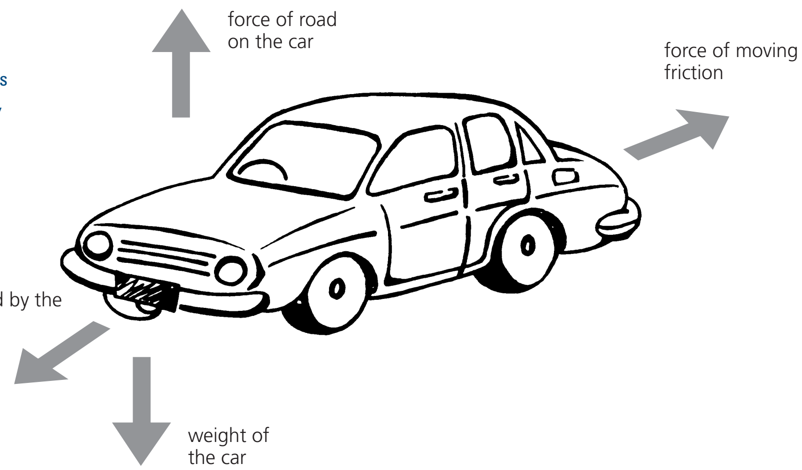
Not wanting to change direction

As a car goes round a corner, your body wants to continue in a straight line so the car door presses against you as it moves sideways. People often say that they get flung against the car door. It is actually the door that gets flung against them.

Balanced forces, constant velocity

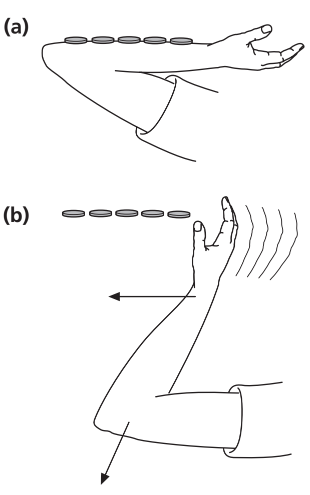
Consider a diagram of the forces acting on a car travelling along a road at constant velocity (Figure 4.6).

Figure 4.6
When this car travels at constant velocity, all the forces acting on it are balanced.



The downward force of the car on the road is balanced by the upward force of the road on the car. The force produced by the engine is balanced by the friction of the tyres on the road and the air resistance. As long as these forces remain balanced the car will not accelerate.

Figure 4.7
Coins stay still long enough for you to catch them.



EI Activity 4.2 INERTIA

Two tricks you can do involve Newton's first law.

1. Place a row of coins along your forearm as shown in Figure 4.7 **(a)**.

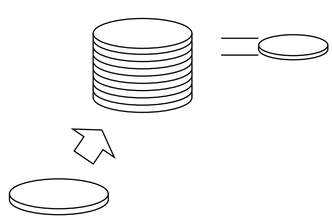
   With practice, as you fling your arm down (Figure 4.7**(b)**), the coins should stay motionless long enough for you to catch them in your hand.

2. Make a pile of 20 cent coins on a smooth surface as shown in Figure 4.8.

   Flick another 20 cent coin towards the stack and with practice you should be able to knock the bottom coin out without disturbing the rest.

Describe how Newton's first law applies in these situations.

Figure 4.8



Questions

3 A thread supports a mass hung from the ceiling. Another identical string is tied to the bottom of the mass (Figure 4.9). Which thread is likely to break if the bottom thread is pulled **(a)** slowly; **(b)** quickly? Explain your prediction.

4 Explain how Newton's first law applies in the following cases:
(a) You flick your hands after washing them, before you use a towel.
(b) You spin your wet umbrella to remove excess water before folding it up.
(c) You can't stay upright on a stationary bicycle without putting your feet down but you have no problem while you ride along.
(d) Falling off a building and accelerating is not dangerous but the deceleration bit at the end is.
(e) Boxers get 'punch drunk' after too many blows to their head.

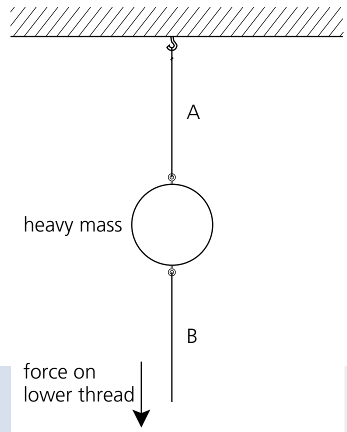


Figure 4.9
For question 3.

4.5 MASS

Everyday experience tells us that a given force will produce different accelerations in different objects. Kick a football and it moves off quickly. Kick a car and it hardly moves. The difference is their **mass**. Obviously, the car has a greater mass than the ball. The word 'mass' was first used in the fourteenth century in this sense. It comes from the Greek *maza* meaning 'barley cake', hence 'lump or mass'. Mass is measured in units of **kilograms** although grams and tonnes are widely used.

What is mass?

Since the word mass is used in everyday language we should have some understanding of it. Is it a body's size, weight or density? The answer is no, none of these, though these characteristics are sometimes confused with mass. The **mass** of a body is a characteristic of its resistance to motion. This is also called its **inertia**. It was astronomer Johannes Kepler who first used the term in physics in the seventeenth century. At the time, in Latin it merely meant 'lack of art' (*in* = not, *ert* = art), 'no skill' or 'idleness'. It was Kepler's wit that saw the term added to our language. Newton's first law of motion may also be called the 'law of inertia'.

Measuring mass

Mass can be defined and measured in two main ways: as inertial mass or as gravitational mass.

Inertial mass is a measure of resistance to motion. If a known force is applied to different objects, then the resultant acceleration is directly related to mass. A 1 kg object will accelerate at twice the rate as a 2 kg object. An 'inertial balance' (Figure 4.10) can be used to measure inertial mass. The object to be measured is placed on the outer section, which is then given a push and allowed to vibrate back and forth. The greater the mass, the slower the rate of vibration. Such a balance is used in the space shuttle to measure astronauts' mass. The laboratory version is sometimes called the 'wig-wag' machine.

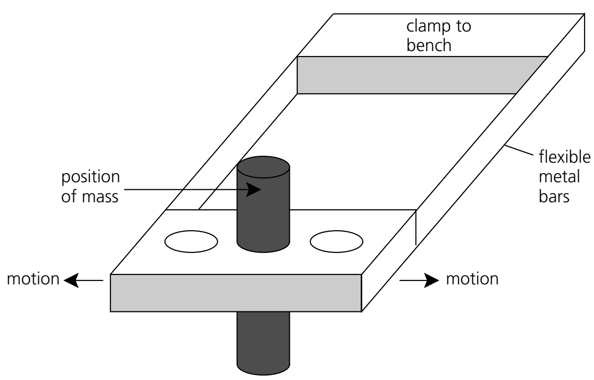


Figure 4.10
The wig-wag machine (inertial balance).

Gravitational mass is a measure of the pull of gravity on an object. A spring balance is often used to measure gravitational mass. It works on the principle that the force of gravity (weight) of an object is proportional to the mass of an object. A beam balance can also be used to compare weights and hence masses of objects (Figure 4.11).

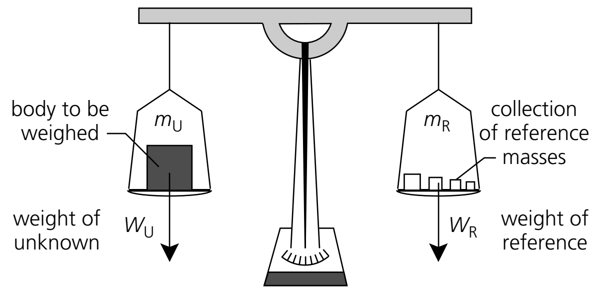


Figure 4.11
An equal-arm balance. When the device is in balance, the masses on the left and the right pans are equal.

NOVEL CHALLENGE

Isaac Newton's mother said that he would fit into a 'quart pot' at birth. If the density of a baby is 1020 kg m^{-3}, calculate his mass.

Gravitational mass and inertial mass are equivalent. (See Chapter 30.)

Density

Which is more dense — milk or cream? Cream is certainly thicker but it floats on the top of milk so it is less dense than milk. You can't go around floating objects on top of each other to compare their densities. This may work for two liquids but what about two solids? A standard definition is needed.

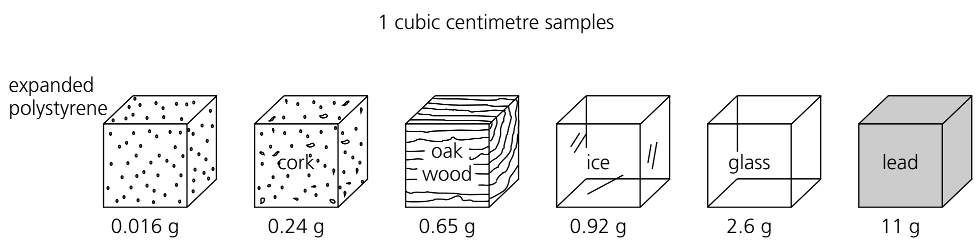


Figure 4.12
Comparing the masses of 1 cubic centimetre of several substances illustrates the different densities.

Density is the mass per unit volume

$$\text{Density} = \frac{\text{mass}}{\text{volume}} \text{ or } D = \frac{m}{V}$$

The units for density will be kg/m^3 or **kg m^{-3}**. Sometimes the unit **g cm^{-3}** is used. In this book we will use both.
Note: the SI symbol for density is the Greek letter 'rho' (ρ) and is preferred to 'D'. It is up to you and your teacher which to use. You may find that in chemistry class the symbol D is used. Density will be dealt with in more detail in the chapter on fluids and buoyancy.

Table 4.1 SOME DENSITIES

MATERIAL	DENSITY (kg m^{-3})	DENSITY (g cm^{-3})
Air (at STP)	1.3	0.0013
Cork	240	0.24
Petrol	800	0.8
Ice	920	0.92
Water	1 000	1.0
Sea water	1 030	1.03
Milk	1 030	1.03
Sand	1 600	1.6
Pyrex	2 230	2.23
Carbon	2 300	2.3
Aluminium	2 700	2.7
Diamond	3 500	3.5
Iron	7 900	7.9
Brass	8 400	8.4
Copper	8 960	8.96
Silver	10 300	10.3
Lead	11 300	11.3
Gold (9 carat)	11 300	11.3
Mercury	13 600	13.6
Pure gold (24 carat)	19 300	19.3

Example

Calculate the density of a cube of copper that has a side of 2.00 cm and a mass of 71.68 g. Give your answer in **(a)** g cm^{-3}; **(b)** kg m^{-3}.

Solution

(a) Volume = $l \times b \times h = (2.00 \text{ cm})^3 = 8.00 \text{ cm}^3$.
Density = mass/volume = 71.68/8.00 = 8.96 g cm^{-3}.

(b) Volume $= (2.00 \times 10^{-2} \text{ m})^3 = 8.00 \times 10^{-6} \text{ m}^3$.
Mass = 71.68×10^{-3} kg.

$$\text{Density} = \frac{\text{mass}}{\text{volume}} = \frac{71.68 \times 10^{-3} \text{ kg}}{8.00 \times 10^{-6} \text{ m}^3} = 8960 \text{ kg m}^{-3} \text{ (or } 8.96 \times 10^3 \text{ kg m}^{-3}\text{)}.$$

Questions

5 The density of iron is 7.86 g cm^{-3}. What is the mass of a cube of iron whose side is 2.50 cm long?

6 Could you lift a ball of cork of diameter 1.5 m? Cork has a density of 0.24 g cm^{-3} (240 kg m^{-3}).

4.6 NEWTON'S SECOND LAW

Newton's first law deals with cases where the forces are balanced, so no acceleration occurs. His second law deals with unbalanced forces and hence acceleration will occur. Here are some examples:

- *Dropping a rock over a cliff.* There is no upward force to balance the force of gravity (assuming we neglect air resistance).

NOVEL CHALLENGE

Isaac Newton's early childhood was marked by rejection and hatred. His step-father refused to allow him into the house. He lived a life of long battles with von Liebnitz about who invented calculus. Robert Hooke hated him. As Master of the Royal Mint Newton frequented brothels and bars trying to catch counterfeiters. From 1678 to 1696 he conducted experiments by heating up heavy metals such as lead and mercury and breathing in the vapour (sweet, saltish, vitriolic) so it was no surprise that he developed mental illness by 1693. Samples of his hair had 200 parts per million mercury where 5 ppm is normal and 40 ppm dangerous. He has been ranked as the second most influential person in the world (influential, not important). *Develop an argument for who might be first and third?*

NOVEL CHALLENGE

(a) *If you took a beaker of wet sand from the beach, would it weigh more or less than the same volume of dry sand?*
(b) *Stand on some wet sand at the beach and it goes dry around your feet. Why is this?*

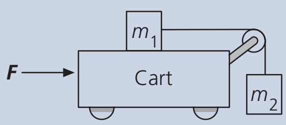
Ch04-Walding 4th 25/8/04 10:30 AM Page 81

NOVEL CHALLENGE

Some people say that Newton discovered the laws of motion but others say he invented them. *Who is right?*

NOVEL CHALLENGE

For m[1] and m[2] to remain in the same positions relative to the cart, what force (***F***) has to be applied?



- *A bullet travelling up a rifle barrel.* The force due to the pressure of hot expanding gases is greater than the friction from the walls of the barrel.
- *Driving away from traffic lights.* The force produced by the engine is greater than the friction of the tyres and air resistance, so a car will accelerate.
  - The heavier the car the more force is needed to accelerate away from the traffic lights.
  - The faster you want to accelerate, the greater the force needed.
  - The acceleration occurs in the direction of the unbalanced force.

Newton's second law summarises these facts:

> The acceleration of an object varies in direct proportion to the external unbalanced force applied to it and inversely proportional to its mass.

Mathematically: $\boldsymbol{a} \propto \boldsymbol{F}$ and $\boldsymbol{a} \propto \frac{1}{m}$ or $\boldsymbol{a} = \frac{\boldsymbol{F}}{m}$.

This can be rearranged to: $\boldsymbol{F} = m\boldsymbol{a}$.

As stated earlier, the unit of force is the newton (N) so we can define a newton as the force needed to give a 1 kg object an acceleration of 1 m s^{-2}.

Example

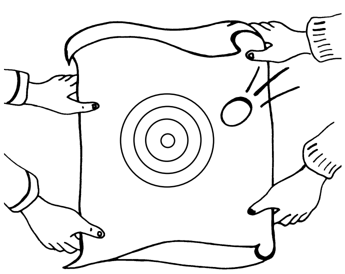
An unbalanced force of 48 N west is applied to a 6.0 kg cart. Calculate the cart's acceleration.

Solution

$$\boldsymbol{a} = \frac{\boldsymbol{F}}{m} = \frac{48\ \text{N west}}{6.0\ \text{kg}} = 8.0\ \text{m s}^{-2}\ \text{west}$$

EI Activity 4.3 THROWING EGGS

Figure 4.13
A soft landing for an egg stops the shell cracking.



Throw some raw eggs full-force at a sheet held as shown in Figure 4.13. Have the bottom curled up to catch the eggs once they have hit the sheet. Explain the result in terms of Newton's second law. If the sheet was wet what do you think would happen? We're not game to try it.

Making your bike go faster

A practical example of Newton's second law involves bicycle racing. In 1986, aerospace titanium–aluminium–vanadium alloys were introduced. These have a density of about 60% of that of steel and so a 300 g steel chain can be replaced by a 130 g one. Hubs can drop by 34 g, and wheels by an amazing 800 g. With over 60 replaceable parts, cyclists can drop about 3 kg of mass from their bikes, which is about 3.5% of total mass of bike plus rider. Tests show that paring just 1 kg from a bicycle can save 2.5 s in a 3 km circuit. That's 16 m at an average speed of 26 km h^{-1}. It also enables them to climb hills at two gears higher than normal and if you consider that the difference between 1st and 10th place is about 0.5% in performance, a 3.5% mass loss can be a winner. The savings are not cheap. A Trek 5500 bicycle with carbon fibre frame (7 kg) and Ti–Al–V alloy throughout is about $4000.

Loss of consciousness

It has been known for a long time that rapid acceleration or deceleration can severely affect the human body. Too high a deceleration can cause loss of consciousness, for example in a sharp loop-the-loop by a jet fighter pilot and crew. Alternatively, it can result in death. Smashing into a power pole can kill a car driver and of course that is why cars have 'crumple zones' to slow the rate of deceleration in an accident.

Questions

7 Calculate the missing quantities in Table 4.2 (do not write in this book).

Table 4.2

	F	*m*	*u*	*v*	*a*	*t*
(a)		1000 kg	rest	25 m s^{-1}		8.5 s
(b)	25 N	15 kg	rest			2.0 s
(c)	1000 N		10 m s^{-1}	40 m s^{-1}	10 m s^{-2}	
(d)		200 g	0.85 m s^{-1}	0.60 m s^{-1}		1.5 min
(e)	150 N			rest	−2.2 m s^{-2}	4 s

NOVEL CHALLENGE

Quick now, could you lift a ball of cork 1.5 m in diameter? Now work out its mass.

8 A car of mass 2000 kg decelerates from 30 m s^{-1} to rest in a distance of 100 m. Calculate the retarding force required to stop the car.

9 Racing driver David Purley (1945–85) survived a deceleration from 173 km/h to zero in a distance of 66 cm in a crash at Silverstone, UK in 1977. He suffered 29 fractures, three dislocations, six heart stoppages and made the *Guinness Book of Records*. Calculate the net horizontal force acting on him in the crash. His body mass at the time was 55 kg.

10 Comment critically on the following claims:

(a) It requires a greater force to accelerate a 2000 kg car from rest to 15 m s^{-1} than from 15 m s^{-1} to 30 m s^{-1} in the same time.

(b) Twice the force is needed to accelerate a 1.5 t car from rest to 60 km^{-1} over 100 m than is required over 200 m.

(c) An object always accelerates in the direction of the net force.

(d) A lower net force is needed to accelerate an object from rest to 10 m s^{-1} than is required to accelerate it from rest to 20 m s^{-1} irrespective of the time taken.

11 In an experiment to find out how the motion of a trolley was related to the force acting on it, a 1.5 kg trolley was accelerated by various forces. The results are summarised below:

Force (N)	0.00	0.10	0.20	0.30	0.40	0.80
Acceleration (m s^{-2})	0.00	0.07	0.13	0.20	0.27	0.53

(a) Plot the data with ***F*** on the *x*-axis.

(b) What relationship is suggested by the data?

12 In an experiment to verify Newton's second law, the equipment shown in Figure 4.14 was set up.

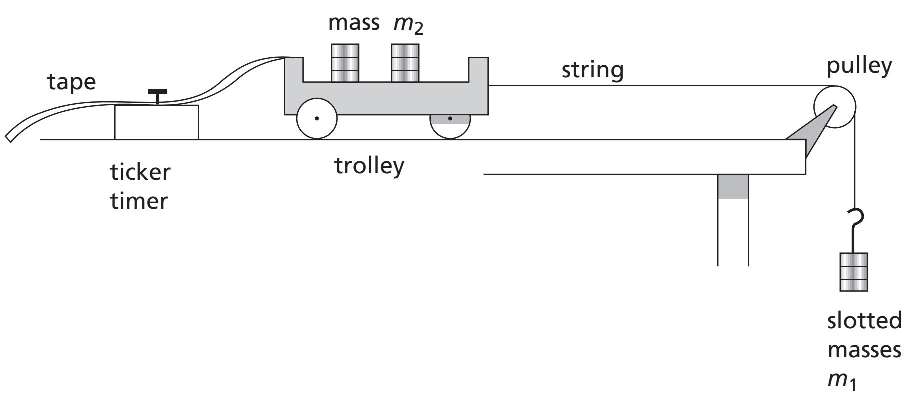


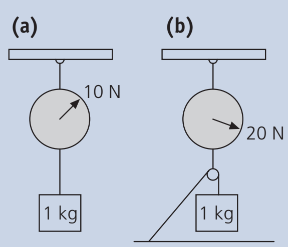
Figure 4.14
As masses are transferred from the trolley to the carrier, the force exerted on the trolley is increased while the mass of the whole system remains constant.

NOVEL CHALLENGE

For most cars, the rear tyres support more weight than the front tyres. For example a Toyota Corolla has 43% of its weight supported by the front tyres and 57% by the rear. When a Corolla brakes, the weight on the front increases to about 69% and reduces to 31% on the rear. Why is this, and why do cars dip when the brakes are applied?

NOVEL CHALLENGE

If a 1 kg mass hanging on a spring balance shows a weight of 10 N, as in diagram **(a)**, will diagram **(b)** be correct? The second diagram shows a 1 kg mass suspended over a pulley by a string tied to the table. Explain.



The mass m_1 hangs vertically and its weight (the force of gravity) is responsible for providing the accelerating force that causes the mass m_2 to move in a horizontal direction. The weight of m_1 is equal to 9.8 N for every 1 kg of hanging mass. As both masses were connected by a light string, the total mass being accelerated by the weight of m_1 is equal to the sum of m_1 and m_2. To keep the total mass of the system constant but to vary the accelerating force (the weight of m_1), the brass masses were shifted from the trolley to the hanger. This made m_1 heavier and m_2 lighter by the same amount. The acceleration was measured by a ticker timer. The results were as shown in Table 4.3.

Table 4.3

HANGING MASS m_1 (g)	MASS OF GLIDER AND m_2 (g)	ACCELERATION a (m s^{-2})
100	750	1.15
200	650	2.45
300	550	3.46
400	450	4.52
500	350	5.76
600	250	6.90

(a) For each case, calculate the force applied ($\boldsymbol{F}_A$) by multiplying the hanging mass (kg) by 9.8 and expressing the answer in newtons (N).
(b) Plot acceleration vs force and comment on whether Newton's second law is confirmed.
(c) How are total mass and the ratio $\boldsymbol{F}/\boldsymbol{a}$ related?
(d) Predict what would happen to the shape of the graph if there was some friction present.
(e) How would you modify the experiment to keep the force the same but vary the total mass?

NEWTON'S THIRD LAW OF MOTION 4.7

PHYSICS FACT

The main gun on the British warship HMS Invincible had a bore (internal diameter) of 16 inches (40 cm) — and that's huge! Its shells could penetrate 24 inches (60 cm) of steel armour on enemy ships. The recoil of the gun was so great that it would buckle up the wooden deck and peel off the paint.

Forces come in pairs.

- If a hammer exerts a force on a nail, the nail exerts an equal and opposite force on the hammer.
- If you lean against a wall, the wall pushes back on you.

These situations can be summed up by the words 'you cannot touch without being touched'. They are examples of **Newton's third law of motion**, which states:

> To every action there is an equal and opposite reaction.

In other words, '*if a body A exerts a force on another body B, then body B exerts an equal and opposite force on body A*'. This second way of expressing the law seems more wordy but is more precise. There can be problems working out the **action** and **reaction** pairs in some situations.

Examples of other action–reaction pairs are given in Table 4.4.

Table 4.4 ACTION—REACTION PAIRS

ACTION	REACTION
• Exhaust gases are pushed out of the rocket	• The rocket pushed forward by the gas
• A sprinter pushes on the starting blocks	• The blocks push on the sprinter
• A vase of flowers presses down on a table	• The table pushes upward on the vase
• A tyre pushes on the road	• The road pushes on the tyre
• A softball is hit by a bat	• The bat slows down as the ball pushes back
• An orbiting satellite is attracted to the Earth	• The Earth is attracted to the satellite

Consider a packet of biscuits resting on a table. The two forces can be labelled as $\boldsymbol{F}_W$, the weight of the biscuits; and $\boldsymbol{F}_N$, the **normal reaction** force that the table pushes up normal to the surface. 'Normal' means at right angles (Figure 4.15).

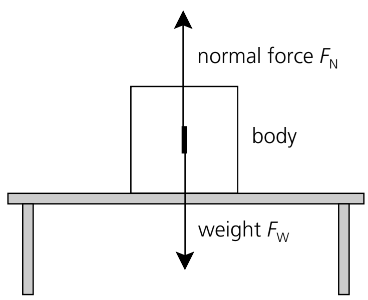


Figure 4.15

In this case they are equal and opposite because no acceleration is occurring. If the mass of the biscuits was 125 g (0.125 kg) for instance, then the weight would be 1.25 N down and the normal reaction force would be 1.25 N upward.

Questions

13 What are the reactions to the following actions: **(a)** a tennis ball is hit by a racquet; **(b)** a horse walks along a road; **(c)** a horse drags a log along a dirt track; **(d)** a click beetle jumps into the air; **(e)** a man falls out of a tree?

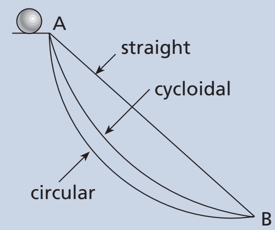
14 A vase of mass 2.5 kg rests on a table.

(a) What is the normal reaction force exerted by the table on the vase?

(b) What if a very large mass such as 500 kg is placed on the table? Explain.

NOVEL CHALLENGE

Brachistochrone



In the summer of 1693, John Bernoulli posed the following problem, which still hadn't been solved 6 months later. The day Newton heard it he solved it. *The problem: you have three paths for a ball to roll down (see figure) — which is the fastest?* The device is called a *Brachistochrone* (Greek *brachy* = 'short', *chronos* = 'time'). Newton didn't give his name but Bernoulli said *'Tanquam ex ungue leonem'* — Latin for 'as the lion is known by his claw'. He recognised the genius of Newton. If you put a ball anywhere on the cycloid it will take the same time to reach B. *Strange — Why? The cycloid is called a* tautochrone *(Gk* tauto *= 'same').*

4.8 FORCE, WEIGHT AND GRAVITY

Objects held above the Earth's surface are attracted towards the Earth's centre by a pulling force called the **force of gravity**. This pulling force causes freely falling objects to accelerate downward. **Gravitational attraction** occurs between all bodies but a general discussion is left until a later chapter.

If you drop a rock, it would accelerate downward at 9.8 m s^{-2} provided that no other external forces interfere with its motion. This is called its **free-fall** acceleration. The force due to gravity is given by $\boldsymbol{F} = m\boldsymbol{a}$. For a 1 kg rock the force of gravity equals 1 kg × 9.8 m s^{-2} or 9.8 kg m s^{-2} directed downward. This is 9.8 newtons. To hold the rock steady in your hand requires an equal and opposite force. This is a measure of the **weight** of the rock. It has the symbol $\boldsymbol{F}_W$ although some teachers and texts prefer to use just $\boldsymbol{W}$. **The SI convention is to use *F* for force and different subscripts for specific types of force.** The word 'weight' has been around for a long, long time, hence the confusion when a common term is given a specific meaning in physics. It comes from pre-historic German where *wegan* meant 'to carry'. Its secondary use meaning 'heaviness' was an Old English adaptation.

NOVEL CHALLENGE

In astronaut Neil Armstrong's biography, it says that on Phobos — one of the two potato-shaped moons of Mars — he would weigh only 3 ounces. *If 'g' on Phobos is one-thousandth that of its value on Earth, what would his mass be in kg?*

INVESTIGATING

You have been provided with a ball, a stopwatch and a tape measure.
How many different ways can you think of to measure the distance from a top floor verandah to the ground below? List sources of error and comment on the most accurate method.

INVESTIGATING

If you weigh yourself on bathroom scales the reading is greater if the scales are on carpet rather than on a hard floor. Now why is this? It would make a great investigation — in fact it was reported in *New Scientist*.

The formula relating mass to weight is then:

$$F_W = mg \quad \text{(where } g = \text{acceleration due to gravity)}$$

On Earth, $g = 9.8$ m s^{-2} but does vary from location to location. The value of acceleration due to gravity on other planets depends largely on the planet's mass and is shown in Table 4.5.

Table 4.5 FREE-FALL ACCELERATIONS

BODY	ACCELERATION DUE TO GRAVITY (m s^{-2})
Jupiter	24.9
Neptune	11.8
Saturn	10.5
Earth (average)	9.8
Earth (poles)	9.83
Earth (equator)	9.78
Venus	8.8
Uranus	7.8
Mercury	3.7
Mars	3.7
Moon	1.6

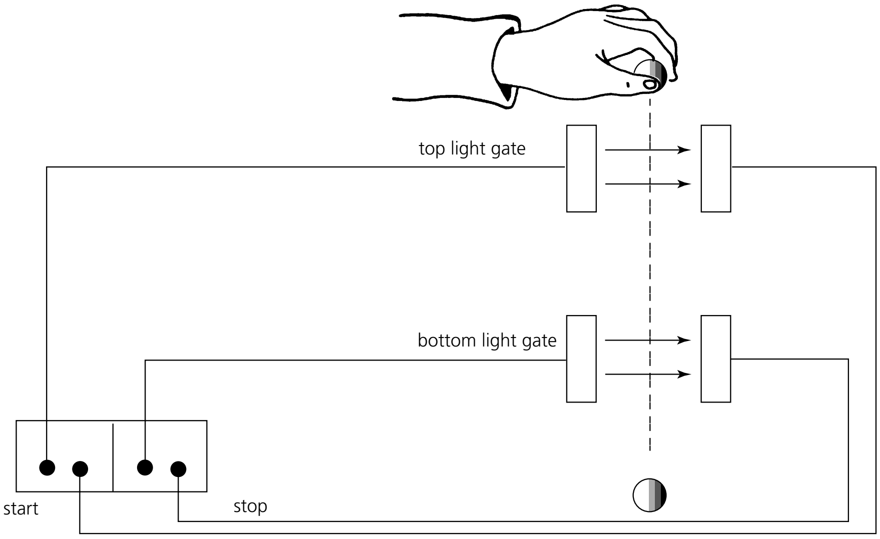
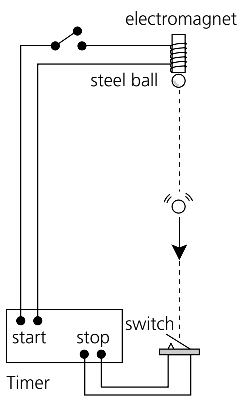


Figure 4.16
Light gates can record the time interval of a falling ball.

Figure 4.17
The ball falls when the electromagnet turns off. This starts the timer.



Several devices are available in the laboratory for measuring acceleration due to gravity. Because you are limited to a vertical distance of about 2 m, the timing of the fall is very short and can't be done by hand with a stopwatch.

Two common methods to measure the time are with:

- light gates (Figure 4.16)
- a mechanical switch (Figure 4.17).

A light gate consists of a light source and a detector. When an object is dropped through the top gate, the counter-timer is started and when the object passes through the lower one the timer stops. Alternatively, a striped piece of plastic can be used as part of a computer interfacing package.

The other method consists of an electromagnet holding a ball bearing high up. When the switch is opened the electromagnet turns off and the ball falls. At the same instant, the opening of the switch starts a timer. When the ball hits the lower plate a switch is closed and the timer stops.

Activity 4.4 BODY MASS INDEX

Body mass index (BMI) is a mathematical formula that correlates mass and height to determine how much body fat you have. See Table 4.6 below. BMI is a better predictor of health risk than simple body mass measurements. BMI should not be used to assess competitive athletes or body builders because of their relatively larger amount of muscle. Neither should pregnant or lactating women, children, or frail sedentary elderly people use BMI.

1. What is your BMI? (BMI = mass/height^2, with mass in kilograms and height in metres.)
2. Is there any evidence that BMI varies during different stages of a woman's menstrual cycle?
3. Are there any other reasons *not* to use BMI as a health index?

Table 4.6

MASS CATEGORY	BODY MASS INDEX
'Underweight'	<18.5
Healthy	18.5 – 24.9
'Overweight'	25 – 29.9
Obese	>30

NOVEL CHALLENGE

A man comes up to a bridge that can just support his weight and 1 of the 4 balls he is carrying. He decides to juggle as he crosses so that only 1 ball will be in his hands at any one time. *What do you think of his solution?*

Weight and mass

Weight is different from mass. *Mass* is a measure of an object's resistance to motion and doesn't vary no matter where the object might be taken to in the universe. *Weight* is a measure of the force of gravity acting on that object and will vary depending which planet it is on or what gravitational forces it is being subjected to. Refer back to Table 4.5. For a body of mass m located at a point where the free-fall acceleration is $\boldsymbol{g}$, then the magnitude of the weight (force) vector is given by $\boldsymbol{F}_W = m\boldsymbol{g}$.

You could measure the weight of a body by placing it on some bathroom scales or by hanging it from a spring balance. With bathroom scales, gravity pulls downwards with a force we've called weight $\boldsymbol{F}_W$ (= $m\boldsymbol{g}$) and the scales push up with a reaction force normal to the surface ($\boldsymbol{F}_N$). $\boldsymbol{F}_N$ could also be called the 'scale reading'.

In this case, as there is no acceleration, the magnitude of the two forces are equal in magnitude, so $\boldsymbol{F}_W = \boldsymbol{F}_N$ and the scales read the true weight of the body.

For example, if you have a mass of 60 kg then your weight on Earth is 60 × 9.8 or 588 N. It is fairly usual to approximate the free-fall acceleration due to gravity on Earth as 10 m s^{-2} so the relationship is simplified to: weight (N) = mass (kg) × 10 m s^{-2}. Your teacher will tell you if you are to use 10 or 9.8 in problems. Certainly for experiments you would use 9.8 m s^{-2}.

The free-fall acceleration on the Moon is 1.6 m s^{-2} so on the Moon $\boldsymbol{F}_W = m \times 1.6$ N = 96 N. If you want to lose weight, fly to the Moon.

Example

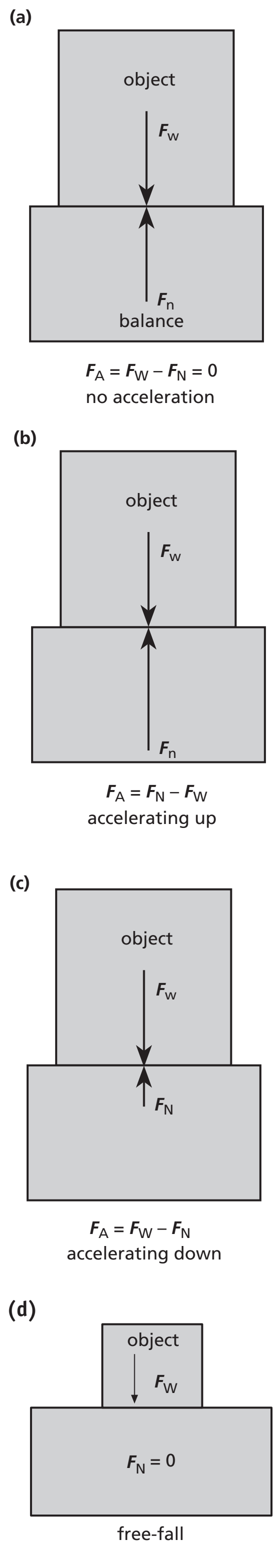
A man has a mass of 65 kg. Calculate his weight on **(a)** the Earth; **(b)** the Moon; **(c)** Jupiter.

Solution

(a) $\boldsymbol{F}_W$ = mg = 65 × 10 = 650 N.
(b) $\boldsymbol{F}_W$ = mg = 65 × 1.6 = 104 N (100 N).
(c) $\boldsymbol{F}_W$ = mg = 65 × 24.9 = 1618 N (1600 N).

Ch04-Walding 4th 25/8/04 10:30 AM Page 87

Figure 4.18
Accelerated frame of reference.



Apparent weight

In the example above, the body was at rest on the surface of Earth or a planet. The resultant force is zero ($\boldsymbol{F}_W = \boldsymbol{F}_N = 0$) and this would hold as long as the body was not accelerating (Figure 4.18**(a)**). If it was accelerating, the resultant force would not be zero, according to Newton's second law. This is considered below.

Your weight ($m\boldsymbol{g}$) can be considered constant at the surface of the Earth. This is sometimes called your true weight but is really just 'weight'. There are slight variations in $\boldsymbol{g}$ due to the different composition of rocks below but it is on average about 9.8 m s^{-2}. Any effects due to the rotation of the Earth will not be considered until Chapter 6.

Your weight ($m\boldsymbol{g}$) can appear to change in two main ways:

- by being in an accelerated frame of reference (skydiving, or going up or down in an elevator)
- by buoyancy effects (floating in a swimming pool).

Accelerated frame of reference

Going up in an elevator

If you were standing on some bathroom scales in an elevator and it accelerated upwards, the reading on the scales would increase (Figure 4.18**(b)**). The reason: as with a non-accelerated (or 'inertial') frame of reference, such as a stationary elevator or one travelling at constant velocity as described above, you are acted on by two forces. One is gravity, which pulls down with a force $\boldsymbol{F}_W$ (= $m\boldsymbol{g}$) called your weight; the other is the normal reaction force ($\boldsymbol{F}_N$), which pushed the scales upwards onto the soles of your feet. But the body is accelerated upwards, so $\boldsymbol{F}_N$ and $\boldsymbol{F}_W$ cannot be equal; in fact, $\boldsymbol{F}_N$ must be greater than $\boldsymbol{F}_W$. The difference in magnitude is the resultant or applied upward force ($\boldsymbol{F}_A$):

$$\boldsymbol{F}_A = \boldsymbol{F}_N - \boldsymbol{F}_W$$
$$\boldsymbol{F}_N = \boldsymbol{F}_W + \boldsymbol{F}_A$$

Apparent weight = weight + applied force or

$$\boldsymbol{F}_N = m\boldsymbol{g} + m\boldsymbol{a} = m(\boldsymbol{g} + \boldsymbol{a})$$

The bathroom scales push upward with a force $\boldsymbol{F}_N$ whose magnitude is the reading on the scales. It is called the apparent weight.

Going down in an elevator

In this case the upward force ($\boldsymbol{F}_N$) is less than the downward force of gravity ($\boldsymbol{F}_W$) (Figure 4.18**(c)**). The difference is equal to the applied force ($\boldsymbol{F}_A$):

$$\boldsymbol{F}_A = \boldsymbol{F}_W - \boldsymbol{F}_N$$
$$\boldsymbol{F}_N = \boldsymbol{F}_W - \boldsymbol{F}_A$$

Apparent weight = weight − applied force

$$\boldsymbol{F}_N = m\boldsymbol{g} - m\boldsymbol{a} = m(\boldsymbol{g} - \boldsymbol{a})$$

Free-fall

Free-fall is another accelerating frame of reference. Here the upward force of the scales is zero so your acceleration is equal to $\boldsymbol{g}$. If you were standing on some bathroom scales on a trap door in a floor and the door opened, the scales would not register a reading so your apparent weight would also be zero as you fell (Figure 4.18**(d)**).

Free-fall is really quite a different case from the accelerated frames of reference mentioned in the above cases of the accelerator going up and down. Floating in a pool is an inertial (non-accelerated) frame of reference. There is a buoyant force acting upward that is equal to your weight so that your resultant downward force is zero. This is a completely different case from free-fall. Buoyancy acts as someone grabbing you under the arms and lifting you off the scale — it has nothing to do with acceleration. The only thing in common is the term 'apparent weight', which is zero in both cases.

Example

What is the weight and what is the apparent weight of a 50 kg person in a lift that is **(a)** accelerating upwards at 1.5 m s^{-2}; **(b)** accelerating downwards at 1.5 m s^{-2}; **(c)** in free-fall?

Solution

The weight in all cases is equal to mg (that is, $\boldsymbol{F}_W = m\boldsymbol{g}$): $50 \times 10 = 500$ N.
Case **(a)**: Apparent weight ($\boldsymbol{F}_N$) = $m\boldsymbol{g} + m\boldsymbol{a} = m(\boldsymbol{g} + \boldsymbol{a}) = 50(10 + 1.5) = 575$ N.
Case **(b)**: Apparent weight ($\boldsymbol{F}_N$) = $m\boldsymbol{g} - m\boldsymbol{a} = m(\boldsymbol{g} - \boldsymbol{a}) = 50(10 - 1.5) = 425$ N.
Case **(c)**: Apparent weight ($\boldsymbol{F}_N$) = $m\boldsymbol{g} - m\boldsymbol{a} = m(\boldsymbol{g} - \boldsymbol{a}) = 50(10 - 10) = 0$ N.
Note: as bathroom scales are calibrated in kg, the scale reading in Case **(a)** would be 57.5 kg; in Case **(b)** 42.5 kg; and in Case **(c)** 0 kg.

Questions

15 **(a)** How many times heavier by weight would a person be on Saturn than on Earth?
(b) Why does Table 4.5 have three values for Earth?

16 In hospitals, newborn babies have their 'weight' recorded in grams but the nurses usually convert this to pounds and ounces for the parents' benefit.
(a) If a physicist had a 7 lb 8 oz baby, what would its weight be? (1 kg = 2.2 lb; 16 oz = 1 lb.)
(b) If the baby was born in the 'weightless' conditions of outer space, how could the parents measure the baby's Earth weight for the benefit of relatives at home?

17 Calculate the apparent weight of a 70 kg person under each of the following conditions: **(a)** Floating in water; **(b)** Free-falling off the stage at a concert; **(c)** Accelerating upward in a lift at 0.5 m s^{-2}; **(d)** Accelerating downward in a lift at 0.5 m s^{-2}.

4.9 TERMINAL VELOCITY

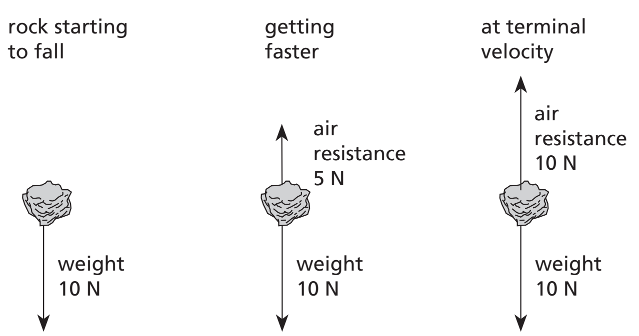


Figure 4.19
Forces acting on a rock falling freely to the ground.

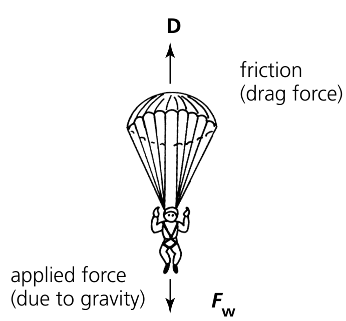


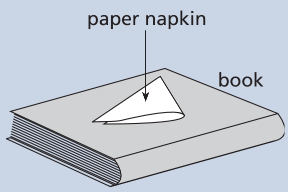
Figure 4.20
Forces acting on a parachutist.

If you drop a rock off a cliff, it gets faster and faster as gravity causes it to accelerate towards the ground. However, as it speeds up, air resistance also increases so eventually the force upwards due to the air equals the force of gravity (down) and the rock stops accelerating and falls at a constant velocity thereafter. This is called its **terminal velocity**. 'Terminal' merely means 'the end' (Latin *terminus* means 'limit').

Without air resistance, objects would accelerate at 10 m s^{-2} toward the ground. In general, the frictional force between an object and the medium through which it is moving is called the **drag force** (Figure 4.20).

NOVEL CHALLENGE

Imagine you were to drop a book that has a paper napkin resting on the top (see Figure). *How will they both fall? Wrong! Try it. Why?*



NOVEL CHALLENGE

If you dropped a marble, a big styrofoam ball and a small one (from a bean bag) together from chest height, two hit the ground at the same time. *Which two and would they be faster or slower than the other one? Why? Try it and see!*

NOVEL CHALLENGE

Two skydivers are freefalling and before their 'chutes open they try to throw a tennis ball back and forth. Propose some reasons why it won't be possible. *Once their parachutes opened would it then be possible?*

Table 4.7 TERMINAL VELOCITY OF SOME OBJECTS

OBJECT	TERMINAL VELOCITY (m s^{-1})
6 kg steel shot	150
Skydiver (typical)	50
Cricket ball	40
Tennis ball	30
Rock (1 cm)	25
Basketball	20
Ping-pong ball	10
Raindrop (3 mm)	7
Insect	6
Parachutist (typical)	5

The drag force (***D***) depends on three factors: the cross-sectional area (*A*), the velocity (***v***) and the density (ρ) of the fluid in which the object is moving. In air you will free-fall for about 12 s and reach 190 km h^{-1} in about 370 m if you are spreadeagled. If you fall head-first you will reach about 300 km h^{-1} in that time.

The world record for free-fall is held by Joseph Kittinger from the US Air Force, who jumped out of a balloon at 31.3 km altitude. He reached a velocity of 988 km h^{-1} before his parachute opened at 5300 m above the ground. His 26 km free-fall lasted 4 minutes 37 seconds. That would be some stunt for your school fête.

Surviving free-fall

- In April 1987, during a jump, parachutist Greg Robertson noticed that fellow parachutist Debbie Williams had been knocked unconscious in a collision with a third skydiver and was unable to open her parachute. Robertson, who was well above Williams at the time and who had not yet opened his parachute for the 13 500 foot (4100 m) plunge, reoriented his body head-down so as to minimise his area (*A*) and maximise his downward speed. Reaching an estimated speed of 200 miles per hour (90 m s^{-1}), he caught up with Williams and then went into a horizontal spread-eagle to increase drag (***D***) so that he could grab her. He opened her parachute and then, after releasing her, his own, with a scant 10 s before impact. Williams received extensive internal injuries due to her lack of control on landing but survived.
- While flying in a British bomber in March 1944, tail-gunner Nick Alkemade bailed out after being hit by enemy fire. With no parachute he fell 5500 m to the snow-covered ground in Germany. He was uninjured.
- Hostess Vesna Vulovic fell 10 000 m after her DC9 plane blew up over Czechoslovakia in 1972. She hit the ground and fractured her spine, legs and arms; she was in a coma for three days and had amnesia for three weeks. She has since completely recovered.
- In June 1985 a teenage boy, while mountain climbing in California, fell 45 m into a pool of water 1.2 m deep. He walked away (to the applause of his mates, no doubt).
- A chimney sweep in 1952 in London fell 45 m into rubble and sustained a deceleration of 162 '*g*'s.

SR Activity 4.5 DRAG COEFFICIENTS

The drag force (***D***) depends on three factors: the cross-sectional area (*A*), the velocity (***v***) and the density of the fluid (ρ). The formula relating them is: $\boldsymbol{D} = \frac{1}{2}C\rho A\boldsymbol{v}^2$.

The drag coefficient *C* is the constant of proportionality and has values ranging from about 0.4 to 1.0.

As an object falls, its acceleration starts at 10 m s^{-2} but gets smaller and smaller until it is zero at the terminal velocity. At this point the weight of the object equals the drag force. The object of this activity is to plot the acceleration against velocity for a tennis ball under free-fall (Figure 4.21).

1. Measure the mass of a tennis ball and its cross-sectional area. Tables of drag coefficients suggest that a tennis ball has a coefficient of 0.7 and the density of air (ρ) is 1.2 kg m^{-3}.

   Note: the resultant force = $m\boldsymbol{a}$ = weight – drag.

$$\boldsymbol{F}_{\mathrm{R}} = m\boldsymbol{a} = \boldsymbol{F}_{\mathrm{W}} - \tfrac{1}{2}C\rho A\boldsymbol{v}^2$$

2. Use a computer spreadsheet or manually calculate the acceleration for every value of $\boldsymbol{v}$ from 0 to its terminal velocity. If you are doing it manually, perhaps you should increase $\boldsymbol{v}$ in units of 5 m s^{-1}.
3. Graph the results.
4. Would it be possible to prepare a table (or spreadsheet) showing $\boldsymbol{a}$ and $\boldsymbol{v}$ as time elapsed increases from 0 to 100 seconds? Pretty difficult huh?

SR Activity 4.6 CATS AND BAD PHYSICS

Investigating cats falling out of high-rise buildings may sound interesting, but if your results seem odd then maybe your data are incomplete. Here's a review of a non-experimental investigation by two veterinarians, Wayne Whitney and Cheryl Mehlhaff, in a report for the *Journal of the American Veterinary Association*. Read it and answer the questions below.

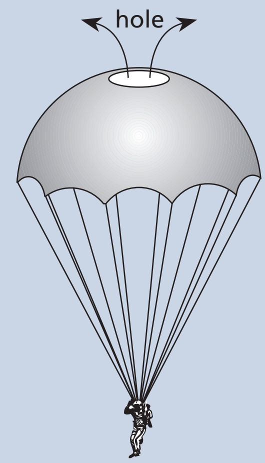
Cats falling from more than seven storeys of high-rise buildings in Manhattan sustained fewer fractures than those that fell from lower levels, according to the researchers. They speculated that a cat relaxes and spreads its limbs more horizontally, like a flying squirrel, after it reaches terminal velocity in a fall. They calculated that the average 4 kg cat reached terminal velocity of about 100 km/h after falling about five storeys. At lower levels, and therefore below terminal velocity, the sense of acceleration may cause them to curl up, making them more prone to injury, Whitney and Mehlhaff suggest.

Only one of 22 cats which fell from more than seven storeys died from its injuries, and there was only one fracture among 13 cats that fell more than nine storeys. In comparison, almost all human falls from over six storeys onto a hard surface are fatal.

Once a cat reaches its terminal velocity, it then begins to slow down. This is because the cat relaxes, changing its position from back arched, head down, and legs pulled tightly underneath its body, to resemble a spreadeagle cat, which slows it down. The reason for this is that our bodies are only sensitive to acceleration, and when you feel acceleration you get scared and curl up in a defensive (foetal) position.

Critics of the study argue that cats do die from great heights and cite the case of Pamela Marx from Brooklyn, New York, who wrote, 'I have had two cats fall from both tenth-floor and fourteenth-floor terraces and both unfortunately died. I never reported these incidents to any medical centre and believe that other people probably don't report their cats' deaths, either. You can add my two cats to your list and report that at least two cats died in fifteen falls over nine storeys.'

NOVEL CHALLENGE



If the objective of a parachute is to slow the descent of a falling object in air, why do parachutes have a hole (the *apex vent*) in the top allowing air to escape? In the Second World War parachutes did not have apex vents and they swung like pendulums as they descended (watch an old war movie, you'll see).
What is the physics behind this?

NOVEL CHALLENGE

If an elephant, a man and a mouse fell from the twentieth storey of a high-rish building, the elephant would splatter on impact and die, the human would be crushed and die, and the mouse would walk away.
Why is this?

PHYSICS FACT

It is often said: 'If a 20 cent coin fell on your head from a high-rise building, it would go through your skull and kill you'. But this is rubbish; it would merely bounce off. Another urban myth bites the dust.

Question

In terms of how you'd carry out a non-experimental investigation, what deficiencies in the data collection and analyses were there? How would you rectify them? Comment critically on the original study.

SR Activity 4.7 RAINDROPS KEEP FALLING ...

The following information enables you to calculate the terminal speeds of raindrops of varying sizes. Raindrops vary in diameter from a maximum of 6.35 mm to a minimum of 0.51 mm.

The drag force is given in Activity 4.5: $\boldsymbol{D} = \frac{1}{2}C\rho A\boldsymbol{v}^2$, where C = the drag coefficient and is 1.2 for a sphere; ρ_a = density of air = 1.2 kg m^{-3}; A = the cross-sectional area of the falling object = πr^2; and $\boldsymbol{v}$ = velocity of the falling object. The radius (r) must be in metres.

At terminal velocity, the drag force equals the weight of the raindrop, and the weight is given by $\boldsymbol{F}_W = m\boldsymbol{g}$. In this equation, $\boldsymbol{g}$ is the acceleration due to gravity (9.8 m s^{-2}) and m is the mass of the raindrop in kilograms. This can be calculated by letting mass = volume x density ($m = \frac{4}{3}\pi r^3\ \rho_W$) where the density of water is 1000 kg m^{-3}. Hence:

$$\boldsymbol{D} = \boldsymbol{F}_W$$

$$\tfrac{1}{2}C\rho A\boldsymbol{v}^2 = \tfrac{4}{3}\pi r^3\ \rho_W\boldsymbol{g}$$

1. Rearrange and simplify this equation to calculate the terminal velocity of a raindrop 5 mm in diameter.
2. How long would it take all possible raindrops to fall from a cloud 2 km above the ground?
3. Draw a graph of raindrop diameter (x-axis) against terminal velocity.

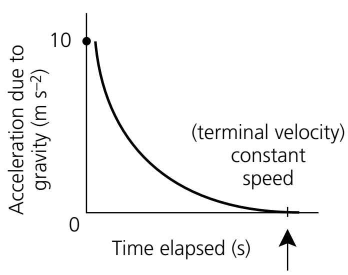


Figure 4.21

SR Activity 4.8 BOATS AFLOAT

The hull speed of a boat is the maximum speed it can reach before drag increases dramatically. For a ship 30 m long it is 25 km h^{-1}; for a 3 m boat it is 9 km h^{-1}; and for a 30 cm hull (e.g. a duck) it is 2.4 km h^{-1}. However, if hull size is too small, surface tension becomes a significant factor. For example, a whirligig beetle has a hull size of 1 cm; its hull speed should be 12 cm s^{-1} (based on extrapolation from the above data) but is actually 25 cm s^{-1} because of surface tension. For a 0.5 cm hull the speed is 30 cm s^{-1} instead of 9 cm^{-1}. But if the hull is too small, the surface tension is too great and the animal can't move. Plot the data and calculate the hull speed of some common animals such as a mouse and a rat. Comment on the accuracy of your analysis.

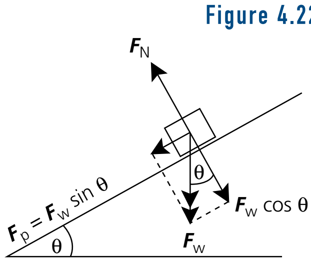


Figure 4.22

INCLINED PLANES 4.10

An object placed on a smooth inclined plane will accelerate down it. The accelerating force is provided by the component of the object's weight in a direction down the plane (Figure 4.22).

Ch04-Walding 4th 25/8/04 10:30 AM Page 92

In the diagram, the weight of the object ($\boldsymbol{F}_W = m\boldsymbol{g}$) has been resolved into two components at right angles — one perpendicular to the plane ($\boldsymbol{F}_\perp = \boldsymbol{F}_W \cos\theta$) and the other parallel to the plane ($\boldsymbol{F}_P$ (or $\boldsymbol{F}_{||}$) $= \boldsymbol{F}_W \sin\theta$). This can be summarised as:

> Parallel component: $\boldsymbol{F}_P$ (or $\boldsymbol{F}_{||}$) $= \boldsymbol{F}_W \sin\theta = m\boldsymbol{g}\sin\theta$
> Perpendicular component $\boldsymbol{F}_\perp = \boldsymbol{F}_W \cos\theta = m\boldsymbol{g}\cos\theta$
> Normal force: $\boldsymbol{F}_N = \boldsymbol{F}_W \cos\theta$

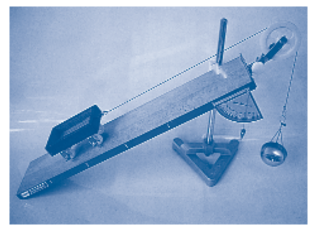
The normal force is equal and opposite to the force perpendicular to the plane because there is no acceleration in that direction.

Using Newton's second law, the resultant force equals $m\boldsymbol{a}$.

Hence: $m\boldsymbol{a} = m\boldsymbol{g}\sin\theta$, so $\boldsymbol{a} = \boldsymbol{g}\sin\theta$.

Note: the mass term cancels out, so we can say the acceleration is independent of the mass of the object in this very specific case. Notice also that this component of gravitational acceleration is down the plane. See Photo 4.3 of an inclined plane apparatus.

Photo 4.3
Demonstration inclined plane used in physics classes.



Limiting cases

From your knowledge of trigonometry you should see that:

- as θ approaches 0°, $\boldsymbol{F}_P$ approaches zero and $\boldsymbol{F}_\perp$ approaches $\boldsymbol{F}_W$
- as θ approaches 90°, $\boldsymbol{F}_P$ approaches $\boldsymbol{F}_W$ and $\boldsymbol{F}_\perp$ approaches zero.

Example 1

A 15 kg wedding cake is allowed to slide freely down a smooth 30° incline. Find **(a)** the resultant force down the incline; **(b)** the acceleration of the object.

Solution

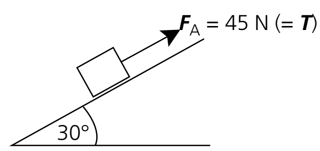
(a) $\boldsymbol{F}_P = m\boldsymbol{a} = m\boldsymbol{g}\sin\theta = 15 \times 10 \times \sin 30° = 75$ N.
(b) $\boldsymbol{a} = \boldsymbol{g}\sin\theta = 10 \times \sin 30° = 5$ m s^{-2}.

Example 2

An 8 kg carton of soft drink is being pulled up a frictionless 30° incline using a rope and an applied force ($\boldsymbol{F}_A$) of 45 N (Figure 4.23). This applied force through the rope is often called the rope **tension** and can be given the alternative symbol $\boldsymbol{T}$.

Calculate the acceleration, if any, up the incline.

Figure 4.23

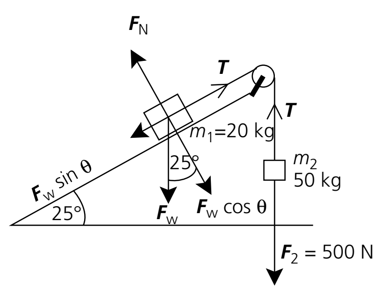


Solution

$$
\begin{aligned}
\boldsymbol{F}_P\ (\text{down}) &= m\boldsymbol{g}\sin\theta = 8 \times 10 \times \sin 30° = 40\ \text{N} \\
\boldsymbol{T}\ (\text{up}) &= 45\ \text{N} \\
\boldsymbol{F}_R &= 45 - 40 = 5\ \text{N up the incline (resultant force or net force)} \\
\boldsymbol{F} &= m\boldsymbol{a} \quad a = \boldsymbol{F}/m = 5/8 = 0.6\ \text{m s}^{-2}
\end{aligned}
$$

Note: in the above case an applied force (tension) in the rope of exactly 40 N would not cause acceleration. The carton would travel at constant velocity or remain at rest. Students often forget that in cases where there is no net force there is no acceleration but that an object can keep travelling at constant speed. In the above case you would need more than 40 N in the rope to start the carton moving up the incline from rest; but once it was moving, a force of 40 N would keep it at constant speed. Less than 40 N in the rope would cause it to slow down, stop and start to move down the incline. In all of these cases friction has been neglected. This, of course, is unrealistic and will be dealt with in Section 4.12.

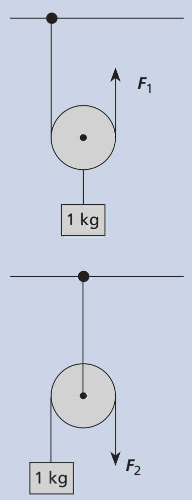
Figure 4.24



Example 3

A 20 kg object is attached by a thin cord to a 50 kg mass that hangs over a frictionless pulley at the top of a 25° incline (Figure 4.24). Calculate **(a)** the acceleration, if any, of the object; **(b)** the tension in the string.

NOVEL CHALLENGE



(a) *Which box requires the smaller force to lift it?*
(b) *Which box requires less work to raise it 1 metre?*

Solution

(a) Let m_1 be the 20 kg mass on the incline; m_2 is the 50 kg hanging mass.

$$\boldsymbol{F}\text{ (down the incline)} = \boldsymbol{F}_P = m_1\boldsymbol{g}\sin\theta = 20\times 10\times\sin 25° = 85\text{ N}$$
$$\boldsymbol{F}\text{ (up the incline)} = \boldsymbol{F}_A = m_2\boldsymbol{g} = 50\times 10 = 500\text{ N}$$
$$\boldsymbol{F}_R = \boldsymbol{F}_A - \boldsymbol{F}_P = 500 - 85 = 415\text{ N up the incline}$$
$$\boldsymbol{F}_R = m\boldsymbol{a} \quad \boldsymbol{a} = \frac{\boldsymbol{F}_R}{m_{(total)}} = \frac{415}{20+50} = 5.9\text{ m s}^{-2}\text{ up the incline}$$

(b)
$$\begin{aligned}\boldsymbol{T} &= \boldsymbol{F}_W\sin\theta + m\boldsymbol{a}\\ &= m\boldsymbol{g}\sin\theta + m\boldsymbol{a}\\ &= 20\times 10\times\sin 25° + 20\times 5.9\\ &= 203\text{ N}\end{aligned}$$

Note: some teachers prefer to use an alternative solution, which uses the tension in the rope ($\boldsymbol{T}$) and produces simultaneous equations in terms of $\boldsymbol{T}$ for each object (m_1 and m_2) separately:

For object m_1:

$$m_1\boldsymbol{a} = \boldsymbol{T} - \boldsymbol{F}_{W1}\sin\theta = \boldsymbol{T} - m_1\boldsymbol{g}\sin\theta$$
$$\boldsymbol{a} = \frac{\boldsymbol{T} - m_1\boldsymbol{g}\sin\theta}{m_1}$$

For object m_2:

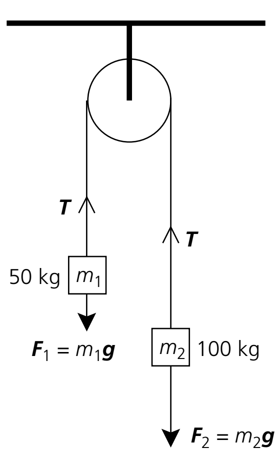
$$m_2\boldsymbol{a} = \boldsymbol{F}_{W2} - \boldsymbol{T}$$
$$\boldsymbol{a} = \frac{m_2\boldsymbol{g} - \boldsymbol{T}}{m_2}$$

Solving simultaneous equations gives $\boldsymbol{T}$ = 203 N as before.

Be guided by your teacher and/or what you may learn in maths. Both methods as you can see produce the same answer.

PULLEYS 4.11

Figure 4.25



Other questions of great interest to physicists and engineers are those about **pulleys**. Pulleys are essential components of many of the machines used in industry and in the home. They reduce friction between two surfaces by converting sliding friction into rolling friction. The word 'pulley' came to the English language in the fourteenth century from medieval Greek — *polos* meant 'pivot'. An understanding of the forces involved enables design engineers to specify sizes and strengths of materials with the knowledge that any loads imposed will be within safe working limits. Typical application examples follow.

Example

An object of mass 50 kg and another of mass 100 kg are tied to the ends of a light inextensible (non-stretching) string. The string passes over a smooth pulley (see Figure 4.25). Determine **(a)** acceleration of the system (magnitude only); **(b)** the tension in the string.

Solution

Again this problem can be solved in two ways: **1.** by developing equations involving the tension in the string on each mass separately and equating the tension ($\boldsymbol{T}$) in each equation; **2.** alternatively, the forces can be considered as a whole. Either method is suitable.

Ch04-Walding 4th 25/8/04 10:30 AM Page 94

The diagram can be rearranged into a linear form (Figure 4.26).

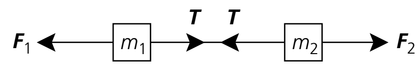


Figure 4.26

$$\begin{aligned} F_R = F_1 + F_2 = m_1 g + -m_2 g &= 50 \times 10 - 100 \times 10 = 500 \text{ N to the right} \\ F_R &= (m_1 + m_2)a \\ 500 &= 150a \\ a &= 3.3 \text{ m s}^{-1} \text{ (the 100 kg mass moves down)} \end{aligned}$$

The tension in the string can be calculated by considering just one side: on the left the tension in the string has to equal the weight of the 50 kg object (pulling down) plus the accelerating force on the 50 kg force upward:

$$T = m_1 g + m_1 a = 50 \times 10 + 50 \times 3.3 = 665 \text{ N upward on the left}$$

Note: the tension in the string on the right has to equal the same value (665 N). You can prove this by showing that the tension has to equal the weight of the 100 kg object (pulling down) *less* the accelerating force (because it is acting in the same direction as the weight):

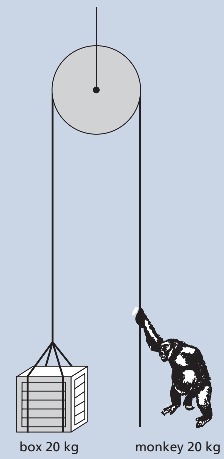
$$T = m_2 g - m_2 a = 100 \times 10 - 100 \times 3.3 = 665 \text{ N upward on the right}$$

There are several ways of approaching these pulley problems. You may find the following method more useful:

$$\begin{aligned} m_1 a &= T - m_1 g \\ m_2 a &= m_2 g - T \\ \frac{T - m_1 g}{m_1} = a &= \frac{m_2 g - T}{m_2} \\ T &= 665 \text{ N} \end{aligned}$$

NOVEL CHALLENGE

A monkey has the same mass as a box. He climbs a rope.
Who will reach the pulley first?



Questions

18 A 30 kg box of vegetables moves down a 35° frictionless incline. Find **(a)** the normal reaction; **(b)** the resultant force down the incline; **(c)** the acceleration down the incline.

19 Two blocks of masses 2 kg (A) and 3 kg (B) respectively rest on a smooth horizontal surface and are connected by a taut string of negligible mass. A force of 10 N is applied to the 3 kg mass as shown in Figure 4.27. Calculate **(a)** the tension in the string between them; **(b)** the acceleration of the system.

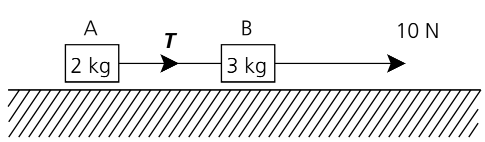
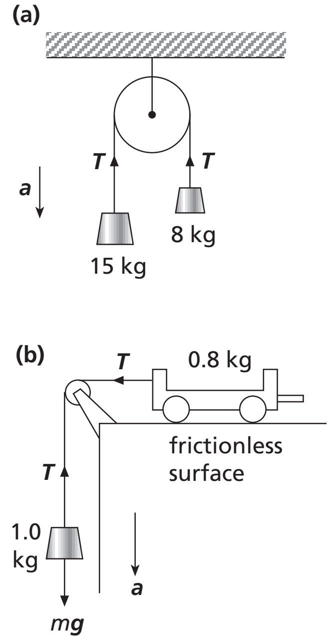


Figure 4.27
For question 19.

20 For each situation shown in Figure 4.28, find **(a)** the acceleration and **(b)** the tension in the string.

Figure 4.28
For question 20.



Ch04-Walding 4th 25/8/04 10:30 AM Page 95

FRICTION 4.12

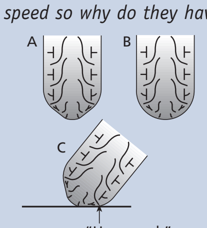
NOVEL CHALLENGE

Motorcycle tyres are of two main types — sport and touring. Sport tyres are v-shaped (A), touring are rounded (B). The figure below shows the profiles of Dunlop's mega-successful D207GP race tyre, which has been cleaning-up in supersport competitions for some time now. On the right (B) is the road version of the tyre (the D207). The unusual tread pattern is called a **cosecant groove**. In maths you may have learnt about the cosecant or cosec trigonometric function, which equals 1/sin. Dunlop scientists found that the wear on a tyre is proportional to the cosec of the lean angle (the angle of the bike to the road). Dunlop engineers incorporated cosecant grooves to minimise the tyre wear problems (see figure).

(a) Plot a graph of the lean angle (from 5° to 50° in 5° increments) against the cosec of the lean angle. *Note*: when $\theta = 30°$, cosec 30° = 1/sin 30° = 2. A computer spreadsheet may speed things up. *How does the curve compare with the groove shape in the D207GP tyre?*

(b) As riders using the v-shaped tyre go into a corner they flip the bike on to its side rather than doing a uniform lean. *Why is this? The upper limit of contact between the tyre and the road is called the 'hero line'. What do you suspect this means?*

(c) Cruising bikes like the Harley Davidson have tyres with a flat profile. *This must impair their cornering at speed so why do they have it?*



In Grade 10 you were probably asked to write about what life would be like without friction. It should have been apparent that it would be a nightmare — in fact, impossible.

- You couldn't swallow, walk, hold a pen or do your homework.
- Your clothes would unravel and your shoe laces would come undone.
- Nails and screws would be useless — your house would fall down.
- Mountains would crumble but TV cameras couldn't record it.
- Childbirth would be by caesarean because uterine pushing would be useless and forceps wouldn't grip.
- Life would last about one day before everyone died. The farewell party would be a disaster because you couldn't hold on to your cup, you couldn't swallow and you couldn't stand up. Ever been to a party like that?

Friction of course is absolutely necessary but it is also a hindrance. The search for ways of altering it has gone on for thousands of years. The search to *understand* it has only gone on for a few hundred years — only since the birth of physics.

What causes friction?

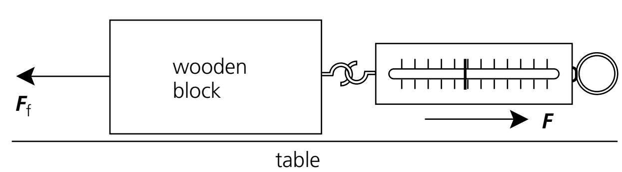
Friction is a force acting between the surface atoms of one body and those of another. If two highly polished and carefully cleaned metal surfaces are brought together in a very good vacuum, they cannot be made to slide over one another. Instead they **cold-weld** together instantly, forming a single piece of metal. Under ordinary circumstances, however, such close atom-to-atom contact is not possible. Even a highly polished metal surface is far from being flat on the atomic scale. Moreover, the surfaces of everyday objects have layers of oxides and other contaminants that reduce this cold-welding effect.

When two surfaces are placed together, only the high points touch each other. It's like a mountain range turned upside down on top of another. The actual area in contact may be only one-thousandth of the total surface area. Many points cold-weld together. When the surfaces are pulled across each other, there is a continuous breaking and remaking of the welds. Strangely enough, it is not always the softer material that gets worn away by friction. In machinery driven by rubber belts, it is often the metal pulley that wears away before the soft rubber belt. Grit becomes impregnated in the soft rubber and grinds away at the metal of the pulley. Pulleys are often replaced before the belts.

EI Activity 4.9 FRICTION IS A DRAG

1 Take a block of wood, note its mass, connect a spring balance to it and place it on the table. Gently pull on the balance horizontally and note the maximum force on the scale before it starts to move (Figure 4.29). This is called the limiting friction, starting friction or static friction.

Figure 4.29



2 Continue to pull it at constant speed across the bench and note the reading. This is called the **sliding friction**. Double check your results.

3 Add a known mass to the top of the block and repeat.

4 Repeat with additional masses, noting the spring balance reading each time.

5 Convert the spring balance reading to an equivalent sliding frictional force (F_f) in newtons.

6 Calculate the normal reaction (F_N) in newtons of the block in each case.

7 Plot F_f (y-axis) against F_N (x-axis) and comment on the relationship.

8 Which is greater, sliding friction or limiting friction?

9 Leave a block on the bench overnight and measure limiting friction the next day. Has it changed? Why?

In the above activity you should have established that F_f was directly proportional to F_N, that is, $F_f \propto F_N$. A constant of proportionality (μ) can be included in the relationship and the formula becomes:

$$F_f = \mu F_N \text{ or } \mu = \frac{F_f}{F_N}$$

The symbol μ is called the **coefficient of sliding friction** and is usually less than 1.0 but can range as high as 7 or 8. It is a ratio — it has no units.

$$\text{Coefficient of friction} = \frac{\text{frictional force}}{\text{normal contact force}}$$

Table 4.8 COEFFICIENTS OF FRICTION

SURFACES IN CONTACT	STARTING FRICTION	SLIDING FRICTION
Steel on ice (ice skates)	0.02	0.01
Teflon on teflon	0.04	0.04
Waxed skis on wet snow	0.14	0.10
Wood on Laminex	0.40	0.30
Glass on glass	0.94	0.40
Steel on steel	0.78	0.42
Wood on wood	0.62	0.48
Rubber tyre on wet road	0.70	0.50
Rubber tyre on dry road	0.90	0.70
Steel on lead	0.95	0.95
Foam rubber on foam rubber	8.0	7.0

Rolling friction

If you've ever tried to slide down a grassy slope on a piece of cardboard you know that you'd go faster on a go-kart with wheels. Wheels and bearings replace sliding friction with rolling friction and this is much lower. Imagine a car parked on a slope. The friction between the brake pads and the wheels prevents the car rolling. Without rolling being available, the tyres would have to slide, and sliding friction between the road and the tyres is great enough to keep the car from moving downhill.

Properties of friction

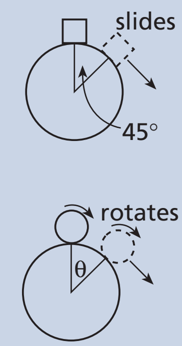
- Starting friction is always more than sliding friction.
- Rolling friction is always less than sliding friction.
- Friction is always in the direction opposing motion.
- Sliding friction is practically independent of surface area.
- $F_f = \mu F_N$

NOVEL CHALLENGE

A piece of pine dowel is placed in an electric drill and rotated against a piece of hardwood. *Which will catch on fire first? Try it.*

NOVEL CHALLENGE

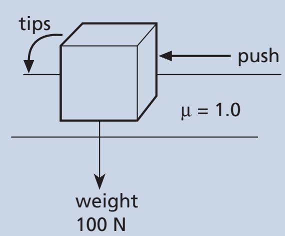
A block of wood is placed on top of a smooth cylinder (see figure).



When the block gets to 45° it slides off. *If a ball was used instead of the block and allowed to roll down the surface, would it fall off at a bigger or smaller angle than 45°?*

NOVEL CHALLENGE

A cubical block of mass 10 kg is placed on a surface where μ=1.0. *Where would you need to push on the block so that it was on the verge of tipping and sliding?*



Reducing friction

About 40% of the fuel used in a car is to overcome friction. This friction may be between the body and the air or between the mechanical components (pistons and cylinders, bearings or gears). Obviously, motor engineers have attempted to reduce those types that are a hindrance by either streamlining the body design or by use of **lubricants** between moving mechanical parts. Lubricants such as the one pictured in the photo may contain teflon (polytetrafluoroethylene — PTFE), which is added to the engine oil lubricant. Another friction modifier is the oil additive molybdenum disulfide (MoS_2). If the oil level in a car gets too low, friction increases, which produces more heat and can damage the engine.

Friction modifiers are not new. The Egyptians built enormous pyramids 4500 years ago using huge stone blocks that were difficult to move by sliding. To move them they used log rollers underneath and took advantage of the fact that rolling friction is lower than sliding friction.

Friction can also be reduced by choosing suitable materials for the contact surfaces; for example, a steel shaft should rotate in a bronze or white-metal bearing. In hi-fi equipment, bearings usually have a nylon 'bush' — a low friction insert inside the bearing. Sometimes bearings are impregnated with graphite granules. Another common bearing is the ball or roller bearing. They can reduce friction by up to 100 times as they convert sliding friction to rolling friction (Figure 4.30). Strangely enough, under load, they work even better.

Human joints are lubricated by synovial fluid (Figure 4.31) between layers of cartilage lining the joint. When you move the efficiency of this lubricant increases. What a piece of work is man (and woman).

Not all lubricants are liquids. For example, hovercrafts float on a cushion of air. A linear air track in a physics laboratory works in the same way. Air can be an excellent lubricant.

Photo 4.4
Friction modifier for a car.

Figure 4.30
Bearings.

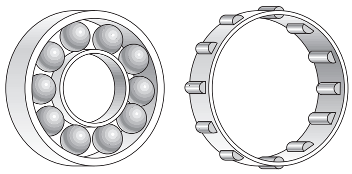
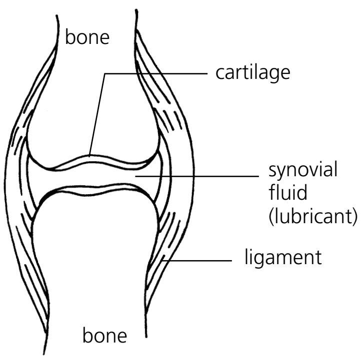


Figure 4.31
A typical joint in the human body.



EI Activity 4.10 VEHICLE FRICTION

To carry out this activity you need the help of a licensed driver.

Part A: Rolling friction

1 The coefficient of rolling friction of a car can be found by pushing a car along a flat, horizontal surface. Have a person sit in the driver's seat ready to apply the brakes. The car should be put in neutral gear and the engine turned off.

2 Put a set of bathroom scales on the bumper bar and note the reading (in kg) needed to keep the car rolling at a slow constant speed. It will probably be between 10 and 20 kg.

3 Stop the car and push from the opposite direction. Average the results and convert to newtons.

4 Look in the owner's manual to find out the mass of the car.

5 Use the friction formula to calculate the coefficient of rolling friction.

6 What is the source of this rolling resistance?

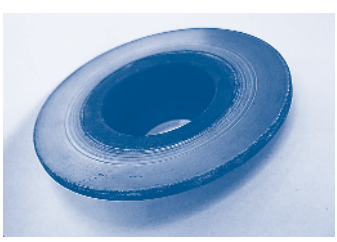
Part B: Engine friction

1 Repeat the above experiment, still with the engine off, but put the car in top gear and let the clutch out. The bathroom scale reading will then be the sum of engine friction and rolling friction. You should get values around 50 kg — a typical value for a 1989 Honda Civic.

2 Subtract this value from the rolling friction (Part A) to get engine drag. Convert to newtons.

3 Comment on the source of engine drag.

Photo 4.5
Disc rotor from a car. When you put your foot on the brakes, pads are pressed against this rotor to slow the car down. The grooves in this one show that the pads wore down to bare metal — dangerous but can be machined flat again.



Ch04-Walding 4th 25/8/04 10:30 AM Page 98

Part C: Air drag

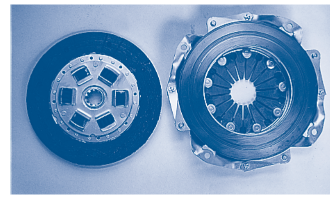
This part requires two people inside the car — a licensed person to give full attention to driving and the other to time and record results.

1. Find a straight, smooth, level road with little traffic at the time. Be careful not to interfere with other vehicles and don't do it at night.
2. Drive at a steady 40 km/h in top gear, push the clutch in and measure the time it takes to slow down to 30 km/h. This is called the 'coast-down' time. We found times of 24 s were average.
3. Use a formula to calculate the deceleration.
4. Use Newton's second law to compute the total frictional force.
5. Subtract rolling resistance and engine resistance to give air drag.

Summary: Total friction = rolling resistance + engine drag + air drag.

Express the three types of friction as a percentage of the total friction.
Comment on your results and comment further on how the activity could be improved or extended.

Photo 4.6
Clutch assembly out of a manual car. The clutch plate (on the left) provides a smooth coupling between the flywheel (not shown) and the pressure plate (on the right). When the clutch fingers in the centre of the assembly are pressed in, the pressure plate moves away so the gears can be changed.



NOVEL CHALLENGE

In an extreme skiing competition in Alaska in 1995, a New Zealand woman tumbled 400 m down a 50° slope and ended up with severe head trauma. (The slope used in the Olympics is 35°.) How fast would she be going if the coefficient of friction was **(a)** zero, **(b)** 0.10?

Increasing friction

Although friction can be a nuisance (as seen in the previous activity), it is also necessary and may even need to be increased. Car tyre tread patterns are designed to increase rather than decrease friction although other factors come into the design as well. Factors such as rate of wear, flexibility, strength, dispersion of road moisture and cost are just as important. The same considerations go into the design of running shoe soles. Some questions at the end of this chapter deal with design of consumer products.

4.13 EXAMINING FRICTION

There are three main situations that you should be familiar with when examining friction between surfaces. They can be best thought of as: horizontal applied force; angled applied force and inclined plane. Important: the term 'at constant speed' means that the applied force in the direction of motion is equal to the friction. If it is greater than friction the object will accelerate. It can never be less than friction!

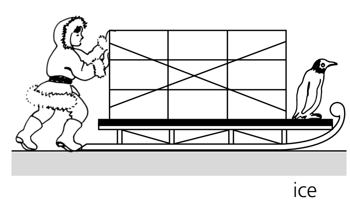
Horizontal applied force

This is the simplest case. The applied force and friction are in a line and oppose each other.

Example 1

What horizontal force has to be applied to a 50 kg sled on an ice surface ($\mu = 0.3$) as shown in Figure 4.32 to make it move at constant speed?

Figure 4.32

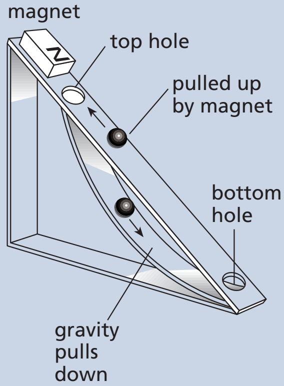


Solution

$$\boldsymbol{F}_W = m\boldsymbol{g} = 50 \times 10 = 500 \text{ N down}$$
$$\boldsymbol{F}_N = 500 \text{ N up}$$
$$\boldsymbol{F}_f = \mu \boldsymbol{F}_N = 0.3 \times 500 = 150 \text{ N}$$
$$\boldsymbol{F}_A = 150 \text{ N to right (constant speed)}$$

INVESTIGATING

Many people claim to have made perpetual motion machines. Find a description of one and say why it couldn't work. Would the one shown here work?



Example 2

A spring balance reads 300 g as it is used horizontally to drag a 750 g block of wood along a laboratory bench. Calculate the coefficient of friction.

Solution

$$F_N = 0.75 \times 10 = 7.5 \text{ N up}$$

$$F_A = F_f = 0.3 \times 10 = 3 \text{ N}$$

$$\mu = \frac{F_f}{F_N} = \frac{3}{7.5} = 0.4$$

Questions

21 A butcher pulls on a freshly cleaned 40 kg side of beef with a horizontal force of 220 N and it slides across the boning table at constant speed. Calculate the coefficient of friction.

22 A horizontal steel cable is used to drag a bucket filled with coal along the ground at constant speed. If the mass of the bucket and coal is 6.1 t and the coefficient of friction is 0.58, calculate the tension in the cable.

Angled applied force

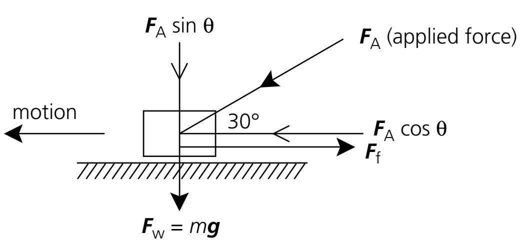
The complication with this case is that the applied force has to be resolved into two components at right angles. The vertical component will change the normal reaction whereas the horizontal component will be responsible for motion along the surface.

Angled forces can be of two types: pushing or pulling.

Example 1: Pushing

A child is pushing a 25 kg box of toys along a carpeted floor at constant speed.

Figure 4.33



If the child's arms make an angle of 30° to the horizontal and she pushes with a force of 100 N, calculate **(a)** the vertical component of the applied force; **(b)** the horizontal com ponent of the applied force; **(c)** the normal reaction; **(d)** the force of friction; **(e)** the coefficient of friction.

NOVEL CHALLENGE

A wooden block is put on an electronic balance, which then reads 100 g (weight = 1 N). A string is attached to the block and a spring balance is attached to the string. A force of 0.4 N is applied at an angle of 45°. *What will the balance read? Try it to check.*

Solution

(a) Vertical component = $F_A \sin\theta = 100 \times \sin 30° = 50$ N.

(b) Horizontal component = $F_A \cos\theta = 100 \times \cos 30° = 87$ N.

(c) Normal reaction = weight + vertical component pushing down $= mg + F_A \sin\theta$
$= 250 + 50 = 300$ N.

(d) Friction = horizontal component (because of constant speed) = 87 N.

(e) $\mu = \frac{F_f}{F_N} = \frac{87 \text{ N}}{300 \text{ N}} = 0.3.$

Note: if the applied force is a pulling force at an angle then the normal reaction is the weight *minus* the vertical component. See the example that follows.

Example 2: Pulling

A boy drags a 15 kg box across a concrete floor at constant speed by means of a cord at an angle of 20° to the floor. If the force applied is 100 N, calculate the coefficient of friction.

Solution

- Vertical component = 100 sin 20° = 34 N.
- Normal reaction = weight − vertical component = 15 × 10 − 34 = 116 N.
- Horizontal component = 100 cos 20° = 94 N.
- Friction ($\boldsymbol{F}_f$) = 94 N.

$$\mu = \frac{\boldsymbol{F}_f}{\boldsymbol{F}_N} = \frac{94\ \text{N}}{116\ \text{N}} = 0.8$$

Example 3: Pulling

What force is needed to be applied to the handle of a 10 kg sled to drag it at constant speed across a horizontal sandy beach? The handle of the sled is at an angle of 30° to the horizontal and the coefficient of friction is 0.5.

Solution

$$\begin{aligned}
\boldsymbol{F}_f &= \mu \boldsymbol{F}_N \\
\boldsymbol{F}_A \cos\theta &= \mu \boldsymbol{F}_N \\
&= \mu(\boldsymbol{F}_W - \boldsymbol{F}_A \sin\theta) \\
\boldsymbol{F}_A \times \cos 30° &= 0.5\ (10 \times 10 - \boldsymbol{F}_A \sin 30°) \\
0.87\ \boldsymbol{F}_A &= 50 - 0.5 \times \boldsymbol{F}_A \times 0.5 \\
0.87\ \boldsymbol{F}_A &= 50 - 0.25\ \boldsymbol{F}_A \\
1.12\ \boldsymbol{F}_A &= 50 \\
\boldsymbol{F}_A &= 45\ \text{N}
\end{aligned}$$

NOVEL CHALLENGE

Here's another Fermi question: What force is required to break a blade of grass by pulling at each end?

NOVEL CHALLENGE

A 10 kg block and a 5 kg block sit side-by-side on a benchtop, just touching each other. *Which is the easier way to push them along: using a 100 N force at 30° to the top edge of the 10 kg block or to the top edge of the 5 kg block? Assume the coefficient of friction is 0.5 for these surfaces.*

Questions

23 A worker drags an 80 kg crate across a factory floor at constant speed by pulling on a rope tied to the crate. The worker exerts a force of 350 N on the rope, which is inclined at 38° to the horizontal. Calculate **(a)** the frictional force; **(b)** the normal reaction; **(c)** μ.

24 A lawn roller of mass 200 kg is being pushed at a constant speed by the handle, which is inclined at 40° to the horizontal. If the coefficient of friction is 0.12, calculate the force being applied to the handle by the pusher.

Inclined planes

Again, with this sort of problem the normal reaction is not equal to the weight but the component of the weight at right angles to the incline (= $\boldsymbol{F}_W \cos\theta$). If the object slides down the plane then the friction acts up the plane, but if the object is dragged *up* the plane, then the friction acts down the plane. There are two situations: sliding down and sliding up.

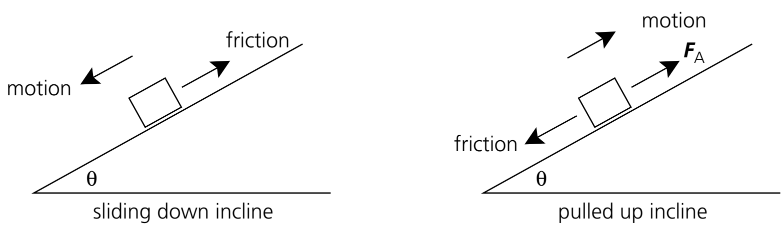


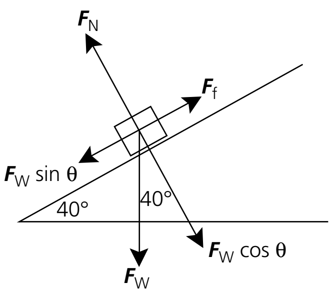
Figure 4.34

Example 1: Sliding down

A 14 kg toolbox is placed on a plank of wood. When one end of the plank is raised, the toolbox begins to slide down the incline at a uniform speed when the angle reaches 40°. Calculate the coefficient of friction.

Solution (See Figure 4.35)

Figure 4.35



$$F_f = F_W \sin\theta$$

$$F_N = F_W \cos\theta$$

$$\mu = \frac{F_f}{F_N} = \frac{F_W \sin\theta}{F_W \cos\theta} = \frac{14 \times 10 \times \sin 40°}{14 \times 10 \times \cos 40°} = \frac{90}{107} = 0.84$$

Note: the previous equation can be simplified. The ratio $\frac{\sin\theta}{\cos\theta} = \tan\theta$, so the coefficient of friction is merely the tan of the angle at which an object slides at constant speed down an incline (tan 40° = 0.84). This is a simple method for finding μ in the classroom. Note also that the value is independent of mass. The mass term cancels out.

Example 2: Sliding up

A bricklayer's apprentice is dragging a tray of bricks up a 50° inclined plank by pulling on a rope attached to the tray. The rope is parallel to the plank. If the load of bricks has a mass of 40 kg and the coefficient of friction is 0.6, calculate the force in the rope.

Solution

As the bricks are moving up the incline, friction acts down the incline. The applied force F_A in the rope equals the component of the weight down the incline plus the frictional force:

$$\begin{aligned} F_A &= F_W \sin\theta + F_f = F_W \sin\theta + \mu F_N \\ &= 40 \times 10 \times \sin 50° + 0.6 \times 40 \times 10 \times \cos 50° \\ &= 306 + 154 \\ &= 460 \text{ N} \end{aligned}$$

ROAD ACCIDENT INVESTIGATION 4.14

NOVEL CHALLENGE

With the automatic gearbox in 'drive' a Toyota 1200 kg RAV 4 will remain stationary facing uphill on a 5° slope. What would its initial acceleration be on the flat (assuming the driver's foot was not on the accelerator)? We got 0.87 m s^{-2}.

Traffic accident investigation is not just about examining the damage to cars and people; it also involves applying physics principles to determine how the accident happened. One of the key pieces of evidence comes from the tyre skid marks.

The speed of a vehicle prior to skidding to a halt can be deduced from the skid mark length and the coefficient of friction between the tyre and the road surface.

Horizontal surface

Suppose a vehicle of mass m travels on a level road at a speed u prior to skidding. During the skid, the only horizontal force acting on the car is friction: $F_f = \mu F_N = \mu mg$. As this is the net force, the deceleration of the car is given by Newton's second law: $F_{net} = ma$ and is equal to F_f. Hence, combining the two equations gives us $\mu mg = ma$. This cancels down to: $a = \mu g$.

If we use our kinematic formula $v^2 = u^2 + 2as$, and assume the vehicle came to rest (i.e. $v = 0$), acceleration a is equal to $u^2/2s$, hence $u^2/2s = \mu g$, or $u = \sqrt{2\mu gs}$

Note that the deceleration is independent of the mass of the vehicle.

Example

Calculate the initial speed of a 1500 kg car that skidded 40 m to a halt on a level road where the coefficient of friction was 0.65.

Solution

$$u = \sqrt{2\mu \boldsymbol{g} s} = \sqrt{2 \times 0.65 \times 10 \times 40} = 22.8 \text{ m s}^{-1} \text{ (82.1 km h}^{-1}\text{)}$$

Different surfaces

When a car skids across different surfaces, the starting speed can be calculated using:

$$u = \sqrt{2\mu_1 \boldsymbol{g} s_1 + 2\mu_2 \boldsymbol{g} s_2 + 2\mu_3 \boldsymbol{g} s_3 + \ldots}$$

where μ_1 is the coefficient of friction on surface 1, s_1 is the length of this surface, and so on.

Example

A car skids with all four wheels locked and leaves skid marks of 19.3 m on dry bitumen, 5.6 m on concrete pavement and 15.4 m on grass. The μ values are 0.74, 0.82 and 0.46 respectively. Calculate the speed of the car at the start of the skidding.

Solution

$$u = \sqrt{2\mu_1 \boldsymbol{g} s_1 + 2\mu_2 \boldsymbol{g} s_2 + 2\mu_3 \boldsymbol{g} s_3 + \ldots}$$

$$u = \sqrt{2 \times 0.74 \times 10 \times 19.3 + 2 \times 0.82 \times 10 \times 5.6 + 2 \times 0.46 \times 10 \times 15.4}$$

$$u = 22.8 \text{ m s}^{-1} \text{ (82.0 km h}^{-1}\text{)}$$

Inclined surfaces

When the skid marks are on a road that has a slope of θ degrees, the simplest way to calculate the starting speed is to use an 'effective' value for μ:

- **up slope:** $\mu_{us} = \mu + \sin\theta$
- **down slope:** $\mu_{ds} = \mu - \sin\theta$

Questions

25 Calculate the skid-to-stop distance of a car travelling on a road at 100 km h^{-1} with a coefficient of friction of 0.68.

26 A car skidded to a stop, producing skid marks of 5.6 m on bitumen (μ = 0.61) and 3.2 m on concrete (μ = 0.79). Estimate the speed of the car prior to skidding.

27 A car skidded to a halt down a road of gradient of 13.8°, producing skid marks of 17.4 m. The coefficient of friction was 0.73. Calculate the speed of the car prior to skidding.

Activity 4.11 SPEED AND STOPPING DISTANCE

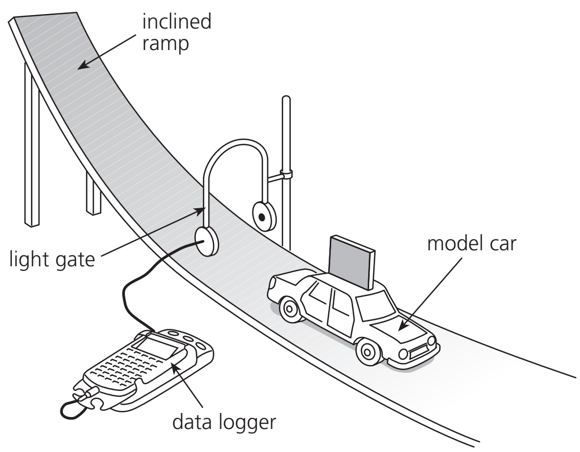
EI

Does it take twice as long to stop a car if its speed is doubled? Let's find out.

A light gate attached to a computer timer is useful for this experiment. The TI graphing calculator and a CBL with a light gate work fine. You will need the 'Data Gate' program.

1. Glue a small piece of card 5 cm long to the top of a small toy car.
2. Construct a ramp and place the light gate at the bottom. (See Figure 4.36.)
3. Allow the car to run down the incline, through the light gate where its speed is measured, and let it run across the floor to a halt.
4. The light gate will measure the time interval for the 5 cm card to pass through. (Calculate $\boldsymbol{v}$ by dividing 10 cm by the time taken.)
5. Measure the distance the car moves before stopping.
6. Repeat with different height inclines. What did you conclude?
7. Plot speed (x-axis) against stopping distance.

Figure 4.36
Using a data-logger to measure car speeds.



Questions

28 In Figure 4.37, the block weighs 50 N and the applied force $\boldsymbol{F}_A$ is 20 N. Calculate the normal force in each case.

Figure 4.37
For question 28.

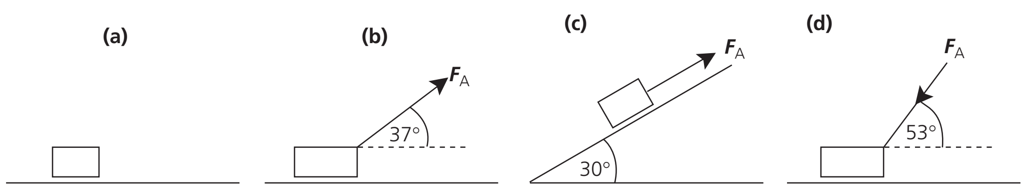
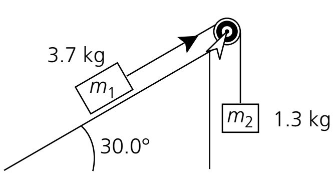


Figure 4.38
For question 30.



29 When a laboratory inclined plane is raised at one end, a small wooden block of mass 80 g slides down the slope at constant speed when the angle reaches 25°. Calculate the coefficient of sliding friction.

30 A block of mass m_1 = 3.70 kg rests on a 30° incline and is held stationary by a mass (m_2) of 1.3 kg hanging vertically over a frictionless pulley at the top (Figure 4.38). Calculate the value of μ.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***31** Calculate the resultant force when the following forces act on the same object:
(a) 24 N north, 18 N south, 19 N north. **(b)** 6.5 N down, 9.2 N up and 7.4 N up.
(c) 55 N north, 35 N west, 65 N south. **(d)** 6 N west, 5 N east, 3 N north, 1.5 N south.

***32** The density of aluminium is 2.7 g cm^{-3}. Calculate the mass of a sheet of the metal 25 cm long by 10.0 cm wide and 3.0 mm thick.

***33** Which of Newton's laws of motion best describes the following situations?
(a) When a car accelerates the occupants feel pressed back into their seats.
(b) When you turn on the garden hose it moves around if it is not held.
(c) A nailfile, when thrown vertically, undergoes uniform deceleration.
(d) A sandwich, when dropped off a cliff, travels straight down.

***34** A handbag is sliding down a 30° incline at constant speed. Which one of the following relationships is false about the situation?
(a) $\boldsymbol{F}_W \sin 30° > \boldsymbol{F}_f$. **(b)** $\boldsymbol{F}_N = m\boldsymbol{g} \cos 30°$. **(c)** $\boldsymbol{F}_f = \mu \boldsymbol{F}_N$. **(d)** $\boldsymbol{F}_W = m\boldsymbol{g}$.

***35** What force is necessary to uniformly accelerate:
(a) a 6.4 kg mass at 2.4 m s^{-2} east; **(b)** a 0.16 kg mass from rest to 2 m s^{-1} in 3 seconds; **(c)** an object weighing 25 N at 9.8 m s^{-2}; **(d)** a 0.50 kg object from rest to 5.0 m s^{-1} over 4.0 metres; **(e)** a 75 kg object from 40 m s^{-1} to 60 m s^{-1} in 5 milliseconds?

***36** A wooden box of bolts has a mass of 250 kg and requires a horizontal force of 2100 N to slide it along a horizontal wooden surface at a constant speed.
(a) Calculate the coefficient of friction.
(b) If the box were to be kept moving constantly at twice this speed what force would be needed to maintain this constant speed?

****37** In a TV tube an electron experiences an unbalanced force of 8.0 femtonewtons over a distance of 20 mm. (Look in the Appendix for the mass of an electron and the meaning of the prefix 'femto'.)
(a) Calculate the electron's acceleration.
(b) Calculate the electron's speed at the end of the 20 mm (starting from rest).

****38** The graph shown in Figure 4.39 is an acceleration–force graph for an experiment with a loaded cart pulled by rubber bands:
(a) What does the intercept of the graph with the force axis measure?
(b) What acceleration would an applied 2.5 N force produce?
(c) What net force would produce an acceleration of 2.0 m s^{-2} and what applied force is this equal to?
(d) What is the mass of the loaded cart?

****39** A bicycle and rider have a combined mass of 65 kg. When travelling at 5.0 m s^{-1} on a level road, the cyclist ceases to pedal and comes to rest in 255 m. What frictional forces must have been acting on the cyclist?

***40** Consult a table of densities to find out:
(a) which, if any, metals are lighter than water, and list their densities;
(b) whether any liquids are heavier than mercury at room temperature.

***41** You are standing on the edge of a frozen pond where friction is negligible. In the centre is a blue circle 1.0 m in diameter. There is a prize of 4.0 L of icecream if anyone can apply all three of Newton's laws of motion to get there. How could you do it?

****42** If friction is independent of surface area, why is it that racing car drivers use wide tyres to improve the grip?

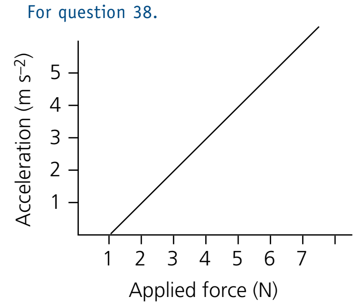
TEST YOUR UNDERSTANDING

When the upward force of air resistance on a parachutist equals his weight (down), shouldn't he be stationary? Explain.

NOVEL CHALLENGE

In 1999, a 19-year-old Gold Coast man tried to stop a car that was starting to roll down a driveway slope. He didn't succeed and was run over. Why couldn't he stop the car — after all, it was only on a 10° slope? How much force can you push with? Try this. Have someone hold a set of bathroom scales against the wall at about hip height while you push the scales with your hands as hard as you can. You'll probably only push to a scale reading of 40 kg (400 N). Calculate the maximum angle of an incline that you could stop a 1500 kg car from rolling down. Surprising, huh?

Figure 4.39
For question 38.



TEST YOUR UNDERSTANDING

(Answer true or false)
- Forces are needed for motion with constant velocity.
- Objects stop moving when the force is removed.
- Inertia is the force that keeps things in motion.
- The normal force on an object always equals the weight.
- Objects thrown in the air start to fall when they run out of force.

NOVEL CHALLENGE

I live at the top of a road that has a 5° downhill slope. When I let my car roll down the slope it reaches 25 km h^{-1} by the time it gets to the bottom, 400 m away. What frictional force must be acting?

****43** Is it better to have high friction between a car tyre and the ground to get good grip or better to have low friction to reduce drag forces? Examine critically.

****44** During an experiment, a linear air track glider was subjected to a single force whose magnitude could be varied. Assume the friction was negligible. The acceleration from various forces was measured and the results tabled as shown below:

Force (N)	5	10	15	20	25
Acceleration (m s^{-2})	1.5	3.0	4.5	6.0	7.5

(a) Draw a graph.
(b) What is the mass of the trolley?
(c) If friction was present how would the shape of the graph have changed?

****45** An experiment was conducted to find the relationship between force, mass and acceleration. A stretched rubber band was used to provide constant force on a trolley to which different masses were added (see Figure 4.40). The trolley was released from rest and timed to move 50.0 cm. A student made the following notes:

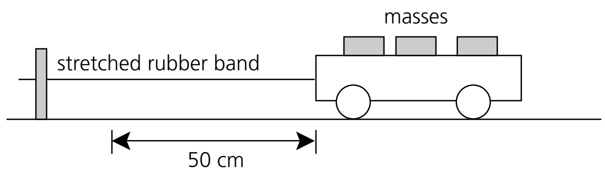


Figure 4.40
For question 45.

'When the trolley had no masses on it, it took 2.00 s. With 300 g added it took 2.39 s and with 900 g added it took 3.02 s. I measured the mass of the trolley by itself and it was 700 g.'

(a) Draw up a data table to show the results.
(b) Calculate the acceleration for each trial and add this to the table.
(c) Draw a graph of total mass (x-axis) versus acceleration.
(d) What relationship is suggested by this graph?
(e) Use Newton's second law of motion formula to calculate the force provided by the rubber band in each case and add to the table.
(f) The student was supposed to have measured the time with 600 g added but forgot. **(i)** What acceleration would the student have calculated? **(ii)** How long would the trolley have taken to cover the 50 cm in this case?
(g) Name one factor that would have had to remain constant during the experiment.

NOVEL CHALLENGE

A lawyer emailed me (RW) wanting to know what was meant when someone was 'in an accident and suffered high-g forces'. I told him in a few paragraphs. How would you explain it? By the way, he never offered to pay, but I bet he charged his client. Hmmm!

****46** Figure 4.41 has been taken from *Scientific American*, December 1994. It shows the fuel needs of the various stages of a car's propulsion system. (One US gallon equals 4.0 L.)

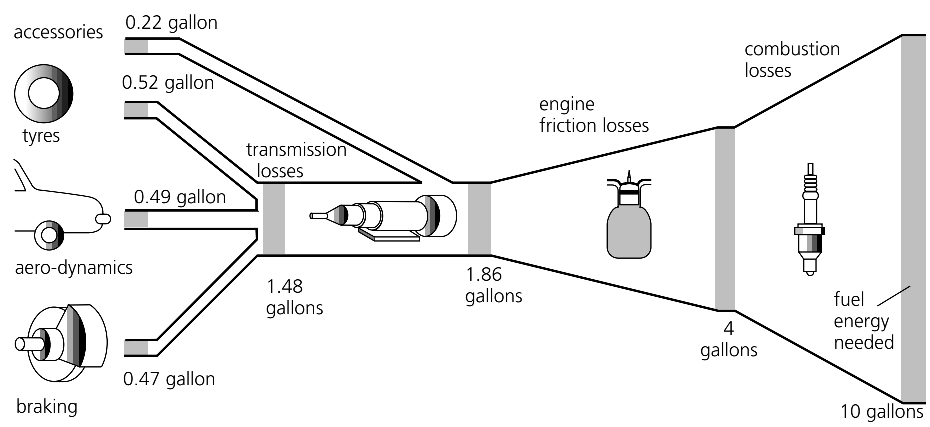


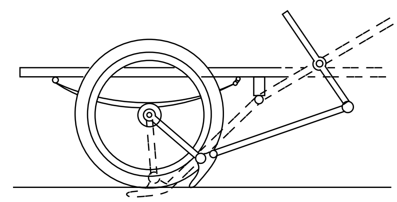
Figure 4.41
Energy losses in a typical automobile.

Ch04-Walding 4th 25/8/04 10:30 AM Page 106

(a) Make a table to show the percentage of fuel use for each stage.
(b) State which stage the following modifications would affect and say whether it would increase or decrease fuel consumption: **(i)** PTFE oil additive; **(ii)** proper engine tuning; **(iii)** low gearbox oil level; **(iv)** glovebox light that stayed on all the time; **(v)** sleeker body shape (of car, not driver); **(vi)** out-of-round disk brake rotors; **(vii)** hood rack and surfboard.
(c) If two cars were identical except that one had a six cylinder engine and the other a four cylinder engine (like the two VB Commodore models of the early 1980s), which stage would be affected (if any)?

****47** A British engineer, Mr Ralph Jackson, was awarded Patent No. 858 in 1901 for his 'Brake for Wheeled Vehicles' (Figure 4.42).
In his patent application he said that 'by a system of levers, the wheel may be raised from contact with the ground'. Explain the physical principles of his device and some good and bad points about it.

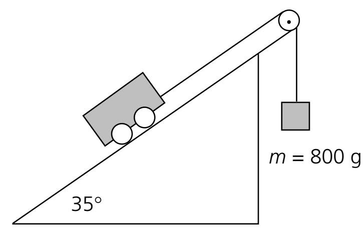
Figure 4.42
Brake for wheeled vehicles.



Extension — complex, challenging and novel

*****48** Consider the system shown in Figure 4.43. The trolley has a mass of 1000 g and is stationary when placed on a slope of 35° under the conditions shown.
(a) Determine the frictional forces acting in this system. **(b)** Calculate μ.

Figure 4.43
For question 48.



*****49** A sphere of mass 0.3 g is suspended from a 30.0 cm cord. A steady horizontal breeze pushes the sphere so that it makes an angle of 37° with the vertical. Find the magnitude of the wind force and the tension in the cord.

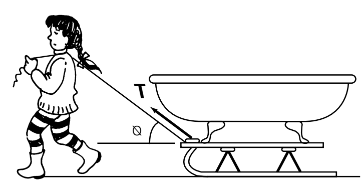
*****50** A cable used to pull mine cars vertically to the pit head has a breaking strain of 3×10^4 N. If the mine shaft is 500 m deep and a full mine car has a mass of 2.5 t, calculate: **(a)** the maximum acceleration the car can attain without breaking the cable; **(b)** the shortest time in which the car can be pulled from rest to the surface in the event of an accident.

*****51** A toboggan of mass 1000 kg starts to move down a 30° slope at an amusement park. In addition to the friction between the runners and the track (μ = 0.2) there is air resistance that has been shown to be equal to 500 N + 80 × (number of people in the toboggan) N. Which would get to the bottom of the slope first — a toboggan with one person of mass 60 kg or a toboggan with two passengers, each of mass 60 kg? Show your working.

*****52** If a car's wheels are 'locked' during emergency braking, the car slides along the road leaving bits of ripped-off tyre and small melted sections of road from the skid marks that reveal the cold-welding during the slide. The record for the longest skid marks on a public road was set in 1960 by a Jaguar on the M1 in England. The marks were 290 m long. Assuming that the coefficient of friction was 0.60, how fast was the car going when the wheels were locked?

*****53** A woman pulls a sled carrying a bath tub along a horizontal surface at constant speed (Figure 4.44). If the mass of sled and bathtub was 75 kg and μ was 0.10 and the angle θ as shown was 42°, calculate the tension in the rope.

Figure 4.44
For question 53.



*****54** A 1500 kg sled is coasting at 20 m s^{-1} on ice where friction is negligible. Suddenly it hits a 22.0 m long rough patch used for ice cricket, which creates a frictional force of 6×10^3 N. With what velocity does the sled leave the end of the rough patch?

*****55** A falling cat reaches a terminal speed of 60 km h^{-1} while it has its legs and head tucked in. When it stretches out its cross sectional area (A) doubles. Calculate its new terminal speed. Refer to Activity 4.5 (page 89) for the appropriate formula.

*****56** An electron is projected horizontally at a speed of 1.2 megametres per second into an electric field that exerts a vertical force of 450 attonewtons on it. The mass of an electron is 9.11×10^{-31} kg. Determine the vertical distance the electron is deflected during the time it has moved 30 mm horizontally. Atto (a) = 10^{-18}.

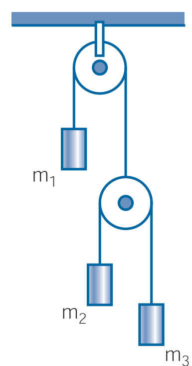
*****57** A crate of tiles of mass m_1 = 14 kg moves up a 30° incline at constant speed when pulled by a crate of cement of equal mass. The crates are connected by a taut, massless cord over a frictionless, massless pulley. Calculate the frictional force and the value of μ.

*****58** A van skidded to a halt up a road which had a slope of 7° and ran into the back of a parked car. The van produced skid marks of 13.7 m on a surface with a μ of 0.71. From the damage to the parked car, police estimated that the van's impact speed was 25 km h^{-1}. Estimate the speed prior to skidding.

*****59** In an accident, a car skidded 9.6 m over bitumen (μ = 0.66) and 2.6 m on concrete (μ = 0.76) before smashing into a fire hydrant. Police estimated the crash speed to be 30 km h^{-1}. Estimate the speed of the car prior to skidding.

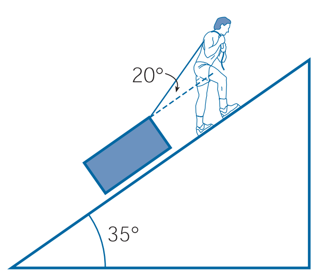
*****60** A mass m_1 hangs over a frictionless pulley and attached to the other end is another frictionless pulley with masses m_2 and m_3 arranged as shown in Figure 4.45. Calculate the acceleration of the three masses and the tension in the strings.

Figure 4.45



*****61** A man is hauling a box of mass 100 kg up a 35° incline by a rope attached to the top of the box (Figure 4.46). If the rope makes an angle of 20° to the incline and the coefficient of friction between the box and the incline is 0.65, calculate the force applied by the man to keep the box moving at constant speed.

Figure 4.46



CHAPTER 05

Projectile, Circular and Periodic Motion

5.1 MOTION IN TWO DIMENSIONS

The previous chapters have considered motion mainly in a straight line. This is called **recti-linear** motion (Latin *rectus* = 'straight' and *linea* = 'line'). This chapter will be looking at motion in two dimensions, that is, **curvilinear** motion.

Projectiles from cannons, a shotput, throwing a cricket ball, motorcyclists jumping rows of cars; and ballet dancers all involve curvilinear motion.

But there are facts and fallacies about such motion:

- Before Galileo, universities taught that when a cannon ball ran out of 'impetus' it would stop in its path and fall vertically to Earth. That's not true, is it?
- Soldiers in war have often reported that enemy bullets fired from miles away fell vertically in to their trenches. How can that be true?
- In the Olympic Hammer Throw, the hammer continues in a circular path for a fraction of a second after it is let go. True or false?
- Bombs and bullets fired at 45° have the greatest range. Well, cricket balls do; so should bullets.
- A pendulum will swing forever in a vacuum because air resistance is nil. True or false?

To make sense of these ideas, it helps if you have first-hand knowledge of some curvilinear motions.

NEI Activity 5.1 THINGS THAT DON'T GO IN STRAIGHT LINES

1 Watch a microwave oven in operation.

(a) Does the carousel rotate clockwise or anticlockwise? Does everyone else in the class get the same result?

(b) Measure the 'period' of the carousel. This is the time for one complete revolution. Time the carousel for five turns to get better accuracy. Is 12 seconds about the class average?

2 If you have a **CD player** and still have the manual, look up the rotation speed of the disk. Is it constant or is there a range of speeds?

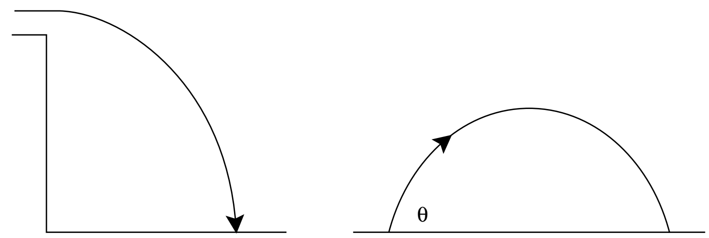
3 **Billiard** players talk about putting 'English' on the ball. What does that mean?

4 The **javelin** design was changed in 1998 so that it couldn't be thrown as far. Consult the *Guinness Book of Records* to find out how this was achieved and by approximately how much its range was reduced.

PROJECTILES 5.2

Good examples of projectiles are 1. a rock thrown straight out from the top of a cliff; 2. a cricket ball thrown across a field. (See Figure 5.1.) The word **projectile** comes from the Latin *jacere* meaning 'to throw' and *pro* meaning 'forward'. Projectile motion can be separated into two components — a **vertical** (up and down) motion and a **horizontal** motion. The vertical motion is the same as discussed in Chapter 2 — the ball is under the influence of gravity and accelerates at −10 m s^{-2} directed downward (the negative direction). In the horizontal direction, there are no net forces acting on the object so the velocity is constant. In all cases we are assuming air resistance is negligible. If you are to ever take air resistance into account in a problem you will be specifically told to do so. The path of a moving object is called its **trajectory** (Latin *trajectus* = 'crossing' or 'passage').

Figure 5.1

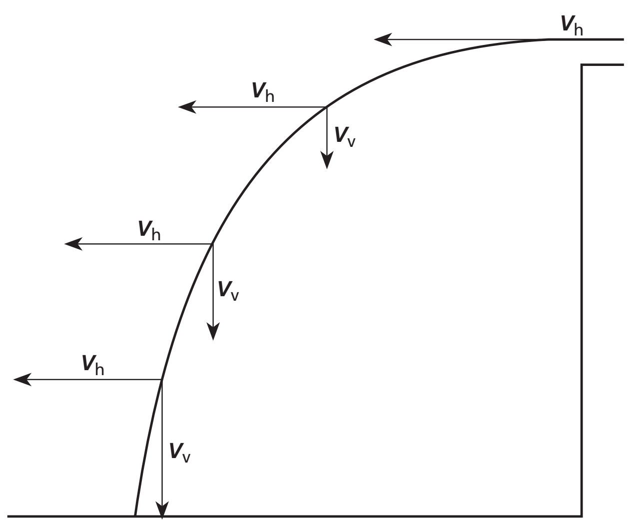


Note: in all examples that follow, the positive direction is upward and the negative direction is downward. You may choose a different convention in your problem-solving. It's up to you and your teacher.

Horizontal projection

This is the example of the rock thrown off the cliff. In this case the value of $\boldsymbol{v}_h$ equals the initial horizontal velocity ($\boldsymbol{u}_h$), which remains constant. The vertical velocity starts at zero ($\boldsymbol{u}_v = 0$) but increases as time passes.

Figure 5.2
The horizontal velocity remains constant while the vertical velocity increases.



Example

A golf ball is thrown horizontally off a cliff at a velocity of 20 m s^{-1} and takes 4 s to reach the ground below. Calculate **(a)** the height of the cliff; **(b)** how far the ball will land from the base of the cliff; **(c)** the impact velocity of the ball.

Solution

(a) In the vertical direction:

$$\boldsymbol{u}_v = 0 \text{ m s}^{-1},\ \boldsymbol{a} = -10 \text{ m s}^{-2},\ t = 4 \text{ s},\ \boldsymbol{s}_v = ?$$

$$\begin{aligned} \boldsymbol{s}_v &= \boldsymbol{u}_v t + \tfrac{1}{2}\boldsymbol{a}t^2 \\ &= 0 + \tfrac{1}{2} \times -10 \times 4^2 \\ &= -80 \text{ m} \end{aligned}$$

(b) In the horizontal direction:

$$\boldsymbol{u}_h = 20 \text{ m s}^{-1},\ \boldsymbol{a} = 0 \text{ m s}^{-2},\ t = 4 \text{ s},\ \boldsymbol{s}_h = ?$$

$$\begin{aligned} \boldsymbol{s}_h &= \boldsymbol{u}_h t + \tfrac{1}{2}\boldsymbol{a}t^2 \\ &= 20 \times 4 + 0 \\ &= 80 \text{ m} \end{aligned}$$

(c) Impact velocity is the sum of horizontal velocity, which remains constant at 20 m s^{-1}, and the final vertical velocity. This is a vector summation. The vertical velocity on impact, $\boldsymbol{v}_v = \boldsymbol{u}_v + \boldsymbol{a}t = 0 + -10 \times 4 = -40$ m s^{-1}.

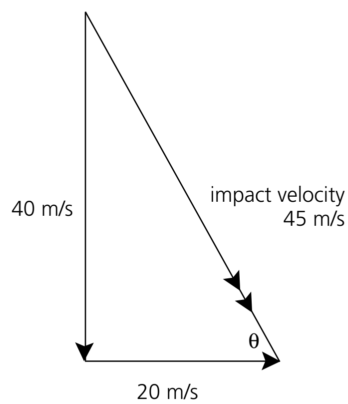


Figure 5.3

Using Pythagoras's theorem:

$$\begin{aligned} \boldsymbol{v}^2 &= 40^2 + 20^2 = 1600 + 400 \\ \boldsymbol{v} &= \sqrt{2000} = 45 \text{ m s}^{-1} \end{aligned}$$

The angle of impact, θ, can be found from $\tan\theta = \frac{40}{20} = 2.0$.

Hence θ = 63°.

Questions

1 A motorcycle is driven off a cliff at a horizontal velocity of 25 m s^{-1} and takes 5 s to reach the ground below. Calculate **(a)** the height of the cliff; **(b)** the distance out from the base of the cliff that the motorcycle lands; **(c)** the impact velocity.

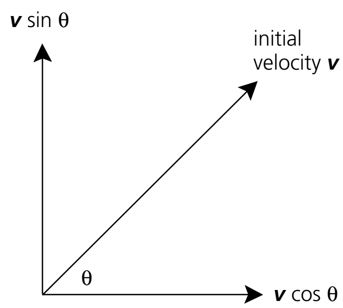
2 A rock is thrown horizontally at 8 m s^{-1} off a 100 m high cliff. Calculate **(a)** how long it takes to hit the ground; **(b)** its impact velocity; **(c)** how far out from the cliff it lands.

PROJECTION AT AN ANGLE 5.3

Not all objects are thrown in a horizontal direction. Cannonballs, footballs and netballs, for example, are often projected upward at an angle.

To study projectile motion, we let θ be the angle at which the object is thrown relative to the horizontal. This is called the **elevation angle**.

Figure 5.4

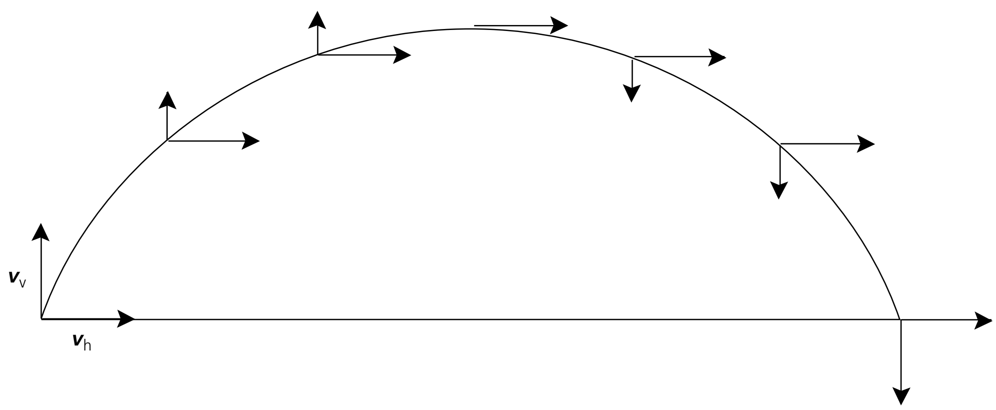


NOVEL CHALLENGE

A flea can jump 18.4 cm high when jumping at 45°. How far horizontally will it go?

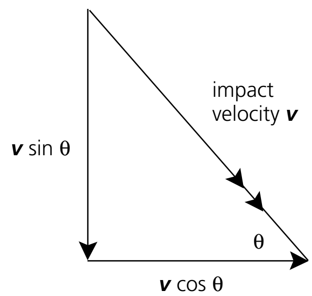
The motion of the projectile is a **parabola** because the vertical displacement varies as a function of t^2 (i.e. $\boldsymbol{s}_v = \boldsymbol{u}_v t + \frac{1}{2}\boldsymbol{a}t^2$) as it is uniformly accelerated motion whereas the horizontal displacement varies with just t (i.e. $\boldsymbol{s}_h = \boldsymbol{v}_h t$) as it is constant velocity. The horizontal displacement is called the **range**.

Figure 5.5 The vertical velocity changes while the horizontal velocity stays constant.



The **impact** velocity will have the same magnitude as the **launch** velocity, but be directed in a general downward direction not upward as at launch (Figure 5.6).

Figure 5.6



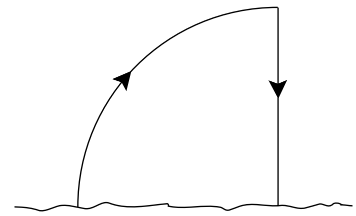
The horizontal component of velocity remains constant for the duration of the flight. The vertical component at launch equals the vertical component at impact but in the opposite direction. Recall from an earlier chapter that for vertical motion, initial speed equals final speed for an object returning to the same horizontal level.

Some old ideas challenged

Until the time of Galileo, the motion of a projectile was based on the teachings of Greek philosopher Aristotle. For example, Albert of Saxony (1316–90), rector of Paris University, taught that the trajectory of a projectile was in three parts: firstly, the upward motion where the initial impetus suppressed gravity; secondly, a period where the projectile's impetus and gravity were compounded; and thirdly, when gravity and air resistance overcame the natural impetus. This produced a trajectory as shown in Figure 5.7.

It wasn't until 1638 that the trajectory of a projectile could be described mathematically. Galileo's description proved to be correct and has been the basis of mechanics since. The mathematical techniques that Galileo pioneered, later refined by Newton, can be seen in the examples that follow.

Figure 5.7
Until the 1600s, people thought that projectile motion was more like this.



Example

The L16 mortar is a weapon currently used by Commonwealth defence forces. If a mortar shell was fired at 200 m s^{-1} at an angle of 40° to the ground, calculate:

(a) the initial vertical and horizontal components of the velocity;
(b) the maximum height reached;
(c) the time of flight (total time taken from start to finish);
(d) the horizontal range;
(e) the impact velocity.

Solution

Let the upward direction be positive: $\boldsymbol{a} = -10$ m s^{-2}.

(a)
- Vertical: $\boldsymbol{u}_v = \boldsymbol{v} \sin\theta = 200 \times \sin 40° = +129$ m s^{-1} in positive direction (up).
- Horizontal: $\boldsymbol{u}_h = \boldsymbol{v} \cos\theta = 200 \times \cos 40° = 153$ m s^{-1}.

(b) At maximum height $\boldsymbol{v}_v = 0$ m s^{-1}.

$$(\boldsymbol{v}_v)^2 = (\boldsymbol{u}_v)^2 + 2\boldsymbol{a}\boldsymbol{s}_v\text{, hence } \boldsymbol{s}_v = \frac{\boldsymbol{v}^2 - \boldsymbol{u}^2}{2\boldsymbol{a}} = \frac{0^2 - (+129)^2}{2 \times -10} = +832 \text{ m}$$

(c) Time of flight can either be calculated by **(i)** determining the time taken to reach maximum height ($\boldsymbol{v} = 0$) and doubling it; or **(ii)** determining time taken until final vertical velocity is equal and opposite to initial vertical velocity; or **(iii)** until vertical displacement is zero again.

By **(i)** $\boldsymbol{v}_v = \boldsymbol{u}_v + \boldsymbol{a}t$, hence $t = \dfrac{\boldsymbol{v} - \boldsymbol{u}}{\boldsymbol{a}} = \dfrac{0 - (+129)}{-10} = 12.9$ seconds. Total time = 25.8 s.

By **(ii)** $\boldsymbol{v}_v = \boldsymbol{u}_v + \boldsymbol{a}t$, hence $t = \dfrac{\boldsymbol{v} - \boldsymbol{u}}{\boldsymbol{a}} = \dfrac{-129 - (+129)}{-10} = 25.8$ s.

By **(iii)** $\boldsymbol{s}_v = \boldsymbol{u}_v t + \frac{1}{2}\boldsymbol{a}t^2$, hence $0 = +129t + -5t^2$; $5t = 129$; hence $t = 25.8$ s.

(d) Horizontal range = horizontal component of initial velocity × time of flight.

$$\boldsymbol{s}_h = \boldsymbol{v}_h \times t = 153 \times 25.8 = 3947 \text{ m}$$

(e) The impact velocity will have the same magnitude as the initial velocity, but will be directed generally downward not up. The angle of impact (θ) will be the same as the angle of elevation (40°). Thus, the impact velocity is 200 m s^{-1} at an angle 40° to the horizontal.

NOVEL CHALLENGE

Acapulco cliff divers jump off a cliff 35 m high and just miss rocks 5 mm out from the base. What is their minimum push-off speed?

NOVEL CHALLENGE

On the Moon, astronauts hit a golf ball 180 m. If they hit the same ball on Earth with the same speed and angle, how far will it go (neglect air resistance)? Note $\boldsymbol{g}_{moon} = 1.6$ m s^{-2}. By the way, there are three golf balls still on the Moon. Learn this off by heart — it could be useful.

Complementary angles of elevation

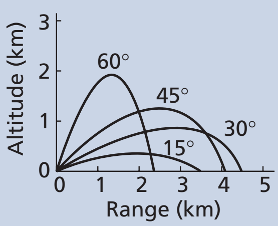
The range of a projectile fired at an elevation angle of 40° will also be the same if it is fired at 50°. The angles 40° and 50° are called **complementary angles** because they add up to 90°. Other examples of complementary pairs are: 30° and 60°; 20° and 70° etc. In other words, the range of a projectile will be the same for elevation angles of θ and 90° – θ. It is interesting that sin θ = cos (90° – θ).

Activity 5.2 TOY CANNON

If you have access to a toy cannon, try firing some projectiles at complementary angles and collect some data. Perhaps you could design a device that uses a rubber band, a mousetrap or a spring to fire small objects up an incline. Then you could vary the elevation angle. Whatever you do, you should aim to confirm or refute the above assertion about complementary angles.

NOVEL CHALLENGE

The following graphs show how the range and altitude of a projectile changes with elevation angle **in the presence of air**. Plot a graph of maximum altitude versus elevation angle and predict maximum altitude for an angle of 90°. *Should the graph pass through the origin (0,0)? Why?*



Example

In the earlier example, an elevation angle of 40° produced a range of 3947 m. If the theory is correct, then an angle of 50° should produce the same range.

(a) Prove this assertion.
(b) By how much do the times of flight differ?
(c) Do the impact velocities differ? (The initial velocity was 200 m s^{-1}.)

Solution

(a) Let $\boldsymbol{a} = -10$ m s^{-2}.

$$\boldsymbol{u}_v = \boldsymbol{v}\sin\theta = 200 \times \sin 50° = +153 \text{ m s}^{-1} \text{ (upward)}$$
$$\boldsymbol{u}_h = \boldsymbol{v}\cos\theta = 200 \times \cos 50° = 129 \text{ m s}^{-1}$$

Impact velocity in vertical direction ($\boldsymbol{v}_v$) = $-\boldsymbol{u}_v$ = –153 m s^{-1}.
Hence, the range is identical.

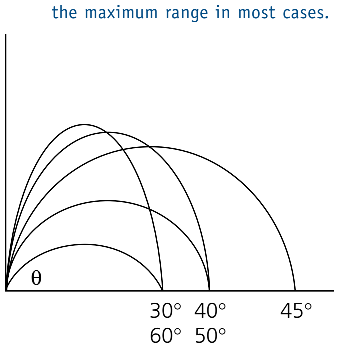
$$\boldsymbol{v}_v = \boldsymbol{u}_v + \boldsymbol{a}t, \text{ hence } t = (\boldsymbol{v}_v - \boldsymbol{u}_v/\boldsymbol{a}) = (-153 - +153)/10 = 30.6 \text{ s}$$
$$\boldsymbol{s}_h = \boldsymbol{v}_h \times t = 129 \times 30.6 = 3947 \text{ m}$$

(b) The times of flight were: for 40°, t = 25.8 s; for 50°, t = 30.6 s; difference was 4.8 s.
(c) Impact velocities are different but only in direction not magnitude.
For 40°, $\boldsymbol{v}_{impact}$ = 200 m s^{-1} at 40° to horizontal.
For 50°, $\boldsymbol{v}_{impact}$ = 200 m s^{-1} at 50° to horizontal.

Maximum range

It was the invention of the cannon in the late 1400s that created a new form of warfare. War at sea using cannons became more common and defence using medieval castles became obsolete. Medieval mechanics also became obsolete. Until then, the motion of a projectile was only of philosophical interest because they all thought they knew how projectiles moved — after all, Aristotle described the motion over 1000 years earlier and no one was prepared to challenge his theories. The theories weren't challenged until they had to be tested in warfare and were found wanting. Aiming was very much a hit-or-miss affair; there was no way of determining the trajectory or even the angle of launch in advance. It wasn't until self-taught engineer Niccolo Fontana published the results of his experiments in 1546 that gunners realised a 45° angle of elevation would give the maximum range.

Figure 5.8 An elevation angle of 45° produces the maximum range in most cases.



In Figure 5.8 the maximum range can be calculated by letting θ = 45°. In this case the horizontal and vertical components of the initial velocity both equal 141 m s^{-1}, the time of flight equals 28.2 s and the maximum range works out to be 3976 m.

Activity 5.3 COMPUTER SIMULATION

SR

If you have access to a computer and spreadsheet you may like to use this exhaustive method of determining the range at different elevations and the elevation for maximum range.

1 The horizontal range (R) can be found by a single formula deduced in the following manner:

(a) Horizontal range $\boldsymbol{s}_{\text{h}} = \boldsymbol{u} \cos \theta \times t = R$; maximum vertical height $\boldsymbol{v}_{\text{v}} = \boldsymbol{u} \sin \theta + \boldsymbol{a}t/2$.

(b) Eliminating t between the equations yields: $R = \dfrac{2\boldsymbol{u}^2}{\boldsymbol{a}} \sin \theta \cos \theta$.

(c) Knowing the identity $\sin 2\theta = 2 \sin \theta \cos \theta$, we obtain $R = \dfrac{\boldsymbol{u}^2}{\boldsymbol{a}} \sin 2\theta$.

2 Set up a spreadsheet and calculate the range for all values of θ from 0° to 90° using a nominal velocity of 100 m s^{-1}.

3 Is the maximum range achieved at an elevation of 45°?

4 Do complementary angles produce the same range? Give an example.

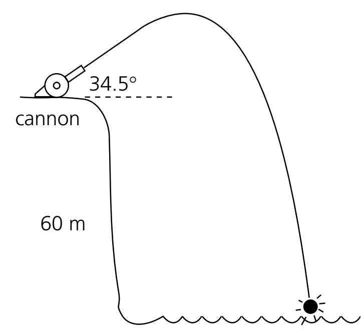
NOVEL CHALLENGE

The world speed record for an archery shot over 100 m is 1.64 seconds (220 km h^{-1}). Calculate the elevation angle of the arrow so that it hits the bull's eye at the same height as that from which it was fired (shoulder high).

More complex situations

If the projectile travels to a point *lower* than its starting point then the situation is more complex. Imagine throwing a ball up and out off a cliff. Another complex situation arises when the projectile lands *higher* up than the starting point, for example throwing a book to someone up on a verandah or shooting a basketball into the hoop.

Figure 5.9(a)



Example: Lower final horizontal displacement

A cannon is fired from the edge of a cliff, which is 60.0 m above the sea (Figure 5.9(**a**)). The cannonball's initial velocity is 88.3 m s^{-1} and it is fired at an upward angle of 34.5° to the horizontal. Determine: **(a)** the time the ball is in the air; **(b)** the impact velocity; **(c)** the horizontal distance out from the base of the cliff that the ball strikes the water.

Solution

- Vertical component of initial velocity $\boldsymbol{u}_{\text{v}} = 88.3 \sin \theta = +50.0$ m s^{-1} (positive is up).
- Horizontal component of initial velocity $\boldsymbol{u}_{\text{h}} = 88.3 \cos \theta = 72.8$ m s^{-1}.
- The final vertical displacement $\boldsymbol{s}_{\text{v}} = -60.0$ m
- Initial vertical velocity $\boldsymbol{u}_{\text{v}} = +50.0$ m.
- Vertical acceleration $\boldsymbol{a} = -10$ m s^{-2}.

$$\begin{aligned}
\boldsymbol{s} &= \boldsymbol{u}t + \tfrac{1}{2}\boldsymbol{a}t^2 \\
-60 &= +50t + \tfrac{1}{2}(-10)t^2 \\
5t^2 - 50t - 60 &= 0 \\
t^2 - 10t - 12 &= 0 \\
t &= \frac{-(-10) \pm \sqrt{(-10)^2 - 4 \times 1 \times -12}}{2 \times 1} \\
t &= 11.1 \text{ s or } -1.1 \text{ s}
\end{aligned}$$

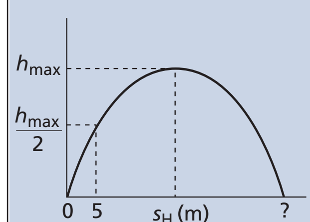
The negative solution is not reasonable, therefore the time of flight is 11.1 s.

(b) The horizontal velocity $\boldsymbol{v}_{\text{h}}$ remains constant at 72.8 m s^{-1}.

NOVEL CHALLENGE

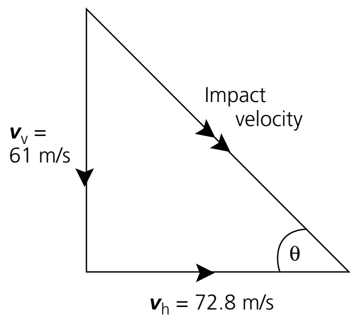
A really hard one! A cannonball is fired and, after travelling 5 m horizontally, it has reached half its maximum height.

At what horizontal distance will it land?



Ch05-Walding 4th 25/8/04 11:27 AM Page 115

Figure 5.9(b)



The vertical component of the velocity will change:

$$\begin{aligned} \boldsymbol{v}_v &= \boldsymbol{u}_v + \boldsymbol{a}t \\ &= +50 + -10 \times 11.1 \\ &= -61 \text{ m s}^{-1} \text{ (downward)} \end{aligned}$$

The total velocity is the vector sum of the two components (Figure 5.9(**b**)). Using Pythagoras's theorem:

- Impact velocity = $\sqrt{61^2 + 72.8^2}$ = 95 m s^{-1}.
- Using trigonometric ratios: $\theta = \tan^{-1} \dfrac{61}{72.8} = 40°$.

(c) Horizontal distance ($\boldsymbol{s}_h$) = horizontal component of velocity ($\boldsymbol{v}_h$) × time of flight (t).

$$\boldsymbol{s}_h = 72.8 \times 11.1 = 808 \text{ m from base of the cliff}$$

Note: you *can't* use the formula $R = \dfrac{\boldsymbol{u}^2}{\boldsymbol{a}} \sin 2\theta$ because the projectile is not landing at a position level with where it was thrown. The range formula is assuming the vertical displacement is zero.

Questions

3 A tennis ball close to the ground is hit by a racquet with a velocity of 30 m s^{-1} at an angle of 25° to the horizontal. Find **(a)** the initial vertical and horizontal components of the velocity; **(b)** the maximum height reached; **(c)** the time of flight; **(d)** the horizontal range.

4 A football is kicked off the ground at an angle of 30° to the horizontal. It moves away at 23.0 m s^{-1}. Calculate **(a)** the vertical velocity after 1.0 s; **(b)** the velocity of the ball after 1.0 s; **(c)** the maximum height reached; **(d)** the time of flight; **(e)** the range of the ball.

5 A rock is thrown off a 100.0 m cliff upward at an angle of 20° to the horizontal. If it has an initial velocity of 15 m s^{-1} and strikes the rocks below, calculate **(a)** the time of flight; **(b)** the impact velocity; **(c)** how far out from the base of the cliff the rock strikes the ground.

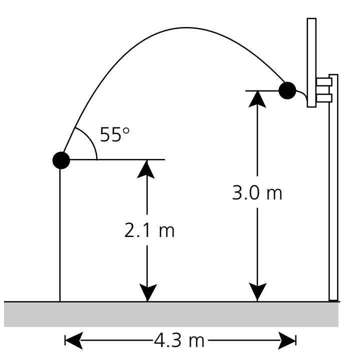
6 A difficult one! A basketball player shoots a ball at an angle of 55° into a hoop on a post 4.3 m away (Figure 5.10). If the ball is released from a height of 2.1 m and lands in the net, which is 10 feet (3.0 m) off the ground, calculate the initial speed of the ball for this foul shot to be successful.

7 Emmanuel Zacchini was a famous American 'human cannonball'. In 1940 he attempted to clear a Ferris wheel 18 m high after being launched from a cannon at an elevation angle of 53° and a muzzle velocity of 26.5 m s^{-1}.

(a) If his point of projection from the cannon was 3.0 m above the ground, did he clear the Ferris wheel?

(b) How far away from the cannon should the net have been placed?

Figure 5.10
For question 6.



The effect of air on projectiles

Aristotle argued that once a projectile ran out of impetus it would fall vertically from the sky. Galileo argued that this was wrong — the trajectory would be parabolic. Galileo was right — or was he? In the discussion so far we have ignored air resistance but when it is taken into account the trajectory is different. Aristotle is almost right but for the wrong reasons. At low speeds air resistance is negligible. But at greater speeds it becomes considerable. For instance, a flyball hit at an angle of elevation of 60° at 45 m s^{-1} will have different trajectories in air compared with those in a vacuum. Table 5.1 summarises the differences.

Table 5.1 TRAJECTORIES OF A BASEBALL

	PATH A (AIR)	PATH B (VACUUM)
Range	100 m	177 m
Maximum height	53 m	77 m
Time of flight	6.6 s	7.9 s

Figure 5.11 shows the difference between the trajectory of a ball as predicted by a computer model **(a)** and that of a bullet as tracked by ballistics experts on a rifle range **(b)**. The differences come about because bullets have more complicated motions than a round ball in flight.

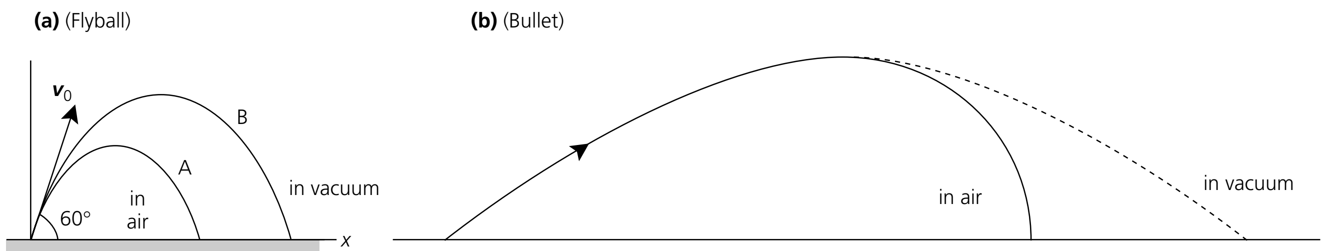


Figure 5.11
(a) The path of a flyball calculated taking air resistance into account (A) and in a vacuum (B).
(b) The dotted line is a trajectory of a bullet in a vacuum. The solid line shows how it is modified by air drag.

Trial-and-error has shown that the maximum range for a bullet fired in air is achieved at an elevation of 33°, a rough rule-of-thumb that works for most guns. As a crude approximation, the angle of descent is $2\frac{1}{2}$ times the angle of launch, so for a 33° elevation of fire, the bullet will arrive at 82.5°, or very nearly vertical. Any greater elevation of the gun merely means that the bullet will actually drop vertically and the last part of the flight will add nothing to the range. So the war veterans were probably right — bullets did fall on them vertically from the sky (and were just as lethal).

Exterior ballistics

Once a bullet leaves the muzzle of a gun, the laws of exterior **ballistics** take over as we have seen above. Ballistics comes from the Greek word *ballein* meaning 'to throw'. Modern high-speed photography enables physicists, chemists and engineers to study the explosion of propellant and the resulting motion of a projectile.

Typically, the bullet exits the muzzle at about 800 m s^{-1}, spinning at some 3000 revolutions per second. At first, it goes off down the range with a slight wobble, which straightens out after about 100 m, whereon it settles down to the main part of its flight, nose first, spinning steadily. This is the useful part of the bullet's life and it is intended that the bullet should hit its target during this stage. In the last part of its flight, the final slowing occurs and the bullet 'drops out of the sky'. The spinning tries to keep the bullet pointing straight ahead but as it falls toward Earth, the bullet cuts through the air sideways and air drag becomes great (Figure 5.12).

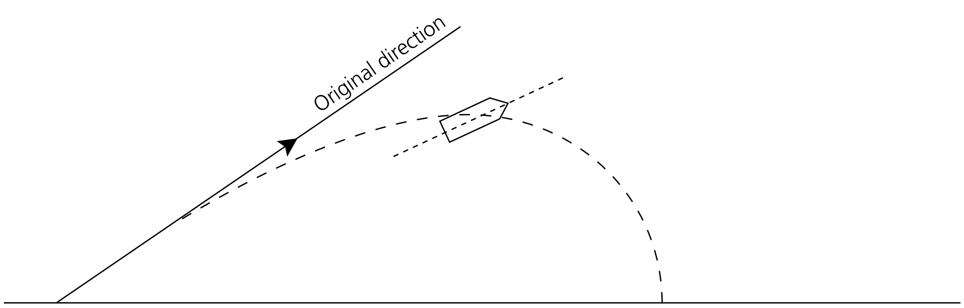


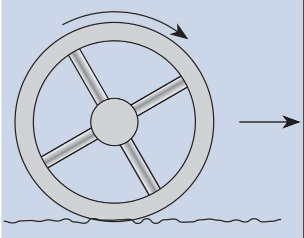
Figure 5.12
An exaggerated view of a wobbling (overstable) bullet, showing how it can fly almost broadside at the end of the trajectory.

The bullet begins to tumble end-over-end and by this stage has a very unpredictable trajectory and is too unreliable. Rifles generally have an effective range of 400–900 m, although weapons like the AR15 Armalite (USA) are designed for modern jungle warfare and are only accurate to 450 m but have an enormous muzzle velocity of 990 m s^{-1} to compensate. Because the bullet is tumbling at the end of this distance, it tears apart whatever it hits.

UNIFORM CIRCULAR MOTION 5.4

NOVEL CHALLENGE

When a wheel rolls along, is any point at rest?



It really wasn't until the 1500s that people began to believe that the Earth rotates on its own axis. Until then, the rate of rotation of objects was of little consequence. Today, rotation and its measurement is of fundamental importance to society, whether it is the rotation of microwave carousels, CDs, car tyres, engines, sewing machines, nuclei or orbiting satellites. In this section we will be looking at **circular motion**, that is, motion in a circle.

A ball on a string

Imagine you are whirling a ball in a horizontal circle on a piece of string. By Newton's first law of motion, the ball is attempting to travel in a straight line but is stopped from doing so by your pull on the string. As your hand is at the centre of the circle in which the ball moves, the force on the string and hence on the ball is always towards your hand and hence towards the centre. This force is called a **centripetal** force (Latin *centrum* = 'centre', *petere* = 'seek'). When the object travels at constant speed in a circle, it is said to be undergoing **uniform circular motion**. Notice that its direction is continually changing so its velocity is not constant.

Figure 5.13

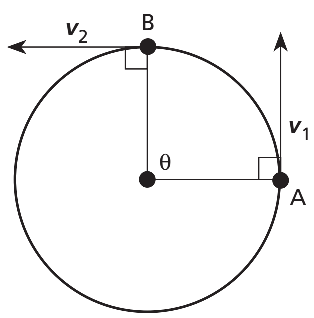
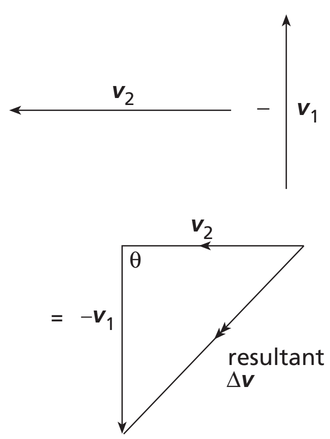


Figure 5.13 shows the motion of a ball moving in a circle of radius r at constant speed. The velocity at any point on the circle is a tangent to the path at that point. For instance, at position A, the velocity vector $\boldsymbol{v}_1$ points up the page. At point B, the velocity vector $\boldsymbol{v}_2$ points to the left but still has the same length as the speed remains the same. As the direction of the velocity has changed, the ball is said to be accelerating (**centripetal acceleration**). The magnitude and direction of this acceleration can be calculated by determining the change in velocity:

Change in velocity ($\Delta\boldsymbol{v}$) = final velocity ($\boldsymbol{v}_2$) − initial velocity ($\boldsymbol{v}_1$).

When we subtract a vector quantity, we turn it into an addition by reversing the direction of the initial vector and adding the arrows head to tail. Hence: $\Delta\boldsymbol{v} = \boldsymbol{v}_2 + {}^{-}\boldsymbol{v}_1$. As can be seen from Figure 5.14, the resultant is directed to the centre of the circle, hence 'centre seeking'.

Figure 5.14



Using similar triangles, it can be shown that the centripetal acceleration is given by:

$$a_c = \frac{v^2}{r}$$

where r is the radius of the circular path in metres. Note that the vector quantities a and v are no longer typed in **bold**. This is because they are not in the same direction. The acceleration is toward the centre whereas the velocity is at right angles to this.

The ball is experiencing a centripetal force to keep it moving in a circle. This is provided by the tension in the string. Using Newton's second law of motion ($\boldsymbol{F} = m\boldsymbol{a}$) we get:

$$\boldsymbol{F}_c = m\frac{v^2}{r}$$

A car going around a curve safely

A racing car travelling around a circular track is similar to a ball being whirled around on a string. A vehicle going round a bend on a level road can be viewed also as going on a circular path. The sideways friction between the tyres and the road provides the force needed to stop the car just going straight ahead. The friction provides the centripetal force. If the car hit a wet patch all of a sudden, the friction would be reduced and insufficient centripetal force could be provided so the car would tend to go straight ahead, possibly even spinning out of control.

Recall from earlier work that friction (F_f) is proportional to the force pressing the surfaces together (the normal reaction F_N): $F_f = \mu F_N$. On horizontal ground, the normal reaction is equal to the object's weight (F_w or mg).

If centripetal force is provided by the friction we can combine the two equations:

$$\boldsymbol{F}_c = \frac{mv^2}{r} \text{ and } \boldsymbol{F}_f = \mu \boldsymbol{F}_N = \mu m\boldsymbol{g}\text{, then } \frac{mv^2}{r} = \mu m\boldsymbol{g}$$
$$\text{i.e. } v_{max} = \sqrt{\mu \boldsymbol{g} r}$$

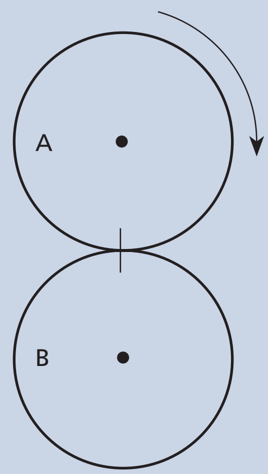
The maximum safe speed to go around a curve is represented by $\boldsymbol{v}_{max}$ in the final equation above. The mass of the car doesn't come into the equation so in this case has no effect on the safe speed. Big cars have the same maximum safe speed as small cars.

NOVEL CHALLENGE

Some coins were placed on a turntable in a line from the centre to the edge. The turntable was then turned on. *What do you predict will happen?*

NOVEL CHALLENGE

How many revolutions will coin A do while rotating around coin B? Try it. You'll be surprised.



Revolutions per second

You probably don't know the speed of the Moon about the Earth in metres per second or even kilometres per hour. But you would know that it makes one revolution in just over 27 days. Engine speeds too are usually expressed in a number of revolutions per minute (rpm). At idle, they might turn at 750 rpm and at cruising speed may reach say 4000 rpm. It depends on the car.

The distance covered in one revolution by an object in uniform circular motion at a distance r from the centre is equal to the circumference of the circle: $s = 2\pi r$. If the time taken to complete one revolution (called the **period**) is T, then:

$$v = \frac{s}{t} = \frac{2\pi r}{T}$$

This velocity is called the **tangential velocity** (Latin *tangere* = 'to touch'). It is sometimes called **radial** velocity. Angular velocities will be dealt with later.

Combining equations we get: $a_c = \dfrac{4\pi^2 r}{T^2}$ $\quad F_c = \dfrac{m4\pi^2 r}{T^2}$

Example

A motorcycle and rider with a total mass of 1250 kg are travelling around a circular track of radius 50 m at a constant speed of 40 m s^{-1}. Calculate **(a)** the centripetal acceleration; **(b)** the centripetal force; **(c)** the time it takes to complete one lap.

Solution

(a) $a_c = \dfrac{v^2}{r} = \dfrac{40^2}{50} = 32 \text{ m s}^{-2}$.

(b) $\boldsymbol{F}_c = m\dfrac{v^2}{r} = 1250 \times 32 = 40\,000 \text{ N}$.

(c) $v = \dfrac{2\pi r}{T}$, or $T = \dfrac{2\pi r}{v} = \dfrac{2 \times 3.14 \times 50}{40} = 7.85 \text{ s}$.

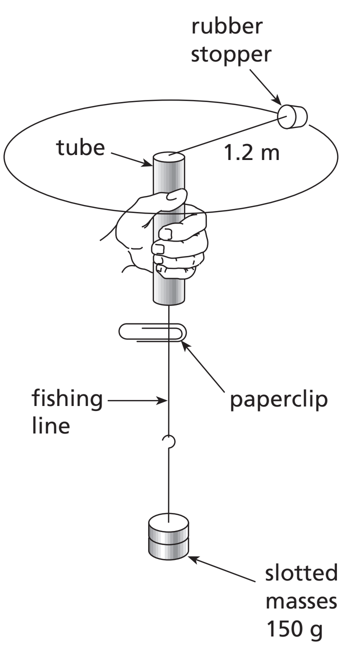
NOVEL CHALLENGE

The government steamer *Relief* attended the lighthouses along the Queensland coast from 1899 to 1952. To cope with the huge spray of seawater on the windows of the steering cabin, a novel approach was taken. The windscreen in part consisted of a circular glass disk about 40 cm diameter that spun at high speed.
How did this keep the seaspray off the window? Why couldn't they use windscreen wipers as in a car? Propose two advantages and two disadvantages of this sytem compared with wipers.

EI Activity 5.4 THE WHIRLING STOPPER

Tie a rubber stopper to a piece of string and whirl it in a horizontal circle above your head. See if you can let it go so that it will hit the wall of your room at right angles. Whereabouts in its circular travel did you have to let it go to achieve this? Which law of motion is confirmed by this?

Figure 5.15



Example 1

In an investigation of uniform circular motion, a student whirled a 50 g rubber stopper above his head in a horizontal circle of radius 1.2 m (Figure 5.15).

The string was passed through a piece of glass tubing and a set of slotted brass masses was suspended off the end of the string. It required 150 g of hanging mass to provide enough force to keep the rubber stopper whirling in a circle at a constant speed. Use $\boldsymbol{g} = 9.8$ m s^{-2} and calculate **(a)** the centripetal force provided by the hanging mass; **(b)** the tangential velocity of the stopper; **(c)** the period of the rubber stopper; **(d)** the time taken for 10 revolutions of the stopper.

Solution

(a) The centripetal force is provided by the weight of the hanging mass:

$$\boldsymbol{F}_c = \boldsymbol{F}_w = m\boldsymbol{g} = 0.150 \text{ kg} \times 9.8 \text{ m s}^{-2} = 1.47 \text{ N}$$

(b)

$$\boldsymbol{F}_c = \frac{m\boldsymbol{v}^2}{r}, \text{ or } \boldsymbol{v}^2 = \frac{\boldsymbol{F}_c r}{m} = \frac{1.47 \text{ N} \times 1.2 \text{ m}}{0.050 \text{ kg}} = 35.3$$

$$\boldsymbol{v} = \sqrt{35.3} = 5.9 \text{ m s}^{-1}$$

(c)

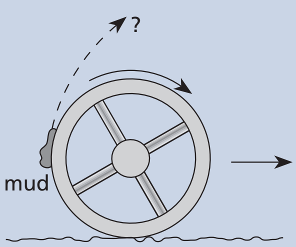
$$T = \frac{2\pi r}{\boldsymbol{v}} = \frac{2 \times 3.14 \times 1.2}{5.9} = 1.3 \text{ s}$$

(d) Time for 10 revolutions = 10 revolutions × 1.3 s/rev = 13 s.

NOVEL CHALLENGE

A wheel is rolling along with constant speed and a lump of mud is thrown off its hindmost point.

Will it touch the wheel again?



Example 2

A car of mass 1750 kg is rounding a curve of radius 70 m at a speed of 20 m s^{-1}. The surface is dry and the coefficient of friction between the tyres and the road is 0.65. The driver then hits a wet patch on the curve where the coefficient of friction is 0.25. Calculate **(a)** how much below the safe maximum speed the car is doing on the dry section of the curve; **(b)** whether the driver has to slow down to safely travel along the wet section and, if so, to what safe maximum speed; **(c)** would a smaller and lighter car allow the driver to go faster around the curve?

Solution

(a) $\boldsymbol{v}_{max} = \sqrt{\mu \boldsymbol{g} r} = 0.65 \times 10 \times 70 = 21.3$ m s^{-1}; the driver is 1.3 m s^{-1} below this speed.
(b) $\boldsymbol{v}_{max} = \sqrt{\mu \boldsymbol{g} r} = 0.25 \times 10 \times 70 = 13.2$ m s^{-1}; the driver has to slow down to this speed.
(c) $\boldsymbol{v}_{max}$ is independent of mass, so a lighter car would make no difference.

Cambered surfaces

Some curved motor car racing tracks are **cambered** or **banked**, that is, tilted in towards the centre of the curve. In this case the component of the vehicle's weight down the slope helps to provide centripetal force so the frictional force need not be as great. Alternatively, the car can safely travel at much higher speeds. Road engineers often camber roads the 'wrong' way for purposes of drainage. You could imagine the effect this has on maximum safe speeds?

Some other examples

- For a **space shuttle** and satellites orbiting the Earth or planets orbiting the Sun, the centripetal force is provided by gravitational forces. This will be dealt with in a later chapter.

- In a **gravitron** or rotor at an amusement park, the person is 'pressed' against the wall. Actually, the person is trying to travel in a straight line but the wall pushes on the person (the centripetal force) and the person pushes back. The centripetal force is the normal force directed radially inward on the rider. At high speeds, this normal force becomes sufficiently great to provide enough friction to stop the rider sliding down the wall under the influence of gravity.
- A **spin-dryer** works on a centripetal force principle. When the tub is spun at high speed, the force of attraction between the water and the clothes is insufficient to keep the water moving in a circle. The liquid moves tangentially and out through the holes in the sides of the tub.

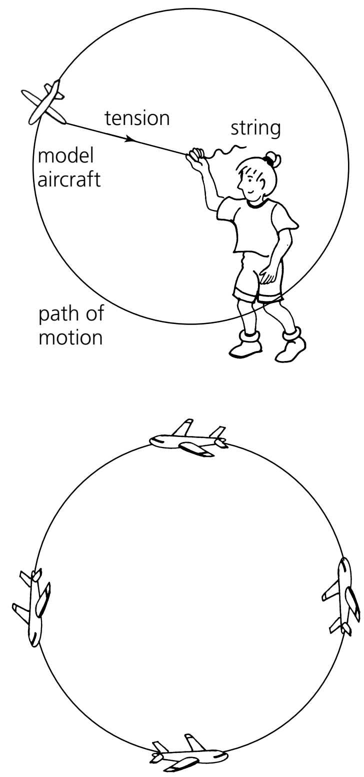
SR

Activity 5.5 SPIN-DRYER CHAMPIONSHIPS

1 Spin-dryers go pretty fast — too fast to see with the naked eye. Design a method to measure the speed of a spin-dryer in revolutions per minute (rpm). You don't have to build it or have the parts at home or school; just design the procedure and instrumentation.

2 If your method is simple, do it and report the result to the class.

Figure 5.16
Looping the loop in a vertical circle.



Questions

8 A car of mass 1900 kg is travelling at a constant speed of 25 m s^{-1} around a level corner of radius 50 m. Calculate **(a)** the centripetal acceleration; **(b)** the centripetal force acting on the car.

9 An aeroplane is travelling at 200 m s^{-1} in a circular path of radius 3000 m. Calculate **(a)** the centripetal acceleration of the plane; **(b)** the time to complete one revolution.

10 The Moon takes a period of 27.3 days to complete one orbit of the Earth. If we consider the path to be circular then its average radius is 3.84×10^8 m from the centre of the Earth. Determine **(a)** the circumference of the Moon's path; **(b)** the speed of the Moon; **(c)** the Moon's centripetal acceleration; **(d)** the centripetal force (the Moon's mass is 7.34×10^{22} kg).

11 Spin-dryers revisited:
- **(a)** Why do clothes that comes out of a spin-dryer still feel damp?
- **(b)** Would continued spinning at the same speed get rid of more water?
- **(c)** Could you spin them completely dry?
- **(d)** How does the water get from the clothes in the middle to the outside (is there a more efficient way)?

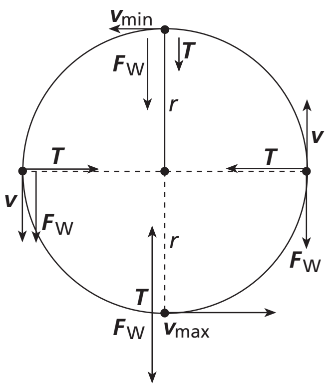
12 A mass of 150 g is whirled in a horizontal circle of radius 95.0 cm on a string. If 10 revolutions at constant speed take 4.5 seconds, calculate the tension in the string.

5.5 NON-UNIFORM CIRCULAR MOTION

The previous section dealt with uniform circular motion. This can be easily achieved by objects travelling in horizontal circles. When they travel in vertical circles it is difficult to keep the speed constant and this is called **non-uniform** circular motion. Two common examples of this are a ball on a string and an aircraft loop-the-loop. Devices that have stiff radial arms such as a bicycle wheel, a ferris wheel or a pulley cannot be considered non-uniform as they are completely rigid and all points on the circumference travel at the same speed.

When a ball is swung on the end of a string in a vertical circle, the speed of the ball is greatest at the bottom of the circle and slowest at the top of the circle. Hence, the centripetal acceleration is smallest at the top and greatest at the bottom. In the following, the symbol $\boldsymbol{T}$ is used to represent the tension in the string, whereas $\boldsymbol{F}_w$ represents the weight of the ball (= mg). Refer to Figure 5.17.

Figure 5.17



- At the top, the string doesn't have to pull as hard (F_c) because the weight is helping it pull down:

$$F_c = \frac{mv^2}{r} = T + F_w \text{ or } T = F_c - F_w$$

- At the side, the weight has no effect on the tension:

$$F_c = \frac{mv^2}{r} = T$$

- At the bottom, the string has to pull harder because it has to support the weight of the ball as well:

$$F_c = \frac{mv^2}{r} = T - F_w \text{ or } T = F_c + F_w$$

The apparent weight of the ball at the top or bottom is given by T.

Minimum velocity

The minimum velocity needed to keep a ball in a circular orbit is found to be the velocity at the instant when the string begins to slacken (i.e. when $T = 0$) at the top. This is when:

$$\frac{mv_{min}^2}{r} = T + F_w = 0 + F_w = mg$$

$$v_{min}^2 = gr \text{ so } v_{min} = \sqrt{gr}$$

Maximum velocity

The maximum velocity occurs at the bottom of the path:

$$\frac{mv_{max}^2}{r} = T - F_w$$

INVESTIGATING

Many factories, laboratories and industries use centrifuges. Locate two places that use centrifuges and write a report comparing and contrasting their uses and performances.

Example 1

The breaking strain of a string is 50 N. A 250 g ball is whirled in a vertical circle of radius 1.2 m. Calculate **(a)** the minimum velocity needed to keep the ball in orbit; **(b)** the maximum speed that the ball can have before the string breaks.

Solution

(a) $v_{min} = \sqrt{gr} = \sqrt{10 \times 1.2} = 3.5 \text{ m s}^{-1}$.
(b) Maximum tension occurs at the bottom of the ball's path:

$$\frac{mv_{max}^2}{r} = T - F_w = 50 - 0.25 \times 10 = 47.5 \text{ N}$$

$$v_{max}^2 = \frac{47.5 \times 1.2}{0.25} = 228, \text{ hence } v_{max} = \sqrt{228} = 15 \text{ m s}^{-1}$$

Example 2

A stunt pilot is diving his plane vertically downwards at a velocity of 200 m s^{-1} when he pulls out of the dive and changes his direction to a circular path of radius of 1000 m. If his mass is 70 kg and he continues to maintain constant speed in the circle, **(a)** what is the maximum centripetal acceleration he experiences; **(b)** what is the maximum force that his seat will exert on him? **(c)** If pilot 'blacks-out' when the acceleration is greater than 3g, will he stay conscious? **(d)** At what circular path radius would he be liable to black out?

Ch05-Walding 4th 25/8/04 11:27 AM Page 122

Solution

(a) Maximum acceleration (at bottom of path) $\boldsymbol{a}_c = \frac{\boldsymbol{v}^2}{r} = \frac{200^2}{1000} = 40\ \text{m s}^{-2}$.

(b) At bottom of loop, the seat provides the equivalent of the tension:

$$\frac{m\boldsymbol{v}^2_{max}}{r} = \boldsymbol{T} - \boldsymbol{F}_w,\ \text{hence}\ \boldsymbol{T} = \frac{m\boldsymbol{v}^2_{max}}{r} + \boldsymbol{F}_w = \frac{70 \times 200^2}{1000} + 70 \times 10 = 3500\ \text{N}$$

(c) $\boldsymbol{a}_c = 40\ \text{m s}^{-2}$. The number of '$g$' this is equal to is $40\ \text{m s}^{-2} \div 10\ \text{m s}^{-2} = 4$ '$\boldsymbol{g}$'. This is greater than 3$\boldsymbol{g}$ so the pilot will black out.

(d) To achieve 3$\boldsymbol{g}$ (30 m s^{-2}), the radius can be calculated:

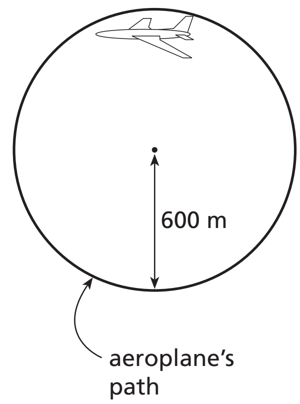
$$\boldsymbol{a}_c = \frac{\boldsymbol{v}^2}{r},\ \text{hence}\ r = \frac{\boldsymbol{v}^2}{\boldsymbol{a}_c} = \frac{200^2}{3 \times 10} = 1330\ \text{m}.$$

Even at this radius, the pilot may black out for a few seconds. Too tight a loop or too high a speed could cause the pilot (and crew) to black out for much longer. This could cause lack of control of the aircraft, death, or both. Some stunt!

Modern military aircraft typically have $\boldsymbol{g}$ limits of around +9.5 to −5.5 $\boldsymbol{g}$, although these boundaries are continually being pushed back. Sensors are fitted into most cockpits to allow the pilot to monitor $\boldsymbol{g}$ values to avoid overstressing the airframe. For additional safety and to cope with crash impacts, cockpit interiors are designed to withstand 20 $\boldsymbol{g}$ in any direction. Ejector seats and escape pods may suffer instantaneous loads (for about 0.1 s) in excess of 30 $\boldsymbol{g}$. The requirement is that a seat shoots a pilot from an aircraft at zero forward speed and zero altitude (the so-called 'zero–zero' seat) to an altitude at which the parachute can open safely. Alternatively, the seat must be able to clear the tailplane of an aircraft travelling at high speed. The record for a human experiencing $\boldsymbol{g}$-loading is around 86 $\boldsymbol{g}$ by the occupant of a rocket-sled. By comparison, civilian airlines experience a modest 1.5 $\boldsymbol{g}$ during take-off acceleration.

The world record for loops is held by David Childs (USA). He did 2368 loops in a Bellanca Decathlon plane over the North Pole on 9th August 1986. Imagine having 'Crazy Dave' in your physics class.

Figure 5.18
For question 13.



Questions

13 A pilot is performing aerial acrobatics at an air show. He drives around a vertical loop of radius 600 m (Figure 5.18). What is the minimum speed at the top of the loop?

14 A 75 kg pilot flies his plane in a vertical circle of radius 600 m and at the bottom of the loop he is travelling at 120 m s^{-1}.

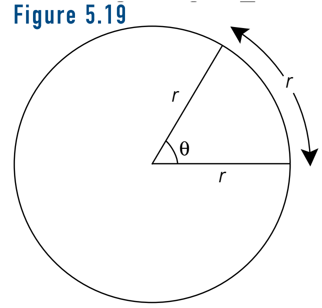
(a) What is the force of the seat on the pilot at this point?
(b) What is the acceleration in m s^{-2}?
(c) If he is known to black out at 5 '$\boldsymbol{g}$', would he black out at the bottom of the loop?
(d) If the plane was travelling at 80 m s^{-1} at the top of the loop, what would the force of the seat on the pilot be?
(e) At what speed would the plane have to travel for the pilot to be just weightless at the top of the loop, that is, his weight equals the centripetal force?

5.6 ANGULAR VELOCITY

When something completes one revolution it has gone through 360°. One revolution per second is the same as 360° per second. This is called its **angular velocity**.

In maths, you may have measured angles in **radians**. One radian (rad) is the angle when the arc length is the same as the radius of the circle (Figure 5.19). There are 2π radians in a circle of 360°, thus one revolution equals 2π radians. Angular velocity (ω) is usually

Figure 5.19



expressed in radians per second (rad s^{-1}). It is a vector quantity. The symbol ω is the Greek letter 'omega'. The word radian comes from the Latin *radius*, meaning the spoke of a wheel.

Tangential velocity ($\boldsymbol{v}$) = angular velocity (ω) × radius (r).

$$\boldsymbol{v} = \omega r \text{ or } \omega = \frac{\boldsymbol{v}}{r}$$

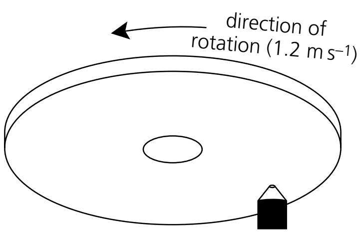
Centripetal acceleration and force can also be expressed in terms of angular velocity:

$$\boldsymbol{a}_c = \frac{\boldsymbol{v}^2}{r} = \omega^2 r = \omega \boldsymbol{v} \text{ and } \boldsymbol{F}_c = \frac{m\boldsymbol{v}^2}{r} = m\omega^2 r$$

The period T of a rotating object is given by: $\omega = \frac{2\pi}{T}$ or $T = \frac{2\pi}{\omega}$.

Why angular velocities?

Figure 5.20
An underside view of a CD showing the objective lens of the laser pickup.



You may wonder what the point of using angular velocities is. A spinning disc such as a CD (Figure 5.20) will have all points on the surface turning at the same angular speed even though different tracks will have different linear velocities. It makes the speed easier to state. Another common way of expressing angular speeds is **revolutions per minute (rpm)**. A microwave carousel does about 5 rpm.

Example 1

A tyre is turning at 20 m s^{-1} as a car travels along a road. If the diameter of the tyre is 62 cm, calculate **(a)** the angular velocity of the tyre; **(b)** the centripetal acceleration of a 2 g stone embedded in the tread of the tyre; **(c)** the centripetal force acting on the stone; **(d)** the rate of rotation of the tyre in rpm.

Solution

(a) Radius = 0.31 m; $\omega = \frac{\boldsymbol{v}}{r} = \frac{20}{0.31} = 64.5$ rad s^{-1}.

(b) $\boldsymbol{a}_c = \frac{\boldsymbol{v}^2}{r} = \frac{20^2}{0.31} = 1290$ m s^{-2}.

(c) $\boldsymbol{F}_c = m\boldsymbol{a}_c = 0.002 \times 1290 = 2.58$ N.

(d) 1 revolution = 2π radians; hence number of revolutions per second = $\frac{64.5 \text{ rad s}^{-1}}{2\pi}$ = 10.2 rps = 616 rpm.

Example 2

A flywheel of radius 2.0 m is rotating at 120 rpm. Calculate **(a)** the angular velocity; **(b)** the linear velocity of a point on the rim.

Photo 5.1
A tachometer. Note the 'red line' from 5 to 7 thousand revs per minute.

Solution

(a) 1 rpm = 2π rad min^{-1}; hence 120 rpm = 120 × 2π rad min^{-1} = $\frac{120 \times 2\pi}{60}$ rad s^{-1} = 4π rad s^{-1}.

(b) $\boldsymbol{v} = \omega r = 4\pi \times 2 = 25$ m s^{-1}.

Everyday examples of angular motion

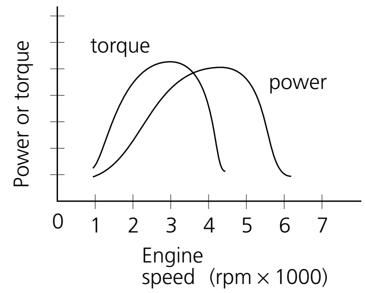
- **Record players** generally have three speeds: 78 rpm for the old bakelite 78s; 45 rpm for vinyl singles and $33\frac{1}{3}$ rpm for LPs. As angular speeds were constant for any particular record, the outside track of a 12 inch (30 cm) LP travelled at a higher linear speed than the inside track, so the outside track gave better sound reproduction. For instance, the outside track at a radius of 14.5 cm had a linear speed of 50 cm s^{-1}, whereas the inside track at a radius of 6.5 cm gave a linear speed of 22 cm s^{-1}.

To overcome the problem of differential track speeds, when compact disc players were developed the track speed was kept constant and the rotation speed was varied. For example, the linear speed of a CD is about 1.2 m s^{-1}, so for an outside track (radius 58 mm), the rotation rate is 200 rpm, whereas for an inside track (r = 23 mm), the rotation rate is 500 rpm (see Figure 5.20). Computer disk drives work on the same principle.

- **Car engines** generally idle at about 800 rpm and cruise at somewhere between 2000 and 4000 rpm. Cars with V8 engines generally have more power and torque (turning force) than either six or four cylinder cars so they can cruise at lower engine speeds. It is unusual for car engines to rev above 6000 rpm because the valves and other components can be damaged. Sportier cars are sometimes equipped with tachometers (Latin, *tachy* = 'swift'), which measure engine speeds in rpm. The maximum recommended speed is indicated with a red line and if you '**red-line**' an engine you are certainly giving it a good thrashing. The power and torque delivered by engines is not constant over the full range of engine speeds (Figure 5.21). Cars are geared so that drivers can maintain the engine speed just below the optimum power and torque range, which usually corresponds to the normal cruising speed in top gear. For instance, a Toyota Landcruiser with a 4.5 L, six cylinder petrol engine develops maximum power at 4600 rpm and maximum torque at 3200 rpm. At a cruising speed of 100 km h^{-1}, the engine turns over at a relatively slow 2100 rpm, leaving plenty of revs in reserve for overtaking.

  The same is true of motorcycles except that they run at much higher revs; a range of 6000 rpm to a red line at 12 000 rpm is typical.

Figure 5.21



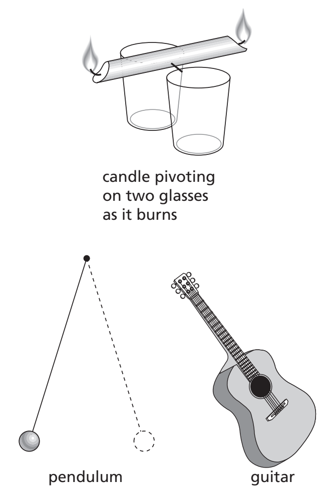
Questions

15 Calculate the force acting on a mass of 3 kg that is rotating at 5 rad s^{-1} in a circle of radius 30 cm.

16 A microwave oven carousel has a diameter of 40 cm and does one revolution in 12 seconds. Calculate **(a)** the angular velocity of the carousel; **(b)** the tangential velocity.

17 While reading the fifth song on a CD, the laser pickup diode is at a radial distance of 50 mm from the centre of the spinning disc. If the linear velocity of the disc directly above the laser pickup is 1.2 m s^{-1}, calculate the angular velocity in **(a)** rad s^{-1}; **(b)** rpm.

Figure 5.22
Three examples of periodic motion.



5.7 SIMPLE HARMONIC MOTION

A swinging pendulum, a vibrating guitar string and a mass **oscillating** on the end of a spring are all examples of **periodic motion** — motion in which an object continually moves back and forth over the same path in equal time intervals (Figure 5.22). The word *oscillate* means to move back and forth. It comes from the Latin *os*, meaning 'mouth' or 'face'. The Greeks used masks of the god Bacchus hung up as charms in vineyards and they swung back and forth in the breeze, hence os-cillate.

Not all periodic motions are as simple as a mask blowing in the bleeze; some are very complex. However, in this chapter we will look at a simple type of periodic motion called **simple harmonic motion** (SHM).

The vibrating mass

Figure 5.23

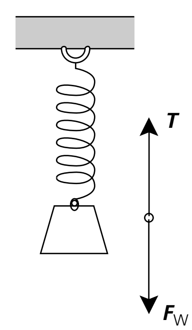


Figure 5.23 shows a mass attached to a spring hooked to the ceiling. When it is at rest, the tension in the spring and the weight are equal and opposite — or equally balanced. This position is called the **equilibrium** position (*equi* = 'equal', *libra* = 'balance'). There is no net

Ch05-Walding 4th 25/8/04 11:27 AM Page 125

force so the mass is not accelerating. The displacement of the mass from the equilibrium position is also called the **amplitude** (*x*) and is zero in this position.

If the mass is pulled down and let go it oscillates up and down as shown in Figure 5.24. A study of the forces and displacements is quite revealing.

Figure 5.24

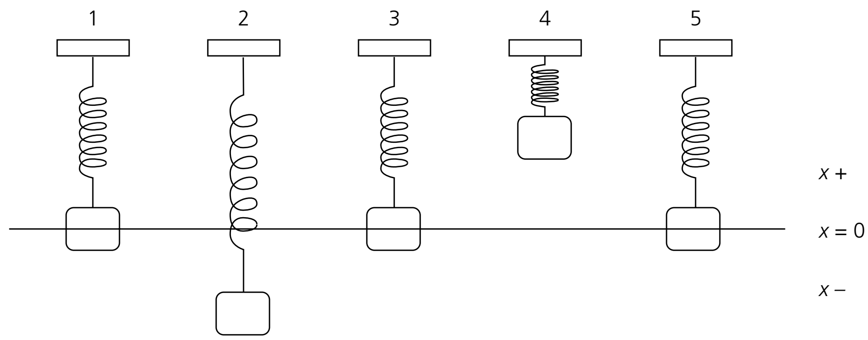


Table 5.2 summarises the variables involved.

NOVEL CHALLENGE

A candle with a nail through the middle is supported on two glasses and lit at both sides. *How could you check if the resulting motion is SHM or just periodic?*

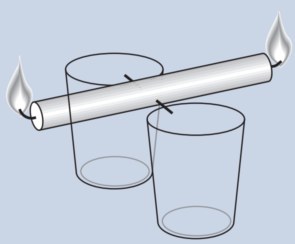


Table 5.2

	POSITION 1	POSITION 2	POSITION 3	POSITION 4	POSITION 5
Net force	0	max. up	0	max. down	0
Acceleration	0	max. up	0	max. down	0
Velocity	max. down	0	max. up	0	max. down
Displacement	0	max. down	0	max. up	0

- Position 1 — the mass is moving downward through its equilibrium position so the net force is zero but it is moving with maximum speed. As there is no net force, the acceleration must also be zero (Newton's second law: $\boldsymbol{F} \propto \boldsymbol{a}$).
- Position 2 — the mass is at its lowest point so displacement is a maximum in the downward or negative direction. The spring is stretched so the tension in it is greater than the weight of the object so the net force is directed upward (positive). Acceleration is also directed up.
- Position 3 — the mass is back to its equilibrium position but is now moving with maximum velocity upward.
- Position 4 — the spring is now unstretched so the tension in the spring is zero. The only force comes from the weight so the net force is a maximum in the downward (negative) direction. Displacement is a maximum in the positive direction.
- Position 5 — equilibrium, with the object moving down at maximum speed.

In summary:

- Simple harmonic motion (SHM) is periodic motion in which $\boldsymbol{F} \propto -\boldsymbol{x}$.
- When the force ($\boldsymbol{F}$) is a maximum, the displacement ($\boldsymbol{x}$) is a maximum but in the opposite direction.
- When the force is a minimum (zero), the displacement ($\boldsymbol{x}$) is also a minimum (zero).

Mathematically:

$$\boldsymbol{F} \propto -\boldsymbol{x} \text{ or } \boldsymbol{F} = -k\boldsymbol{x}$$

The constant (k) is called the **spring constant**. Its units will be N m^{-1}. The stiffer the spring the larger the spring constant.

Example

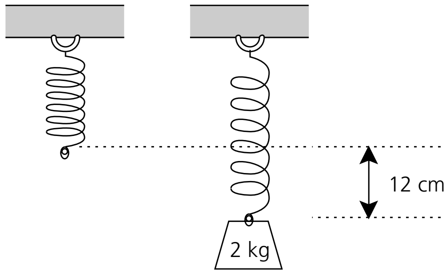


Figure 5.25

When a mass of 2.0 kg is attached to a spring it stretches by 12 cm (Figure 5.25).

(a) Calculate the spring constant.

(b) What would the stretch be if a further 1.0 kg was added?

Solution

(a) $\boldsymbol{F} = -k\boldsymbol{x}$ or $k = -\dfrac{\boldsymbol{F}}{\boldsymbol{x}} = -\dfrac{20}{0.12} = 167 \text{ N m}^{-1}$.

(b) $\boldsymbol{x} = \dfrac{\boldsymbol{F}}{-\boldsymbol{k}} = \dfrac{30}{-167} = 0.18 \text{ m} = 18 \text{ cm}$.

Experiments show that if the spring is of negligible mass compared with the object hanging on it, then the period (T) of the motion is given by:

$$T = 2\pi\sqrt{\frac{m}{k}}$$

Example

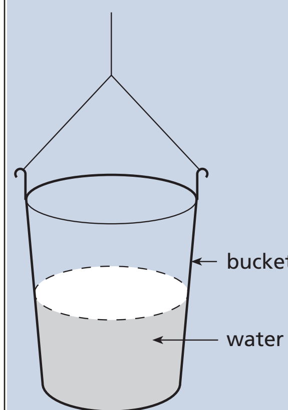
A light spring has a mass of 100.0 g attached to it. If it has a spring constant of 4.5 N m^{-1}, calculate the period of the vibrating spring.

Solution

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{0.1000}{4.5}} = 0.93 \text{ s}$$

NOVEL CHALLENGE

A pendulum has a bucket for the bob and it is half-filled |with water. *When it freezes, predict what happens to the period of the pendulum.*



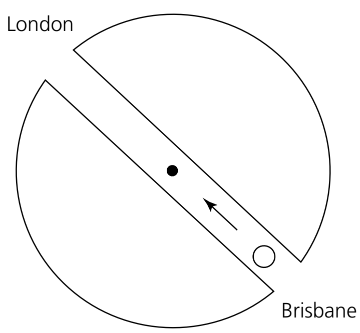
Journey to the centre of the Earth

An idea that has intrigued people for years is a hole through the Earth. Imagine a hole from Brisbane to London — it would be about 1.3×10^7 m long (Figure 5.26). If you dropped a parcel in one end it would come out the other some time later. A 1 kg parcel dropped into the hole at Brisbane would experience an initial force due to gravity of 10 N and would be pulled to the centre of the Earth some 6.5×10^6 m away. SHM would apply and we could calculate the force constant (k) $= \dfrac{\boldsymbol{F}_w}{\boldsymbol{x}} = \dfrac{10}{6.5 \times 10^6} = 1.5 \times 10^{-6} \text{ N m}^{-1}$.

Using the SHM formula: $T = 2\pi\sqrt{\dfrac{m}{k}} = 2\pi\sqrt{\dfrac{1}{1.5 \times 10^{-6}}} = 5066$ s (for one oscillation).

The time to get to the other side of the Earth would be half that or 2532 s (= 42 minutes).

Figure 5.26



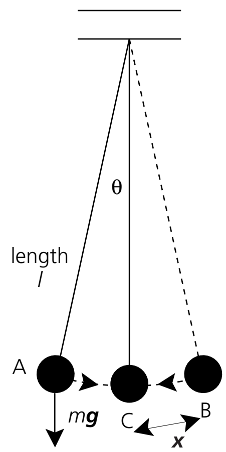
Questions

18 What assumptions have been made in the above example about the hole through the Earth that would make it an impossibility to achieve? List as many as you can.

19 A light spring has a mass of 200 g attached to it. When it is set oscillating, its period is measured to be 1.2 s. Calculate its spring constant.

20 To measure the mass (M) of an astronaut in the weightless conditions of space, an oscillating chair (mass m) bound to a spring is used. The body mass measuring device (BMMD) has a period of oscillation of 0.901 49 s when no one is in it. When one of the Skylab astronauts sat in it its period increased to 2.088 32 s. If the spring constant for the BMMD is 605.6 N m^{-1}, calculate the mass of the chair and of the astronaut.

Figure 5.27



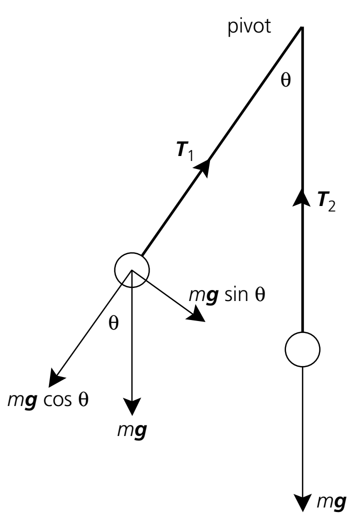
The pendulum

If you hang an apple on the end of a long thread fixed at its upper end, and then set it swinging, you can see that the motion is periodic (see Figure 5.27). It is also simple harmonic motion. Such an arrangement is called a **pendulum** (Latin *pendulus* = 'swinging'). The weight on the end is called the 'bob'. Why 'bob'? It comes from the Old French *bober*, meaning 'to mock'. When you mock someone your head moves up and down as you laugh.

As the pendulum, of length l, moves from A to B and back again to A, it makes a complete oscillation. The time required is the **period** (T). The number of oscillations per second is called its **frequency** (f). The sideways displacement ($\boldsymbol{x}$) is the sideways distance from the vertical or equilibrium position. The maximum displacement during the oscillations is called the **amplitude** (*amplus* = 'large'). The position C is called the **equilibrium** position. The forces acting on a pendulum during its travel are shown in Figure 5.28.

At an angle of θ as shown, the restoring force is equal to the component of the weight (= $m\boldsymbol{g}$) directed back to the equilibrium position (= $m\boldsymbol{g}\sin\theta$). The tension in the string ($\boldsymbol{T}_1$) is equal to the component (= $m\boldsymbol{g}\cos\theta$). At the equilibrium position, the restoring force is zero as θ equals zero and the component of the weight pulling the bob sideways is therefore also zero ($\sin 0° = 0$). The tension in the string ($\boldsymbol{T}_2$) is now equal to $m\boldsymbol{g}$ as $\cos 0° = 1$. The tension $\boldsymbol{T}_2$ is greater than $\boldsymbol{T}_1$.

Figure 5.28



The pendulum formula

Experiments show that the period of a pendulum is given by:

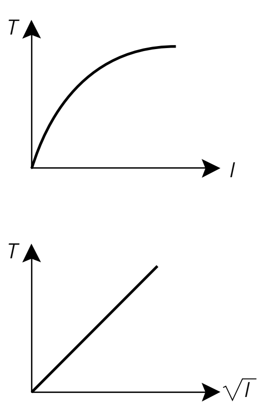
$$T = 2\pi\sqrt{\frac{l}{g}}$$

Note that the period is independent of the mass of the bob and amplitude (if it is fairly small, e.g. less than 20°) but as the graphs in Figure 5.29 show, T is proportional to $\sqrt{l}$.

Galileo is said to have confirmed that the period of a pendulum is independent of its amplitude. He observed the gentle swaying of a sanctuary lamp in the cathedral at Pisa. Using his pulse as a timer he found that successive oscillations were made in equal times, regardless of the amplitude. He later verified these observations in his laboratory.

The simple pendulum can be used to calculate $\boldsymbol{g}$ at any place by measuring T and l for a pendulum oscillating at that place. Countless thousands of such measurements have been made in the course of geophysical prospecting.

Figure 5.29



Example

The period of a simple pendulum 50.0 cm long is 1.42 s. Determine the acceleration due to gravity at that location.

Solution

$$T = 2\pi\sqrt{\frac{l}{g}}, \text{ or } T^2 = 4\pi^2\frac{l}{g}$$

$$\text{Hence } \boldsymbol{g} = \frac{4\pi^2 l}{T^2} = \frac{4\pi^2 \times 0.50}{1.42^2} = 9.79 \text{ m s}^{-2}$$

Activity 5.6 A SPRING PENDULUM

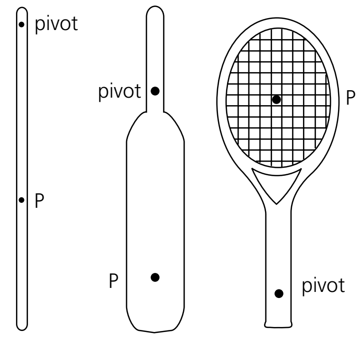
Make a pendulum out of a spring instead of a piece of string. Set it swinging and you'll soon see that T_2 is greater than T_1 as it bobs up and down as well as oscillating back and forth. The motion is fascinating. It almost makes you go to sleep.

Activity 5.7 THE SWEET SPOT

Any object that can vibrate like a pendulum is called a **physical pendulum** as distinct from a **simple pendulum**, which is a bob on a string. A wooden ruler, a cricket bat and a squash racquet can oscillate back and forth if allowed to pivot.

1. Suspend a **metre ruler** on a pin or nail through the hole in its end. Make sure it can vibrate freely. Set it in motion and measure the time it takes to make 10 swings. Calculate its period (T) and then calculate the effective length (l) using the pendulum formula. Assume g = 9.8 m s^{-2}. Mark this distance on the ruler. It is probably at about the 60 cm mark. This is called the centre of oscillation or centre of percussion (Latin *percussio* = 'striking'). You'll see why in the next part.
2. Repeat the above but use a **cricket bat** this time. Use two pins stuck into the handle at a point where your main grip would be and suspend the bat between the backs of two chairs. The pins can act as a pivot (Figure 5.30). Time it for 10 swings and calculate the effective length. Mark the centre of percussion (P). This is also called the 'sweet spot' because there is no sting in your hands if you hit the ball at this point. If the ball hits at any other point, the bat rotates about some other point than P, which accounts for the sting.
3. Try the same for a **squash racquet**. An effective length of 49 cm is common, which puts the sweet spot right at the middle of the head area. However, why do some world champion players hold their racquets where the grip joins the shaft? The answer is that the racquet is not rigid but flexes like a guitar string about the midpoint of the shaft while the end of the handle stays still. That's where they grip it to avoid the jarring. But designers also have to consider the power centre — the point at which maximum power is transferred to the ball. This is another complication that also applies to cricket, baseball and softball bats. This will be discussed further in Chapter 8, Momentum.
4. Over the past few years the sweet spot in **tennis racquets** has become higher up the head of the racquet. As a result, players can reach higher for the ball when they serve, opening up more of the opponent's court. This is a huge advantage because players can smack the ball that much harder instead of aiming carefully. For example, the world's fastest servers can now reach more than 200 km/h — speeds that were unheard of several years ago. Commentators have argued that speeds over 200 km/h are basically unplayable (and therefore boring) and that tennis balls need a 20% diameter increase to slow the maximum speed to a playable 180 km/h. If you can get hold of an old tennis racquet and a new one, compare the position of the sweet spots by the pendulum method. Is the above assertion correct?

Figure 5.30
The 'sweet spot'. The centre of percussion can be measured experimentally.



Questions

21 Determine the period of a pendulum with a length of 67.2 cm at a place where **(a)** g = 9.81 m s^{-2}; **(b)** g = 9.78 m s^{-2}.

22 **(a)** If you were accelerating upwards in a lift at 1.5 m s^{-2} what would the apparent acceleration due to gravity be?
(b) What would the period of oscillation of a 30 cm pendulum be in this lift?
(c) If the period of oscillation was 0.95 s, what acceleration upward would the lift be undergoing?

NOVEL CHALLENGE

Two side-by-side pendulums are oscillating. One has a period of 6 s and the other a period of 7 s.
If the bobs are touching at one time, how much longer must you wait until they come together again?

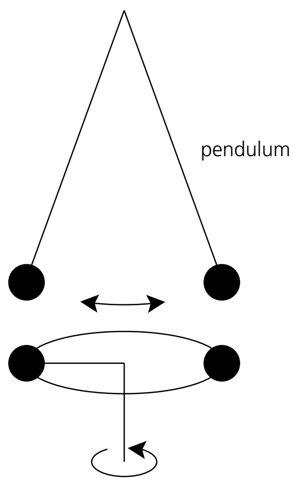
NOVEL CHALLENGE

You have been asked by your employer to write an instruction manual for a swing set in which you have to explain how a user can make it go higher.
What would you say? Now explain the physics behind your instructions.

SHM AND CIRCULAR MOTION COMPARED 5.8

There is a very clear relationship between SHM and circular motion. Galileo was the first person to make observations in this respect. In 1610 he was using his newly constructed telescope and discovered the four principal moons of Jupiter. Over weeks of observation, each moon seemed to be moving back and forth past the planet in what we now call simple harmonic motion. This has been confirmed by plotting his data. But actually, the moons move in an essentially constant circular motion around Jupiter. What Galileo saw — and what you can see with a pair of binoculars — is this circular motion edge on, and they look as though they are oscillating back-and-forth beside the planet.

Figure 5.31



Observing the two motions together

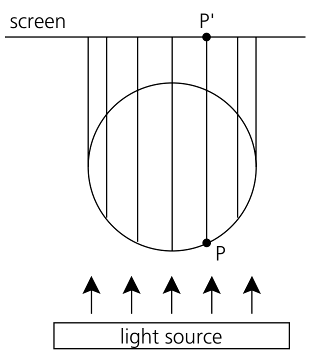
If you could set a pendulum swinging above an object moving in a horizontal circle at constant speed, you could get the two moving side-by-side if the speeds were right (Figure 5.31).

If a light was used to project an image of the oscillating objects on to a wall, the shadows of the two objects would move in exactly the same manner (Figure 5.32).

Consider point P making a complete revolution of the circle in Figure 5.32. The point P′ makes a complete oscillation on the straight line of the pendulum. Equally spaced points on the circle project as points on the line as shown. This illustrates that maximum acceleration occurs at the maximum amplitude of the pendulum, and minimum acceleration occurs when the amplitude is a minimum. This is a characteristic of SHM.

In more formal language: Simple harmonic motion is the projection of uniform circular motion on the diameter of the circle in which the circular motion occurs.

Figure 5.32



Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

*23 A boy sitting in a train carriage moving at constant velocity throws a ball straight up in the air.
(a) Will the ball fall behind him, in front of him or into his hands?
(b) What happens if the train accelerates while the ball is in the air?
(c) What happens if the train turns a corner while the ball is in the air?

*24 A motorcycle is driven off a cliff at a horizontal velocity of 15 m s^{-1} and takes 2.5 seconds to reach the ground below. Calculate **(a)** the height of the cliff; **(b)** the distance out from the base of the cliff that the motorcycle lands; **(c)** the impact velocity.

*25 When a wedding ring is thrown horizontally out of a fifth floor window 15 m off the ground, it lands 7.5 m out from the base of the building. Calculate **(a)** the throwing speed; **(b)** the impact velocity; **(c)** how long the marriage will last.

*26 A golf ball is hit by a club and moves off with a velocity of 30 m s^{-1} at an angle of 55° to the horizontal. Find the following:
(a) The initial vertical and horizontal components of the velocity.
(b) The maximum height reached.
(c) The time of flight.
(d) The horizontal range.

***27** A soccerball is kicked off the ground at an angle of 20° to the horizontal. It moves away at 30.0 m s^{-1}. Calculate **(a)** the vertical velocity after 0.5 s; **(b)** the velocity of the ball after 1.0 s; **(c)** the maximum height reached; **(d)** the time of flight; **(e)** the range of the ball.

****28** The world record for fresh hen's egg throwing is 96.90 m, set in 1981. Assuming no air resistance, what would have been the **(a)** throwing speed; **(b)** elevation angle; **(c)** maximum height; **(d)** time of flight?

***29** A car of mass 2250 kg is travelling around a circular track of radius 90 m at a constant speed of 30 m s^{-1}. Calculate **(a)** the centripetal acceleration; **(b)** the centripetal force; **(c)** what time it takes to complete one lap.

****30** An aviator, pulling out of a dive, follows the arc of a circle and is said to have experienced 3 '$\boldsymbol{g}$'s. Explain what this means.

***31** In the Bohr model of a hydrogen atom, an electron orbits a proton in a circle of radius 5.28×10^{-11} m with a speed of 2.18×10^{6} m s^{-1}. What is the acceleration of the electron in this model?

***32** Convert the following:
- **(a)** 1 rad to degrees;
- **(b)** 8.5 rad to degrees;
- **(c)** 90° to rad;
- **(d)** 5 rpm to rad s^{-1};
- **(e)** 100 rad s^{-1} to rev per second;
- **(f)** 2 revolutions of a 50 cm radius circle to metres;
- **(g)** 20 rad s^{-1} of a 1.5 m radius circle to linear m s^{-1}.

***33** An amusement park Ferris wheel moves in a horizontal circle of 15 m radius and completes five turns every minute.
- **(a)** What is the acceleration of a passenger at **(i)** the highest point; **(ii)** the lowest point?
- **(b)** If the passenger has a mass of 65 kg, what would her apparent weight be at these two points?

****34** The maximum breaking strain of a piece of cord is 250 N. What is the maximum rpm at which the line can retain a 3 kg mass swung in a 1.8 m radius circle?

****35** A flywheel of radius 65 cm is rotating at 2000 rpm. Calculate **(a)** the angular velocity; **(b)** the linear velocity of a point on the rim.

***36** A light spring stretches by 20 cm when a mass of 200 g is hung vertically from it.
- **(a)** Calculate its spring constant.
- **(b)** When it is set oscillating, what would be its period?
- **(c)** What would be its frequency be?

***37** Determine the period of a pendulum with a length of 45.0 cm at a place where: **(a)** $\boldsymbol{g}$ = 9.805 m s^{-2}; **(b)** 9.785 m s^{-2}.

****38** When travelling upwards in a lift at constant speed a pendulum has a period of 1.30 s. When accelerating, however, the period becomes 1.22 s. Calculate **(a)** the length of the pendulum; **(b)** the acceleration of the lift. Assume $\boldsymbol{g}$ = 9.8 m s^{-2}.

****39** A centripetal force experiment was conducted to find relationships between some of the variables.

Part A was conducted to determine the relationship between centripetal force ($\boldsymbol{F}_c$) and velocity ($\boldsymbol{v}$) in horizontal circular motion. Using the experimental set-up as shown in Figure 5.33, a rubber stopper was swung at constant speed in a horizontal circle. The hanging mass, which provided the centripetal force, was varied and the time for 10 complete revolutions of the rubber stopper was noted. In all cases the radius of revolution (r) was kept at 1.5 m and the same rubber stopper was used each time. The mass of the rubber stopper (m_s) was 50 g.

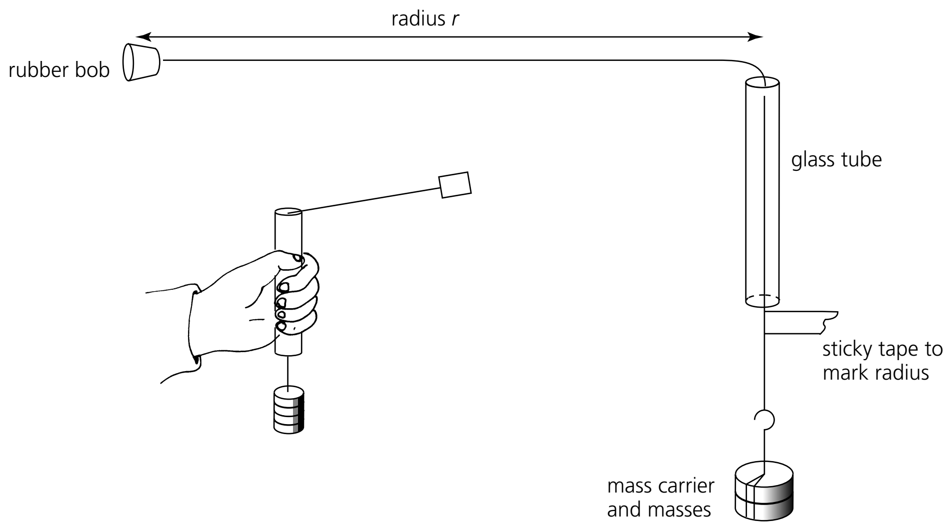


Figure 5.33
For question 39.

The results shown in Table 5.3 were obtained.

Table 5.3 CENTRIPETAL FORCE DATA (PART A)

m_h(g)	m_s(g)	RADIUS (m)	TIME FOR 10 REVOLUTIONS (s)
50	50	1.5	24.6
100	50	1.5	17.4
150	50	1.5	14.2
200	50	1.5	12.3
250	50	1.5	11.0

(a) Calculate the centripetal force ($\boldsymbol{F}_c$) provided by the hanging mass for each stage.
(b) Calculate the period (T) and the linear velocity ($\boldsymbol{v}$) of the rubber stopper for each stage.
(c) Plot $\boldsymbol{F}_c$ vs $\boldsymbol{v}$ using the x-axis for $\boldsymbol{v}$.
(d) Suggest a possible relationship between $\boldsymbol{F}_c$ and $\boldsymbol{v}$. Plot the appropriate data to confirm or refute the suggested relationship. Does it agree with the centripetal force formula?

Part B Relationship between radius and velocity. The above experiment was repeated with a 100 g rubber stopper. This time the hanging mass was kept constant at 100 g while the radius of revolution was varied. Again, the time for 10 revolutions was measured and the data recorded in Table 5.4.

Table 5.4 CENTRIPETAL FORCE DATA (PART B)

m_h(g)	m_s(g)	RADIUS (m)	TIME FOR 10 REVOLUTIONS (s)
100	100	0.8	17.9
100	100	1.0	20.1
100	100	1.2	22.0
100	100	1.5	24.8

(e) Calculate and plot r vs $\boldsymbol{v}$ using $\boldsymbol{v}$ for the x-axis again.
(f) Suggest a relationship and plot to confirm.
(g) Does it agree with the formula?
(h) What would the shape of an $\boldsymbol{F}_c$ vs r graph look like (r on the x-axis) if m_s and $\boldsymbol{v}$ were kept constant?

**40 An experiment was carried out to establish the relationship between length and period of a simple pendulum. A brass bob was tied to a length of cotton thread and as its length was increased, the time for 10 oscillations was noted. The results are as follows:

Length (cm)	20.0	25.0	35.0	40.0	45.0
Time for 10 swings (s)	9.0	10.0	11.9	12.7	13.6

(a) Plot a graph to establish the possible relationship.
(b) Plot another graph to confirm the suggested relationship.
(c) From either graph, determine the time for 10 swings if the length was **(i)** 30.0 cm; **(ii)** 60 cm.

**41 The following data (Table 5.5) were taken from *Overlander 4WD* magazine's road test of some four wheel drives.

Table 5.5 FOUR WHEEL DRIVE ENGINE DATA

	LANDCRUISER	LAND ROVER	PAJERO	NISSAN PATROL
Capacity	4.477 L	3.528 L	2.972 L	4.169 L
Maximum power	158 kW at 4600 rpm	114 kW at 4700 rpm	109 kW at 5000 rpm	129 kW at 4000 rpm
Maximum torque	373 Nm at 3200 rpm	271 Nm at 3000 rpm	234 Nm at 4000 rpm	330 Nm at 3200 rpm

Comment critically on the following assertions by referring to the data.
(a) The bigger the engine capacity the greater the power and torque.
(b) Smaller capacity engines have to rev at a higher rate (rpm) for their maximum power and torque than do bigger engines.
(c) Engines have to turn at a higher rpm to get maximum power than they have to for maximum torque.

Extension — complex, challenging and novel

***42 A dart is thrown horizontally towards a bull's eye of a dart board but it strikes the 3 on the bottom of the board directly underneath, 0.19 s later (Figure 5.34). What is the distance from the bull's eye to the 3?

***43 An arrow is fired off a 50 m cliff at an angle of 20° above the horizontal. If it has an initial velocity of 35 m s^{-1} and strikes the rocks below, calculate **(a)** the time of flight; **(b)** the impact velocity; **(c)** how far out from the base of the cliff the arrow strikes the ground.

***44 In the 1968 Olympics in Mexico City, Bob Beamon shattered the world long jump record with a jump of 8.90 m. His speed on take-off was measured at 9.5 m s^{-1}, about equal to that of a sprinter. How close did he come to achieving maximum range for this speed in the absence of air resistance? The value of $\boldsymbol{g}$ in Mexico City is 9.78 m s^{-2}.

***45 A plane, diving at an angle of 53.0° to the vertical, releases a projectile at an altitude of 730 m. The projectile hits the ground 4.50 s after being released (Figure 5.35).
(a) What is the speed of the plane?
(b) How far did the projectile travel horizontally during its flight?
(c) What is the impact velocity?

Figure 5.34
For question 42.

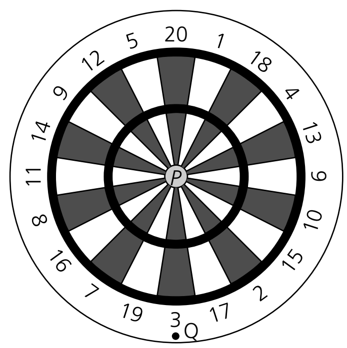
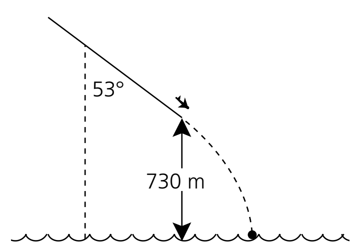


Figure 5.35
For question 45.



*****46** A person stands against the vertical walls of a cylindrical rotor in an amusement park. As it rotates, she feels pressed against the walls of the rotor and she remains suspended there as the floor moves away. The centripetal force is the normal force with which the wall pushes on the person.

(a) If the rotor has a radius of 2.1 m and the coefficient of friction between the person and the wall is 0.40, calculate the minimum speed of the rotor to just keep the person suspended on the wall.

(b) If the person has a mass of 49 kg, calculate the centripetal force acting on her.

*****47** A pilot of mass 80 kg who has been diving his plane vertically downwards with a velocity of 120 m s^{-1} pulls out of his dive by changing his course to a circular path of radius 800 m. If he maintains his constant speed,

(a) what will be his maximum acceleration;

(b) if he can stand 4.5 ***g*** without blacking out, will he remain conscious;

(c) what is the maximum force that his seat exerts on him?

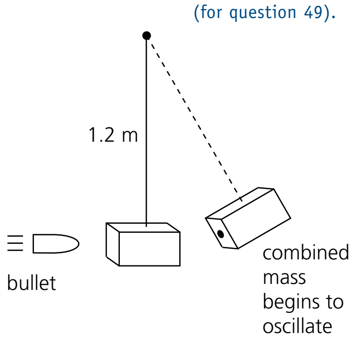
*****48** For a simple pendulum undergoing four oscillations:

(a) Draw graphs showing the relationships between the following variables **(i)** displacement vs time; **(ii)** velocity vs time; **(iii)** acceleration vs time; **(iv)** velocity (*y*-axis) vs displacement (*x*-axis). Remember that ***s***, ***v*** and ***a*** are vector quantities so have + and – direction.

(b) Repeat the question above but this time imagine that the pendulum is 'damped', that is, friction causes it to slow down as it moves.

(c) The ***v*** vs ***s*** graph for damped motion is said to be a 'strange attractor'. Look up a book on chaos theory to find out what this means.

Figure 5.36
The ballistic pendulum (for question 49).



*****49** A bullet of mass 10.0 g is fired into a 'ballistic pendulum' — a wooden block, which has a mass of 1.000 kg. The wooden block is suspended from a string 1.20 m long as shown in Figure 5.36. The bullet enters the stationary block and remains embedded in it. Using the value of 9.80 m s^{-2} for ***g***, calculate the period of the pendulum.

*****50** *Courier-Mail* correspondent Dave Barry wrote about an exciting new sport taking off in Florida, USA. It's called 'car bowling' where you go up in an airplane and drop bowling balls on cars. He wrote: 'Women think — "You drop what, on what, from what?" whereas men think "When can I do this?" You fly over an old car on a private runway at 145 km/h at an altitude of 20 m and attempt to hit the car with a bowling ball. The beauty of car bowling is that even if you miss, you get to watch a bowling ball bounce along a runway. It's amazing.'

(a) How far horizontally should you be from the car when you drop the ball?
(b) What would your 'sight angle' be at this point? (Sight angle is the angle between the line to the target and the vertical at the drop point) **(c)** Assume that the impact angle on contact with the runway equals the launch angle after contact, but with a 20% reduction in speed. Calculate **(i)** the maximum height and **(ii)** the distance the ball travels before its next impact.

TEST YOUR UNDERSTANDING

(Answer true or false)

- The period of a pendulum depends on the amplitude.
- A pendulum accelerates through the lowest part of its swing.
- An object moving in a circle with constant speed has no acceleration.

CHAPTER 06

Astrophysics

From the very earliest days, humans have looked into the sky and wondered what it's all about. Thousands of years ago priests in Babylon (now present-day Iraq) stood gazing into the night sky, not realising just how big it was. All sorts of theories, all sorts of myths and legends have grown out of attempts to understand how the universe works.

The big questions on everyone's mind concerned the solar system (Sun and planets) and the universe in general. The old **geocentric** view of our solar system said that the Sun went around the Earth (Greek *geo* from *gaia* = 'Earth' and *centro* = 'centre'). The modern view belongs to the Polish astronomer and priest Nicolas Koppernigk (**Copernicus**, as he was better known in Latin), who published his **heliocentric** or Sun-centred theory in 1540. Not that it was a new concept, for the Greek philosophers Heraclides and Aristarchus put forward a similar view in 300 BC, but after they were threatened with death they kept quiet.

The modern view of the entire universe came much later. The general opinion up until the 1920s suggested that the universe was infinite in size and in a steady unchanging 'static' state, made up of fixed stars that had always shone and would continue to shine forever. This view was well entrenched — even Einstein believed it! But in 1929 an astronomer made a finding that was to shake the foundations of this steady state model forever. His name was Edwin Hubble and his story follows later. The **steady-state** theory clung on until the 1970s when it died and was buried.

Like those before us, do you ever wonder about these:

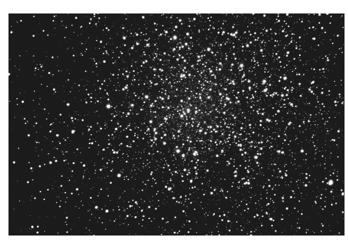
- What makes the Universe tick? How do the four forces work together?
- What is the Universe made of? We don't know what's out there.
- Was Einstein's anti-gravity theory really a great mistake or ahead of its time?
- Why do we live in a three-dimensional world; is it just a fluke?
- Can we travel in time and could we come back?
- Can black holes collapse to infinite density? How would you know?
- Where does consciousness come from; where does life come from?
- Are we alone?

6.1 THE NATURE OF THE UNIVERSE

Here's what *The Hitch Hiker's Guide to the Galaxy* has to say about the size of the universe:

> Space is big. Really big. You just won't believe how vastly hugely mindbogglingly big it is. I mean you may think it's a long way down the road to the chemist, but that's peanuts to space ... It's just so big that by comparison, bigness itself looks small.

Photo 6.1
The sky at night is almost black — but so what?



Astrophysicists agree. They can also answer some of the other questions above.

- The universe is between 10 and 15 billion years old, with most scientists agreeing on 13.4 billion years (a billion is 10^9).
- The remotest object from us is the quasar PC 1247+3406 at 13 200 million light years (1.25×10^{23} km). One light year is the distance light travels in one year or 9.46×10^{12} km. Hence, the edge of the universe is believed to be 15 000 million light years from us.

Some of the other questions will be answered later in this chapter. Some may never be answered but physicists will keep on trying. Many laws have been developed in this quest.

To appreciate the size of the universe, imagine the Earth is the size of a pinhead. The Sun would be the size of a grape about $1\frac{1}{2}$ metres away. Jupiter would be a pea 8 metres away, and Pluto a grain of dust 70 m in the distance. That's our solar system. Whew!

Now imagine the whole solar system shrunk down so that the Sun is now a pinhead. The Earth would orbit a few centimetres away, and Pluto about 60 cm away. On this scale our nearest star system — containing Proxima Centauri — is 3 km away and the size of a tiny sand grain. Other stars are also like sand grains and they reach out a distance equal to the distance from us to the moon. That's our galaxy. Big in anyone's language!

Lastly, imagine our galaxy shrunk down to the size of a dinner plate. Our nearest neighbouring galaxy is Andromeda — another dinner plate just a few metres away. The edge of the visible universe is many kilometres in every direction. But extending past that are more galaxies that we can't see because light has yet to reach us. Scientists believe that there are approximately 100 billion galaxies, with each galaxy containing between 100 and 200 billion star systems. That's our universe. To better understand the universe as it is today, you have to appreciate three fundamental observations: Olbers's paradox, Hubble's law, and the Cosmic Microwave Background Radiation, of which more later.

Astronomical distances

It takes light about 10 billion years to get from the edge of the observable universe to us. That's a huge distance and the units metre and kilometre seem inadequate. Astronomers use the unit megaparsec (Mpc) for distance. A parsec (pc) is a distance based on how far away a star would be if it appeared to change position by an angle of one second when viewed from the Earth at 6-month intervals. It sounds complex but astronomers assure us it is eminently suitable for their work. They also use light-years (ly) for distance: the distance light travels in a year. They use km s^{-1} for velocity or express it as a fraction of the speed of light (c).

1 light-year (ly)	9.47×10^{15} m
1 parsec (pc)	3.262 ly
	3.09×10^{16} m
1 megaparsec (Mpc)	1 million parsec
	3.262×10^{6} ly = 3.262 Mly
	3.09×10^{19} km
Speed of light (c)	3×10^{8} m s^{-1}
Temperature (T)	K = °C + 273

Example

The Hydra galaxy is 1960 million light-years (Mly) away and has a radial velocity of 60 500 km s^{-1}. Convert the distance to Mpc and the speed to units of c.

Solution

Mpc = Mly/3.262, hence Mpc = 1960/3.262 = 600 Mpc

$$\boldsymbol{v} \text{ in units of } c = \frac{\text{m s}^{-1}}{3 \times 10^{8}} = \frac{60\,500 \times 1000}{3 \times 10^{8}} = 0.20 \text{ c}$$

Olbers's paradox

In 1823, a German astronomer Heinrich Olbers stumbled on a contradiction that could not be easily explained. The following activity poses this contradiction.

SR Activity 6.1 THE NIGHT SKY

You don't really need to carry out this experiment — a *gedanken* (thinking) experiment will do!

Have a look at the sky at night. Is it black or white? Of course it is mostly black (see Photo 6.1), with about 4000 tiny stars twinkling away; but why doesn't the night sky look uniformly bright? If there were an infinite number of stars which had been glowing for an infinite time, no matter where you looked you'd see a star and so the night sky should be ablaze with light. But it's not!

We now know that the old-fashioned idea of an infinite, static universe is simply wrong. The universe has a finite age, and is not just three-dimensional as we perceive things on Earth. Because only 10 billion years have elapsed thus far, we can only observe stars out to a large, but strictly finite, distance of 10 billion light-years or so. This 'observable universe' contains a large but finite number of stars, about 1000 billion (10^{12}). These stars contribute to the observed brightness of the night sky, which glows very faintly.

SR Activity 6.2 LIGHT AT A DISTANCE

Here's a good experiment you could try using a computer-based laboratory such as the TI graphing calculator and the CBL. There are plenty of other ways to do it as well.

Set up equipment as shown in the Figure 6.1.

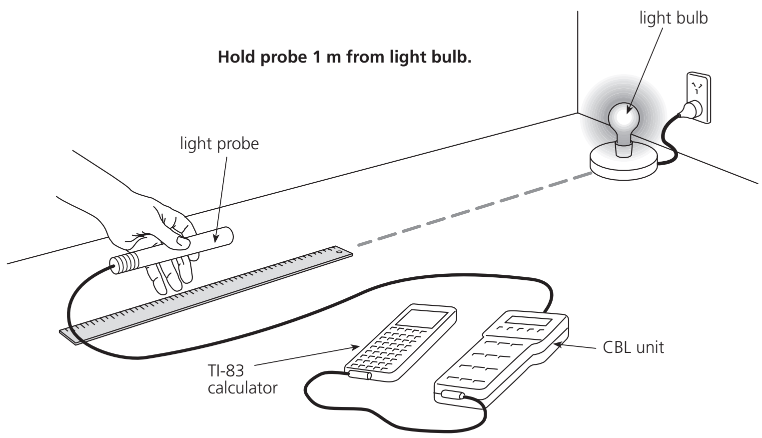


Figure 6.1
Apparatus to measure how light intensity varies with distance.

1. Mark off distances of 1 m and 2 m from the light socket. Then divide the distance into 10-centimetre intervals between the one-metre and two-metre marks.
2. While you are taking intensity readings during the activity, the light sensor should be pointed directly at the illuminated bulb with the end of the sensor held a certain distance from the bulb, as specified in the calculator program.
3. Darken the room, with the exception of the light source.
4. Collect light intensity data for different distances.
5. The data you collected will be modelled with a power relation of the form $y = ax^b$. First, you will need to find the values of A and B. The rest is up to you and your graphing calculator. Good luck.
6. According to scientific theory, the correct model for light intensity against distance is an inverse square relationship. This relation is expressed mathematically as: $y = a/x^2$ (inverse square law).

   If this equation is expressed in the form $y = ax^b$, what would be the value of b? Is this consistent with the models you found earlier?

Hubble's law

During the 1920s astronomers looked at starlight through spectrometers and noticed that the spectral lines of elements such as hydrogen and helium seemed to be occurring at longer wavelengths than normal (more on this in Chapter 29). If you look at the centre colour photos in this book you will see the normal spectra of many elements. The shift in wavelength was towards the red end of the spectra and so the term '**red shift**' was coined for this phenomenon. Physicists deduced that the shift in wavelength meant that the star was moving relative to the observer on Earth, similarly to the way the sound of an ambulance siren seems to change as it moves towards or away from you. This is called the **Doppler effect** and is discussed fully in Chapter 16 (section 16.3). Red shift is also treated comprehensively in Chapter 29.

In 1929, US astronomer Edwin Hubble used his 100 inch (2.5 m) diameter telescope to show that the universe is expanding. This was a monumental breakthrough. And if galaxies are moving away from each other, there must have been a time when they were all together. This was about 10 billion years ago — the time of the Big Bang — when the whole universe was the size of a dot (.).

Hubble combined his knowledge of galaxy red shifts with an estimate of the distance to these galaxies, and determined that galaxies more distant from us were moving away from us more rapidly than closer galaxies. This relationship has become known as **Hubble's law**.

Mathematically the law is written as $v = H_0D$, where v is the **radial velocity** — that is, how fast the galaxy is moving directly away from us; D is the distance to the galaxy; and H_0 is the **Hubble constant**. The radial velocity is sometimes called the ***recession velocity*** (Latin *recessio* = 'recede' or 'withdraw').

SR Activity 6.3 HOW OLD IS THE UNIVERSE?

A plot of distance and radial velocity gives the Hubble constant, which is a measure of the rate of the expansion of the universe. It can be used to calculate the age of the universe.

1 Use the data in Table 6.1 to plot distance (Mpc) on the *x*-axis and radial velocity (km s^{-1}) on the *y*-axis. The table also shows the laboratory values for the three common spectral lines (e.g. H_α is 656.3 nm).

TABLE 6.1

GALACTIC CLUSTER	SPECTRAL LINES (nm)			RADIAL VELOCITY (km s^{-1})	DISTANCE (Mly)	DISTANCE (Mpc)
	Ca-K (393.4)	Ca-H (396.9)	Hα (656.3)			
Virgo	394.9	397.0	656.3	1 140	38	12
Perseus	400.5	397.1	656.4	5 430	179	55
Hercules	407.0	397.4	656.5	10 400	360	110
Pegasus II	410.2	397.5	656.6	12 800	490	150
Ursa Major 1	413.1	397.9	656.8	15 000	750	230
Gemini	424.1	398.2	657.0	23 400	980	300
Ursa Major 2	446.4	398.9	657.3	40 400	1 500	460
Hydra	472.7	399.5	657.6	60 500	1 960	601
3C295	574.4	404.4	660.1	138 000	5 700	1 747

2 Calculate the slope of the line produced by these points (= Hubble constant, H_0). It should be given in (km/s)/Mpc. Most modern estimates put it somewhere between 70 (km/s)/Mpc and 75 (km/s)/Mpc. Show your calculations.

3 We can use this to determine the age of the universe. At the instant of the Big Bang, all the matter in the universe was together. It has had all of the intervening time to fly apart to its present positions. A galaxy that is more distant from us is so because we have been separating from it at a faster rate in that time. (Remember, Hubble's law states that the farther a galaxy is from us, the more rapidly it is receding from us.) Choose a point on the line in your graph that you used to find Hubble's constant. This point represents the distance and recession velocity of a galaxy following Hubble's law. The only problem is that Mpc and km are different units, and seconds aren't terribly good units for measuring the age of the universe. Convert Mpc to km (1 Mpc = 3.09×10^{19} km).

Use the relationship $\boldsymbol{v} = d/t$ or $t = d/\boldsymbol{v}$ to calculate the time taken for the galaxy to cover the distance at the given velocity. Convert seconds to years. It should be about 10 billion years. Alternatively, take the reciprocal of the Hubble constant and multiply by 3.09×10^{19} km to get the age in seconds. Convert to years.

Example

Using the data for the cluster Virgo, calculate **(a)** the Hubble constant; **(b)** the age of the universe.

Solution

(a) $H_0 = \frac{\boldsymbol{v}}{D} = \frac{23\ 400 \text{ km/s}}{300 \text{ Mpc}} = 78 \text{ km/s/Mpc}$

(b) $300 \text{ Mpc} = 300 \times 3.09 \times 10^{19} = 9.27 \times 10^{21} \text{ km}$

$$t = \frac{9.27 \times 10^{21} \text{ km}}{23\ 400 \text{ km/s}} = 3.96 \times 10^{17} \text{ s} = 1.256 \times 10^{10} \text{ years} = 12.56 \text{ billion years.}$$

Alternatively: $t = \frac{1}{H_0} \times 3.09 \times 10^{19} \text{ km} = 3.96 \times 10^{17} \text{ s} = 1.256 \times 10^{10}$ years.

Some people say that the Earth was created 6000 to 8000 years ago, because that's what they get when they add up all of the 'begats' in the genealogy of the Old Testament. However, many people agree that the 'days' of creation are not literal 24-hour days but could be translated from the Hebrew as billion-year 'eras'. You decide — but be aware that scientists reject anything less that about 10 billion years for the age of the universe and $4\frac{1}{2}$ billion years for the age of the Earth.

Cosmic microwave background radiation

The entire universe exists in a sea of background radiation. During the early days of creation, a great deal of radiation was present. As the universe continues to expand and cool, this radiation should still be present, although at a stretched wavelength due to the expansion. The presence of this 'microwave' radiation is unmistakable evidence of the Big Bang fireball of creation. In 1955, astronomer George Gamow predicted background radiation of 5 K, which was subsequently confirmed by Arno Penzias and Robert Wilson at Bell Laboratories in New Jersey USA in 1965. The Cosmic Background Explorer (COBE) satellite measured the radiation more accurately in 1989 — to a wavelength of 1 mm, equating to a temperature of 2.3 K. (See Photo 6.5, page 151.)

ACTIVITY 6.4 THE TEMPERATURE OF DEEP SPACE

SR

In the early 1900s, German scientist Max Planck found that the peak wavelength of black-body radiation was related to temperature by the formula: $\lambda = 0.2\ hc/kT$ where h = Planck's constant = 6.63×10^{-34} Js, k = Boltzmann's constant = 1.38×10^{-23} J K^{-1} (see Section 11.4), c = speed of light = 3×10^8 m s^{-1}, and T = Kelvin temperature.

Example

Calculate the peak wavelength of light emitted by the brightest visible star in the sky (Sirius) with a temperature of 9000 K.

Solution

$$\lambda = \frac{0.2 \times h \times c}{k \times T} = \frac{0.2 \times 6.63 \times 10^{-34} \times 3 \times 10^{8}}{1.38 \times 10^{-23} \times 9000} = 3.20 \times 10^{-7}\ \text{m} = 320\ \text{nm}$$

(a) Use the formula to confirm that the background temperature of outer space is 2.73 K, given that the Cosmic Background Explorer (COBE) satellite measured the cosmic background radiation to have a wavelength of 1 mm. This provided dramatic evidence in 1989 to support the prediction of the Big Bang theory.

(b) Show, by dimensional analysis (cancellation), that the unit for the right-hand side of the equation is metres.

(c) A simplification of Planck's law is known as Wein's law. Wein expressed it as $\lambda = 0.0029/T$. Show that it is the same as the original formula.

(d) In 1992, COBE detected *fluctuations* in the background radiation, and this sent tingles of excitement down astronomers' backs. Why was this, and where does the 'recombination' of 300 000 years ago fit in? Off to the library.

Questions

1 The Leo cluster is 251 Mpc away. Prove that this is equal to a distance of 7.8×10^{24} m.

2 Hercules has a radial velocity of 0.035c. What is this in km/s?

3 The Bootes cluster is 457 Mpc away from us. Based on an average Hubble constant of 75 km/s/Mpc, calculate the recession speed of Bootes.

4 If the Hubble constant was found to be 100 km/s/Mpc, what would the age of the universe be?

5 Calculate the wavelength of maximum intensity of the red star Proxima Centauri, which is 4.3 times the diameter of the Sun and has a temperature of 2870 K.

6 The Stefan–Boltzmann law can be used to calculate the power output (luminosity, L) of a star based on its radius (R) and temperature (T): $L = (7.125 \times 10^{-7})\ R^2\ T^4$, where L is in watts, R in metres and T in kelvins.

   (a) Our star, the Sun, has a temperature of 5775 K and a radius of 1.39×10^6 km. Calculate its luminosity.

   (b) The star Rigel has a temperature three times that of the Sun, and a luminosity 64 000 times that of the Sun (one very bright star). If the Sun's temperature is 5775 K, calculate Rigel's radius.

7 Some stars are red and some are blue. Which are hotter? Why?

8 Comment critically: 'A star with a temperature of 0°C will not give off any radiation.'

9 Does a star at 6000 K emit twice as much radiation as it does when it drops to 3000 K? Explain.

Ch06-Walding 4th 25/8/04 11:08 AM Page 140

6.2 HISTORY OF THE UNIVERSE

'This is the way the world ends
Not with a bang but a whimper' — T. S. Eliot

The best picture we have at present suggests that the past, present and future of the universe can be arranged into the five 'ages' proposed by creative American astrophysicists Fred Adams and Greg Laughlin. It starts with the Big Bang and ends up expanding forever 'like a whimper', as poet T. S. Eliot put it. Like all scientific theories, it could be wrong, but it does explain why the universe is like it is today. Many people opposed to these theories say 'You can't re-create the Big Bang so how can it be scientific?' However, the theory is consistent with all known laws and principles of physics, and many parts of the theory can be tested experimentally. It also offers testable predictions about what we should find as time goes by. All that makes it scientific.

Admittedly, we can't tell what is happening beyond a certain distance (>10 billion light-years away) because light has not had time to reach us. We call this the 'visible horizon'. As we get older this horizon will get further away. Maybe it's where the wild things are.

The Primordial Era: $t = 0$ to 3 minutes

The very beginning

About 13 billion years ago, the universe was just a point in space with infinite density and temperature. This was the beginning of time ($t = 0$) as we know it. We have no understanding yet of what the universe was like inside this point. The four fundamental forces that we know today (gravity, electromagnetic, weak and strong nuclear) were rolled up into one 'super force' known as the Grand Unified Theory (GUT) force.

The Big Bang and inflation

An enormous explosion ('the Big Bang') occurred at $t = 0$, and after 10^{-43} s had elapsed the universe began to expand and cool at a fantastic rate. Adjacent points in space rushed away from each other at speeds greater than that of light, and the small dot (.) that was the universe inflated about 10^{30} times — all within a period of about 10^{-35} s (see Figure 6.2).

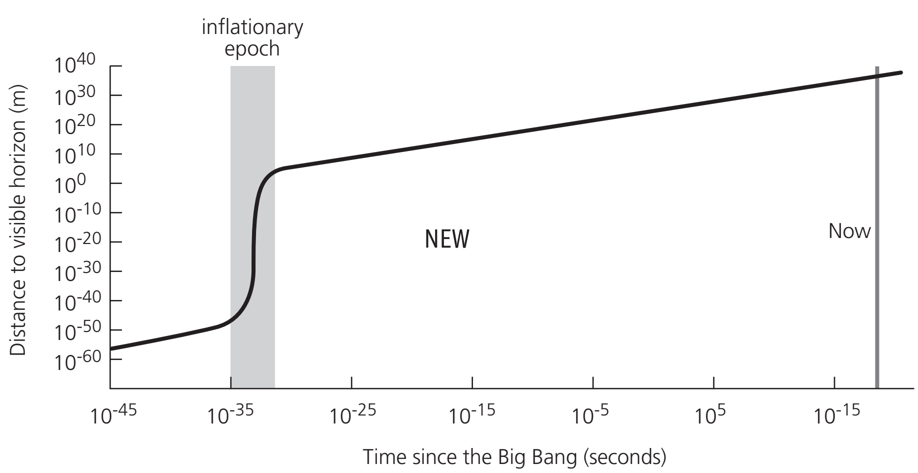


Figure 6.2
The expansion of the universe throughout time. Note the exponential scale (powers of 10) on the *x*-axis.

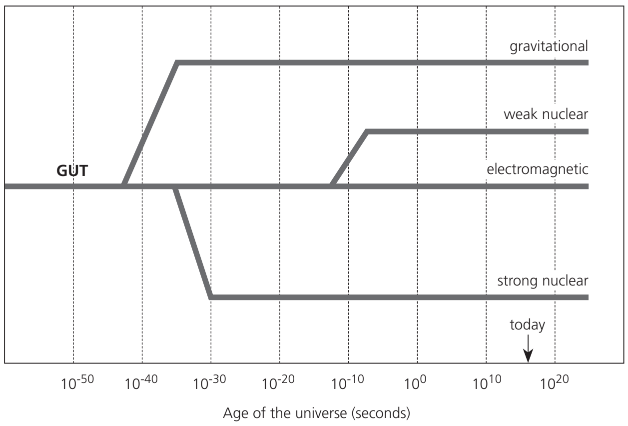
In this very short time the temperature fell from 10^{32} K to 10^{20} K. Most of the energy of the universe at the time of **inflation** was in the form of electromagnetic radiation because it was like a blast furnace — too hot for atoms and molecules. It was too hot for even protons and neutrons. During this period the GUT forces began to decouple (separate out) into the four fundamental forces (Table 6.2). Chapter 29 deals with these forces in more detail.

Table 6.2

FUNDAMENTAL FORCE	TIME OF SEPARATION (S)	STRENGTH	TEMPERATURE (K)	PARTICLE THAT CARRIES THE FORCE
Gravitation	10^{-43}	10^{-38}	10^{32}	graviton
Strong nuclear	10^{-35}	10^{0}	10^{27}	gluon
Weak nuclear	10^{-12}	10^{-13}	10^{15}	W and Z
Electromagnetic	10^{-12}	10^{-2}	10^{15}	photon

Figure 6.3 shows the decoupling graphically.

Figure 6.3
Decoupling of the GUT forces into the four forces we know today.



Baryogenesis

The universe was now a vast sea of radiation with a small mixture of **quarks** and other particles called **gluons** which acted between these quarks. This froth of quarks and gluons is known as the Quark Gluon Plasma (QGP). Quarks consist of both ordinary matter and antimatter, with a slight excess of the former. For every 30 million antimatter quarks there were 30 million and one quarks made of matter. As the universe cooled, the matter and antimatter quarks annihilated each other, leaving the excess fraction of matter quarks to survive. This process is called **baryogenesis** (Greek *baros* = 'weight', *genesis* = 'origin'); the general name given to particles that form matter is *baryons*.

Physicists have speculated that in other universes there must be an equivalent excess of antiquarks so that the total amount of matter and antimatter are equal, as stipulated by the conservation law. During this period any quarks that came together were unable to combine to form larger particles (protons and neutrons) as the high-energy gamma rays blasted them apart. By the time the universe was 10^{-12} s old the temperature was down to 10^{15} K and the GUT forces had completely decoupled. It was now sufficiently cool for all quarks to condense (in groups of three) to form **hadrons** (such as protons and neutrons). Because there was an excess of quarks over antiquarks, more protons and neutrons formed than antiprotons and antineutrons. These particles and antiparticles continued to annihilate each other, but as the temperature dropped fewer and fewer particles and their antiparticles were created, which just left an excess of protons and neutrons.

These basic building blocks of matter, synthesised in the first microsecond of the universe's history, live not only to the present time, some ten billion years later, but will endure for another 10^{32} years — give or take a few years. You are made up of quarks from the Big Bang.

PHYSICS FACT

During 'inflation' the universe expanded at a rate faster than the speed of light. How can this be, if nothing can go faster than light? The answer is that it is space that is expanding without any objects moving. In the Big Bang model the distance between galaxies increases, but the galaxies don't move. Since nothing is moving, there is no violation of the restriction that nothing can move faster than light. Hence, it is quite possible that the distance between two objects is so great that the distance between them expands faster than the speed of light.

Ch06-Walding 4th 25/8/04 11:08 AM Page 142

Nucleosynthesis

The next major achievement of the early universe was the production of small compound nuclei. About one second after the Big Bang, when the temperature had dropped to 10 billion kelvin, the universe was cool enough to allow the fusing of protons and neutrons to synthesise (Greek *syn* = 'together', *tithenai* = 'to place') light atomic nuclei, mainly helium ($^{4}_{2}He$) with some traces of deuterium ($^{2}_{1}H$), and lithium ($^{7}_{3}Li$). This **nucleosynthesis** continued for about three minutes until the temperature of the ever-expanding universe dropped to a mere 1 billion kelvin. Nuclear reactions abruptly stopped and nucleosynthesis thus came to an end. Although huge amounts of helium and the other light nuclei were produced, about 75% of the observable mass of the universe remained as hydrogen nuclei (protons, $^{1}_{1}H$). This was a busy three minutes.

The Stelliferous Era: $t = 3$ minutes to 10^{14} years

Stelliferous comes from the Latin and means 'star-producing' (*stella* = 'star', *fero* = 'bring forth' or 'produce'). We are in the middle of this era right now. Our Sun ignited some 4.5 billion years ago (4.5×10^9 years) and has enough hydrogen to last another 6 billion years. We are all children of the stars because ten billion years ago every atom in our bodies was once near the centre of a star.

Once the explosive first three minutes ended, the universe settled into a much calmer phase. For the next 300 000 years, the universe consisted of a sea of hydrogen and helium nuclei, photons, free electrons and the mysterious **dark matter**. The universe kept expanding and cooling but the intense radiation caused the disintegration of any atoms. After further expansion and cooling, this period eventually ended when the sea of photons was not energetic enough to keep electrons from joining with nuclei to form atoms.

The universe had cooled down to about 3000 K by now. Atoms — mostly hydrogen and some helium — began forming. The photons no longer had much to do, so they travelled unhindered through space. It was as if the universe became transparent to photons. Gravity started pulling the hydrogen and helium together and this collapse produced the vast aggregations of gas and other matter we now call galaxies. Scattered within these galaxies were sub-condensations of gas. As this gas continued to condense inwards, magnetic fields provided a pressure acting outwards which slowed the collapse. Enormous heat was generated as gravitational energy was transferred to heat energy and this caused them to glow with infrared radiation. As the swirling and rotating gases collapsed, the speed of rotation increased (we see this principle stated in the law of conservation of angular momentum; that is, as the radius decreases, the speed increases). Jets of material blew outwards and stopped any more material condensing on the newly created star (the 'protostar') and this outflow separated the young solar system from its parental core.

These protostars give off huge amount of heat, but not all ignite to become true visible stars. Depending on their mass, they can go one of several ways: The low-mass protostars (<0.08 solar masses) will fail to ignite and just glow dull red. These 'failed' stars are called 'brown dwarfs' and lock up enormous amounts of hydrogen for billions of years. Ones that are more than 8% of the mass of our Sun (>0.08 solar masses) will sustain nuclear reactions. Low-mass stars (0.08–0.5 solar masses) fuse hydrogen into helium but at a fairly slow rate. These are called 'red dwarfs' and will last for an extraordinarily long time — possibly a trillion (10^{12}) years.

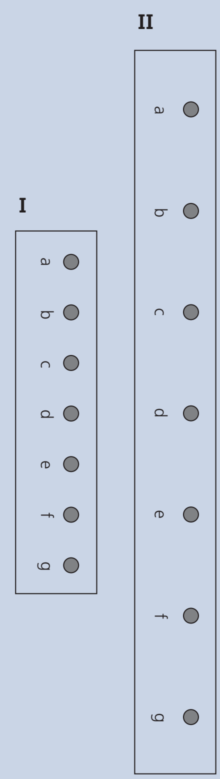
Stars with masses in the range from 0.5 to 8 solar masses will burn like our Sun. The core temperature of our Sun is currently about 16 million K, but when the hydrogen becomes depleted in its centre the core will lack an energy source and will cool. There will not be enough heat to support its overlying bulk. The core will shrink to become a white dwarf but the outer layers will evaporate as a massive solar wind of energetic particles — and become a red giant. The central core will continue to shrink and heat up, and when it reaches

NOVEL CHALLENGE

The figure below shows the relative positions of some ink dots on the surface of a balloon as it is inflated. If you keep dot 'd' directly in front of you, all the other dots appear to be moving away from 'd' as the balloon gets bigger. Imagine diagram II shows the positions 1 second after diagram I.

A Calculate the speed of dots 'e', 'f' and 'g' by measuring the distances with a ruler. Plot a graph of speed (y-axis) versus distance from 'd'. Relate this graph to Hubble's law.

B Imagine we are object 'd' in the Milky Way galaxy and object 'g' is the Hydra galaxy. If Hydra is moving at a speed of 60 500 km s^{-1}, calculate the time elapsed between diagrams I and II.



PHYSICS FACT

The shortest 'day' in our solar system has just been discovered by NASA. The asteroid 1998KY26 (about 30 m across) spins at 1 revolution every 10.7 second. That means its day is 0.09 seconds long. KY26 consists of 4 tonnes of water which must be frozen otherwise it would fly apart.

100 million K a new series of nuclear reactions will occur. Helium will fuse into carbon and release more energy. The core of the giant will become an enormous helium bomb and for a time will produce more energy than all the stars in the universe combined. The resulting helium flash will be followed by a quiet burning for about 100 million years.

Stars that are superheavy (>8 solar masses) will burn up their central stores of hydrogen in about 10 million years (recall that lighter ones like our Sun will last for about 10 *billion* years from ignition). The helium quickly fuses into carbon and the star begins to collapse under its own weight. The core reaches 100 billion K and huge numbers of neutrinos are produced, leaking away energy and allowing the collapse to accelerate. Temperatures rise and carbon fuses into magnesium, and by a complex maze of nuclear reactions neon, oxygen, silicon, sulfur and iron form. Once the chain reaches iron no further fusion is possible as iron is so stable. A star with an iron core is doomed as it cannot squeeze any more energy out by fusion. The star cannot support itself any longer. In a single second the star collapses, compressing the central regions to a density of 10^{14} g cm^{-3} (and that is dense — water is only 1 g cm^{-3}). If Earth were this dense it would be about 400 m in diameter. Electrons and protons are squashed up to form neutrons and the whole star resembles one big nucleus. A shock wave ensues and the outer layers are fused into heavy elements such as gold, lead and uranium and blown away, leaving a dense core of neutrons behind. This explosion is a **supernova** and the dense core of neutrons remaining becomes a **neutron star**. For some very massive stars (8–50 solar masses), the neutron cores cannot support their own weight and collapse to become a **black hole**. Black holes are strange beasts, with gravitational fields so strong that light itself cannot escape. They are so fascinating that we have devoted a whole section to them later on. (See section 6.9.) The matter produced by the supernova is thrown back into the interstellar medium, where it eventually condenses to form stars like our own Sun and its planets like our Earth. The heavy elements on Earth (e.g. uranium) originated in a supernova.

Superheavy stars of more than 100 solar masses are likely to blow themselves apart instantly so that nothing is left.

Many stars will continue to be born, but after a few trillion years even the hardiest of stars will have died and returned much of their mass back to the interstellar medium. The universe will all of a sudden be a dark place.

The Degenerate Era: $t = 10^{15}$ to 10^{39} years

At the beginning of the Degenerate Era, a thousand trillion years will have elapsed since the Big Bang. The universe will be a pretty boring place with just the remnants of the Stelliferous Era scattered all over the place. There will be brown, red and white dwarfs, neutron stars and black holes. Galaxies will drift around but the mutual force of attraction will cause them to combine and form mega-galaxies. Our Milky Way will combine with the Andromeda galaxy to form a sort of thick-shake. However, stellar collisions will be rare because space is so big (remember?). About once every billion years two stars will collide, and this will be a big event in the Degenerate Era. The most boring will be two brown dwarfs in collision. They won't light up but will just form some planets which are likely to have the conditions necessary to support life. White dwarf collisions will be more spectacular. Their combined mass will enable them to ignite and burn for a million years before depleting their nuclear fuel and turning off. Black holes will grow larger and more massive throughout this era as they capture stars and gas that have come too close.

Proton decay

Up until the 1980s it was thought that protons were infinitely stable. Now we know that protons do not last forever; in fact they last for only about 10^{37} years. In the final stages of the Degenerate Era protons will start to decay. In a white dwarf for instance, the protons decay into pions and positrons; the pions decay into high-energy gamma rays and the positrons pair up with stray electrons to form more gamma rays. The net result is that the white dwarf slowly evaporates and finally nothing is left. Kaput! Neutron stars suffer the same fate. Neutrons decay into protons, electrons and antineutrinos, which all end up as gamma radiation. By the end of the era, 10^{40} years will have passed since the Big Bang. All that will remain will be a vast sea of radiation, mostly photons and neutrinos with a smaller admixture of electrons and positrons. And Black Holes — millions of big Black Holes.

The Black Hole Era: $t = 10^{40}$ to 10^{100} years

Now the black holes reign supreme. The universe has reached a volume of 10^{100} cubic metres and is inhabited by 10^{46} black holes in an otherwise almost empty universe (see Figure 6.4). These holes grew larger and larger during the previous two eras — sweeping up everything in sight. But black holes radiate energy, and slowly — very slowly — evaporate. Their last moments are dramatic. After shedding 95% of their mass through evaporation, their surface will be as hot as our Sun. During the last second the black hole explodes, giving off 10^{22} J in gamma-ray energy and a blast of electrons, positrons, protons, antiprotons and other exotic particles such as the dark matter WIMPS. The protons and antiprotons annihilate themselves immediately to produce more gamma photons. By 10^{100} years the black holes have all evaporated and a final enveloping night moves in. This is getting creepy!

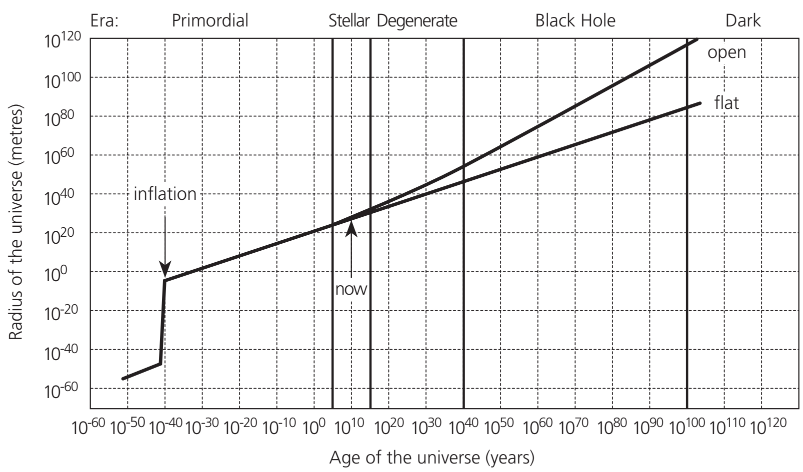


Figure 6.4
Changes to the radius of the observed universe over the five eras. Note the exponential scales.

The Dark Era: $t = 10^{100}$ years to eternity

In this cold and distant future, activity in the universe has almost stopped. Energy levels are low and the expanses of time and space are almost beyond belief. Much depends on whether the universe will continue to expand at the same rate (an 'open' universe — the most favoured model) or slow down to an almost negligible expansion (a 'flat' universe). The third alternative — the 'closed' universe — now seems most unlikely, as there appears to be insufficient mass in the universe to stop expansion and begin a contraction phase. (See Figure 6.5.) This possibility has the universe expanding to a maximum volume in about 20 billion years and then contracting back to a point in another 20 billion years — the Big Crunch. Not much is mentioned of the Big Crunch these days. If it does happen, there is speculation that another Big Bang could ensue and another cycle begins. Physicists say that if this happens, the cycles will get bigger and bigger as each one occurs. Some say that maybe there was a cycle before the present one. They say that if there was there could have been up to 100 previous cycles (but no more). They are adamant that there was a definite start to the cycles if they occurred – that is, the universe hasn't been going for an infinite time.

At the start of the Dark Era the volume of the universe will be 10^{182} m^3 if the 'flat' scenario happens or 10^{272} m^3 if it is 'open'. Either way, space will be occupied by about one electron and positron in every 10^{192} m^3 or so. With no protons left, electrons cannot form

NOVEL CHALLENGE

A marathon runner starts off at the same time as a radar signal leaves the earth for Jupiter. He stops when the echo is received back on Earth. How many kilometres does he run?

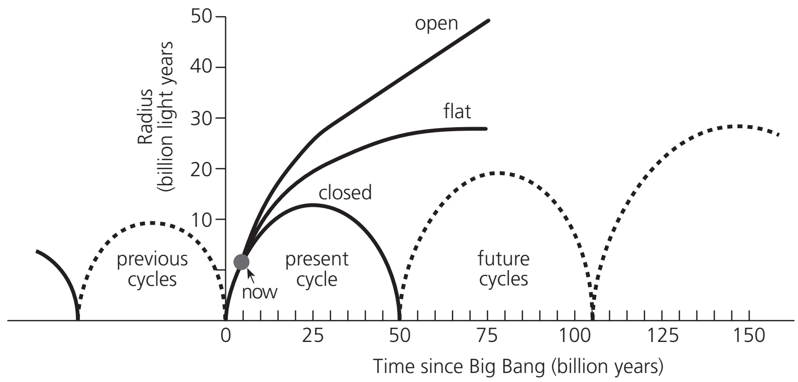


Figure 6.5
Three models of the future of the universe. The 'open' model is the current favourite.

normal atoms. Instead they form **positronium** atoms in which the electron and positron spiral around each other in orbits trillions of light-years across. By the time 10^{145} years have passed since the Big Bang, even these atoms will have decayed. The temperature is now exceedingly low — just billionths of a degree above absolute zero. The photons of light that inhabit the universe are so lacking in energy that their wavelength is 10^{41} light-years — bigger than the observable universe today. The universe has run down and now heads for heat death but never quite dies; it hangs on in eternal death throes. But wait — bizarre things could happen. Physicists say that processes could happen where new 'child universes' are created spontaneously within our universe and then undergo inflation just like ours did at the Big Bang. Who knows?

Questions

10 By what factor did the universe expand in the 'inflation' phase?
11 What was the cause of the universe cooling?
12 Comment critically: 'The Big Crunch is the favoured theory of the future of the universe.'
13 Discuss the longevity of protons.
14 Discuss critically: 'All stars containing hydrogen will undergo fusion at a rate dependent on their mass.'
15 Inward and outward forces maintain a star in a state of equilibrium. Explain what happens to the forces as a star turns into **(a)** a brown dwarf; **(b)** a white dwarf; **(c)** a neutron star; **(d)** a black hole.

Let's now look at our own Solar System in more detail.

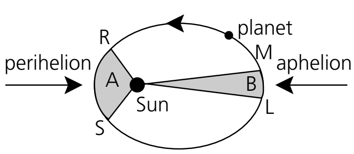
KEPLER'S LAWS 6.3

German astronomer **Johannes Kepler** (1571–1630) spent years working on data provided by Tycho Brahe, his predecessor as Imperial Mathematician to Emperor Rudolph in Prague. However, in 1614, Galileo described Kepler's writing as 'so obscure that apparently the author did not know what he was talking about'. In 1618 Kepler published the first two of his three laws and was hailed as a hero by those who wanted to do away with the old geocentric (Earth-centred) universe of Aristotle and Ptolemy. Kepler's laws were these:

First law: the law of orbits All planets move in elliptical paths, the Sun being at one focus.

Second law: the law of areas The speed of a planet along this path is not uniform, but varies with its distance from the Sun in such a way that a line drawn from the planet to the Sun would sweep out equal areas in equal times; or, in other words, the area swept out in a given time by the radius vector is always constant (Figure 6.6).

Figure 6.6
Planets move in elliptical orbits, with varying speeds around the Sun. The planet takes the same time to move from L to M as it does from R to S. The shaded areas A and B are equal (Kepler's law).



Both these laws were in contradiction to conventional wisdom of the time. The view of the universe being taught in universities in the 1500s and early 1600s was based on the ideas of Greek philosophers Pythagoras, Aristotle and Plato, who declared that the Earth was at the centre of the universe and the path of planets must be circular and the speed uniform. The second-century Arab astronomer Ptolemy produced a comprehensive theory of planetary motion based on these ideas (the Ptolemaic theory), which held sway for 1400 years. In fact, all Christian astronomers held this view, using a literal interpretation of the Bible for support ('Joshua commanded the Sun to stand still, not the Earth'). As a result, the Catholic Church in 1616 put Copernicus's works on the forbidden list. Galileo and Tycho refused to accept elliptical motion, preferring to believe in Copernicus's Sun-centred model but with uniform circular orbits. Copernicus too said that his 'mind shuddered' at the idea of elliptical orbits. Kepler had earlier agreed, saying that non-circular orbits were 'a cartful of dung', but was eventually won over by the beauty and simplicity of elliptical orbits. Kepler was truly the first astrophysicist.

Third law: the harmonic law or the law of periods In 1619, a more remarkable hypothesis followed. Galileo couldn't cope with this one either. The law stated:

For orbiting satellites or planets of any system, the ratio of the radius of orbit cubed, r^3, to period squared, T^2, is constant for all satellites of that system. (Table 6.3)

$$\frac{r^3}{T^2} = \text{constant, or } \frac{r_a^3}{T_a^2} = \frac{r_b^3}{T_b^2}$$

The table below shows the r^3/T^2 values for the Earth and its neighbouring planets. Note how constant the ratio is (average = 3.36×10^{18} m^3/s^2).

Table 6.3

PLANET	AVERAGE RADIUS OF ORBIT (m)	PERIOD OF REVOLUTION (s)	r^3/T^2 (m^3/s^2)
Mercury	5.97×10^{10}	7.60×10^6	3.68×10^{18}
Venus	1.08×10^{11}	1.94×10^7	3.35×10^{18}
Earth	1.49×10^{11}	3.16×10^7	3.31×10^{18}
Mars	2.28×10^{11}	5.94×10^7	3.35×10^{18}
Jupiter	7.78×10^{11}	3.74×10^8	3.36×10^{18}
Saturn	1.43×10^{12}	9.30×10^8	3.38×10^{18}

Example 1

Triton and Nereid are the two moons of Neptune. Triton is 353 000 km from Neptune and has a period of 5.87 Earth days. Nereid is 5 560 000 km from Neptune and its period is 359.9 Earth days. Find out if these data are consistent with Kepler's third law. Neptune has a radius of 24 750 km.

Solution

For Triton: $\dfrac{r^3}{T^2} = \dfrac{(353\,000 + 24\,750)^3}{5.87^2} = 1.56 \times 10^{15}$.

For Nereid: $\dfrac{r^3}{T^2} = \dfrac{(5\,560\,000 + 24\,750)^3}{359.9^2} = 1.34 \times 10^{15}$.

Conclusion: values are close — Kepler's Law confirmed.

NOVEL CHALLENGE

Another new planet outside our solar system has recently been discovered. It lies within the Oort Cloud with an orbital radius of 0.4 ly. It has a mass 1.5 to 6 times of Jupiter and a period of 6 million years. How does it r^3/T^2 ratio compare to that of our solar system?

Ch06-Walding 4th 25/8/04 11:08 AM Page 147

Example 2

Consider a satellite launched from Earth to have a geosynchronous orbit (period = 1 day) around the Earth. Knowing that the natural satellite of Earth (the Moon) has a period of 28 days and an orbiting radius of 3.8×10^8 m, calculate the desired radius for the artificial satellite so that it has a period of 1 day.

Solution

$$\frac{r_m^3}{T_m^2} = \frac{r_s^3}{T_s^2}$$

$$r_s^3 = \frac{r_m^3 \times T_s^2}{T_m^2} = \frac{(3.8 \times 10^8)^3 \times 1^2}{28^2} = 7.0 \times 10^{22}$$

$$r_s = 4.1 \times 10^7 \text{ m } (41\ 000 \text{ km})$$

Questions

16 The average radius of the orbit of Uranus is 2.87×10^{12} m. Use the average value for r^3/T^2 from Table 6.3 to calculate the period of Uranus in **(a)** seconds; **(b)** years.

17 Neptune takes 164.8 y to orbit the Sun. Use the average value of r^3/T^2 from Table 6.3 to find the average radius of its orbit.

18 It was once thought that the planet Vulcan existed between Mercury and the Sun. What would its period have been if it was at a mean radius of 40 million km from the Sun?

NEI Activity 6.5 FACTS AND FIGURES

Use a dictionary or encyclopaedia to answer the following:

1. **Planet** means 'wanderer' but what exactly was meant by wandering and what is **retrograde** motion?
2. **Galaxy** comes from the Greek *galas* meaning 'milk'. What has our galaxy got to do with milk?
3. What is the difference between a **pulsar** and a **quasar**? Name one of each and state their distance from Earth.
4. Locate r and T values for Pluto and determine its r^3/T^2 value as in Table 6.3.

NEI Activity 6.6 COMPETING THEORIES

Select one of the following topics for library research and write a short (one page) response.

- The Ptolemaic system used circular motion but still allowed planets to move in non-circular orbits. Show how **Ptolemy** used 'epicycles' to contrive his system.
- Why was **Copernicus** so reluctant to publish his theory but finally relented on his deathbed? What was he scared of?
- Why was the Church so annoyed with **Galileo**? Describe how he was treated. Was he better off than fellow astronomer Bruno in the hands of the Inquisition?

Ch06-Walding 4th 25/8/04 11:08 AM Page 148

6.4 NEWTON'S LAW OF GRAVITATION

Kepler had shown that the Sun was the centre of our solar system but the question remained: what was the nature of the force that the Sun exerted? Kepler believed that the Sun moved the planets by sending out rays like wheel-spokes, which carried the planets around.

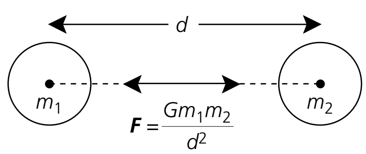
In 1666, **Newton** united the ideas of Copernicus and Kepler with the laws of falling bodies, developed by Galileo. Newton provided a new worldview, based on a new physics, uniting heaven and Earth in one mathematical structure. A revolution in thought had begun.

Newton used the word **gravity** to describe the force between the Sun and the planets. It came from the Latin *gravitas*, meaning 'weight'. In his most famous work, *Philosophiae Naturalis Principia Mathematica*, published in Latin in 1687, he wrote: 'I deduced that the forces which keep the planets in their orbs must vary reciprocally as the squares of their distances from the centres about which they revolve and in direct proportion to their masses'. In the form of an equation, this becomes:

$$\boldsymbol{F}_g = \frac{Gm_1m_2}{d^2}$$

where $\boldsymbol{F}_g$ is the force of gravitational attraction, m_1 and m_2 are the masses of the attracting objects, d is the distance between the objects' centres and G is a proportionality constant called the **universal gravitational constant** or more commonly known by astrophysicists as *Big G* to distinguish it from 'little' g — the acceleration due to gravity. In SI units, G has the value 6.67×10^{-11} N m^2 kg^{-2}. (See Figure 6.7.)

Figure 6.7



Example

Determine the force of attraction between the Earth (mass = 5.98×10^{24} kg) and the Moon (mass of the Moon is 7.35×10^{22} kg), given that the Earth–Moon distance is 3.8×10^8 m.

Solution

$$\boldsymbol{F}_g = \frac{Gm_1m_2}{d^2} = \frac{6.67 \times 10^{-11} \times 7.35 \times 10^{22} \times 5.98 \times 10^{24}}{(3.8 \times 10^8)^2} = 2.07 \times 10^{20}\ \text{N}$$

Questions

19 Calculate the force between the Sun ($m = 2.0 \times 10^{30}$ kg) and the Earth ($m = 5.98 \times 10^{24}$ kg) assuming they are about 1.5×10^8 km apart.

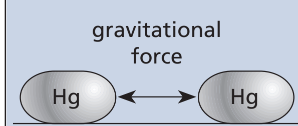
20 Black holes are super-massive collapsed stars. The closest anything can get to one and still escape the gravitational force is called the 'event horizon'. Calculate the force acting on a 15 t spacecraft at the event horizon of 30 km from a black hole that has a mass equal to 10 times that of the Sun. Mass of Sun = 2×10^{30} kg; 1 t = 1000 kg.

NOVEL CHALLENGE

Newton said *rationem vero harum Gravitatis propietartum ex phenomenis nondrum potui decucere* ('But I have not been able to discover the reason for this property of gravitation from the phenomena'). *What did he mean?*

NOVEL CHALLENGE

Two spherical drops of mercury are resting on a frictionless surface. The only force between them is that of gravitation. *What would you need to know to be able to calculate how much time it would take for them to touch?*



PHYSICS UPDATE

In 1998, the International Panel on Physical Constants deemed G to be 6.672 59 ± 0.008 $\times 10^{-11}$ N m^2 kg^{-2}.

CIRCULAR MOTION 6.5

The two main types of periodic motion in space are:

- **elliptical motion** — planets about the Sun and some artificial satellites
- **circular motion** — moons about their planets and some artificial satellites.

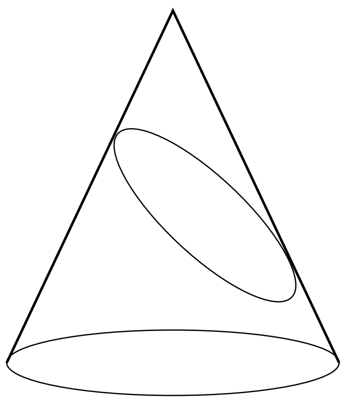


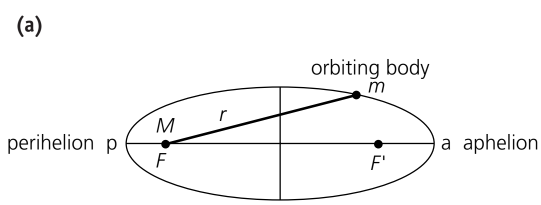
Figure 6.8
An ellipse is an angled segment through a cone.

Elliptical motion

An ellipse is produced if you make a sloping cut through a conical pyramid (Figure 6.8).

A planet of mass *m* moves in an elliptical orbit around the Sun. The Sun, of mass *M*, is at one focus F of the ellipse (Figure 6.9**(a)**). The other, or 'empty' focus is F′. The point closest to the Sun is called the **perihelion** and the opposite point farthest from the Sun is called the **aphelion**. When referring to an elliptical orbit in general, these points are called the **perigee** and **apogee** respectively. The words come from the Greek *peri* meaning 'around'; *apo* meaning 'away'; *helios* meaning 'Sun'. The suffix 'gee' is derived from the Greek *geo*, the Earth. You should be able to deduce where words like perimeter, geometry and apology come from.

The **eccentricity** is a measure of how much out-of-round the ellipse is (Figure 6.9**(b)**). The eccentricity, e, is the difference between the distance from focal point to aphelion, F_a, and focal point to perihelion, F_p, divided by the total perihelion to aphelion distance, ap: $e = \dfrac{F_a - F_p}{ap}$. An eccentricity of zero means the ellipse is circular. At perihelion, the Earth is 147 097 800 km from the Sun; at aphelion it is 152 098 200 km, a difference of 5 million km. The eccentricity of the Earth's orbit about the Sun is 0.0167, which means it is almost circular. In a circle of diameter 100 cm, this difference corresponds to the centre being 0.8 cm off-centre. Not much! In the drawings below, the shape of the ellipse has been distorted for clarity. It should look more like a circle.



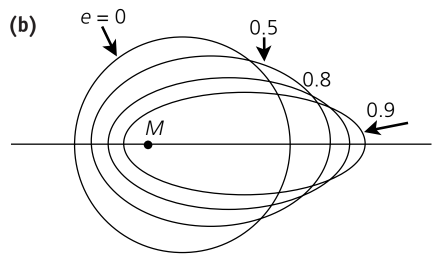


Figure 6.9
Planets move in elliptical orbits.

Circular motion

Planets may have elliptical orbits about the Sun, but satellites that orbit the planets mostly have **circular** paths.

- The natural satellite of the Earth (the Moon) has a circular path.
- Most artificial Earth-orbiting satellites have circular paths, thus keeping their speed constant.
- The Earth rotates on its own axis, so a point on its surface travels in a circular path too.

In order to understand surface and satellite motion, it is necessary to revise the physics of circular motion.

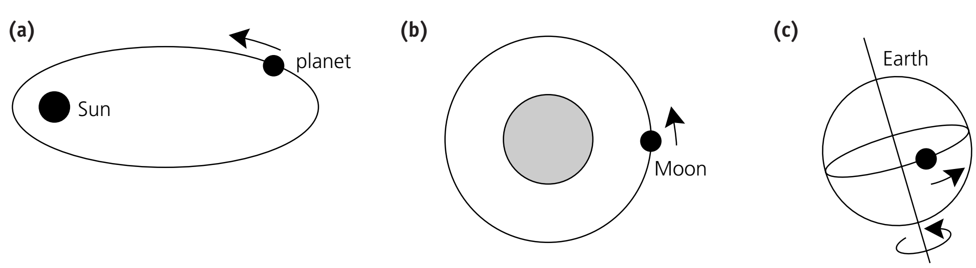


Figure 6.10
(a) Planets travel around the Sun in elliptical orbits. **(b)** The moon and most artificial satellites have circular orbits about the Earth. **(c)** Points on the Earth's surface have circular paths as the Earth turns on its own axis.

Centripetal acceleration

In the previous chapter it was shown that objects travelling in circular orbits at constant speed had an acceleration directed toward the centre of the circular path. This acceleration was called **centripetal acceleration:**

$$\boldsymbol{a}_c = \frac{v^2}{r}$$

Centripetal force

Because of this centripetal acceleration, an object of mass m experiences a **centripetal force** $\boldsymbol{F}_c$ also directed toward the centre of the circular path:

$$\boldsymbol{F}_c = m\boldsymbol{a}_c = \frac{mv^2}{r}$$

Period of circular motion

If the object travels at uniform speed $\boldsymbol{v}$ in a circle of radius r, the distance travelled during one full revolution $s = 2\pi r$. The time taken to complete one full revolution (the period, T) is:

$$T = \frac{\boldsymbol{s}}{\boldsymbol{v}} = \frac{2\pi r}{\boldsymbol{v}} \quad \text{or} \quad \boldsymbol{v} = \frac{2\pi r}{T}$$

Since $\boldsymbol{a}_c = \frac{v^2}{r}$, we obtain $\boldsymbol{a}_c = \frac{4\pi^2 r}{T^2}$ and $\boldsymbol{F}_c = \frac{4\pi^2 rm}{T^2}$.

Example

The radius of the Earth at the equator is 6.4×10^6 m and its mass is 6.0×10^{24} kg. Calculate **(a)** the centripetal acceleration of a point at the equator; **(b)** the centripetal force acting on this point; **(c)** the linear velocity of this point.

Solution

- T = 24 hours = 86 400 s

  (a) $\boldsymbol{a}_c = \frac{4\pi^2 r}{T^2} = \frac{4\pi^2 \times 6.4 \times 10^6}{86\,400^2} = 0.034 \text{ m s}^{-2}$.

  (b) $\boldsymbol{F}_c = m\boldsymbol{a}_c = 6.0 \times 10^{24} \times 0.034 = 2.04 \times 10^{23}$ N.

  (c) $\boldsymbol{v} = \frac{2\pi r}{T} = \frac{2\pi \times 6.4 \times 10^6}{86\,400} = 465 \text{ m s}^{-1}$.

Questions

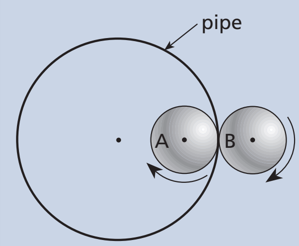
21 A 750 kg spacecraft is in a circular orbit of radius 700 km and is travelling at 6500 m s^{-1}. Calculate **(a)** the centripetal acceleration; **(b)** the centripetal force; **(c)** the period.

22 A person living on the equator of the Earth makes one complete revolution around the Earth's axis in 24 hours. For a 65 kg person, find **(a)** their centripetal acceleration; **(b)** the centripetal force; **(c)** the linear velocity. The radius of the Earth is 6.4×10^6 m.

NOVEL CHALLENGE

Two balls are rolled around the inside surface and outside surface of a pipe whose diameter is 3 times the diameter of the balls.

How many times will the outside ball turn to complete one rotation of the pipe compared with the inside one?



6.6 SATELLITES

Satellite means 'neighbour' or 'companion'. Town planners talk about satellite cities. To an astrophysicist, satellites can either be natural (moons) or artificial (e.g. communications satellites). Artificial satellites are generally used for:

- science — research, weather, mapping
- industry — agriculture, mining
- communications — radio, TV, phone and computer networks; position coordinates
- military and political — defence planning, spying.

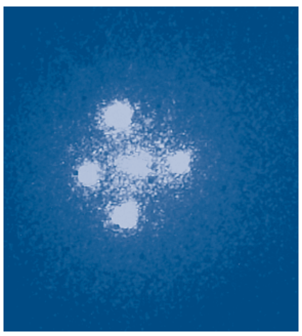
Examples of satellites

A big new eye in the sky

Photo 6.2
The Hubble telescope.

A new era in astronomy began in 1990 when the **Hubble space telescope** went into orbit around the Earth (see Photo 6.2). The telescope initially had blurred vision due to a faulty main mirror and also had problems with shaking solar panels. These were eventually fixed in 1993 and it now sends back a rich harvest of observations including compelling clues to the existence of some super-heavy massive black holes and amazing close-ups of the Orion nebula, where new stars are being born from clouds of gas. Hubble also beamed back details of a new type of cosmic object — a gigantic concentration of stars produced by two colliding galaxies 200 million light years away called the 'Starburst galaxy'. Probably the most spectacular Hubble image shows the famous 'Einstein cross' (see Photo 6.3). As light from a quasar 8000 million light years away grazes a galaxy at only 400 million light years away, it is bent in new directions. From Earth we see four images of the distant quasar, with the foreground galaxy in the middle. This effect, called 'gravitational lensing', was predicted by Albert Einstein 70 years ago and is a remarkable confirmation of his theory of gravity.

Photo 6.3
Einstein's Cross.



The Hubble space telescope has a mass of 11 t and is in a circular Earth orbit 610 km above the surface. With its unrivalled ability to measure cosmic distances, it could help to answer one of the biggest questions of all: how big and how old is the universe? If Hubble could see 15 billion light years away then it would see the moment of creation.

ERS-2 This satellite was launched in April 1995 after the huge success of ERS-1 (see Photo 6.4). CSIRO placed a temperature sensor aboard to gather data about global warming. The sensor also measures sunlight reflected from the ground, providing better estimates of bushfire risk and crop yields.

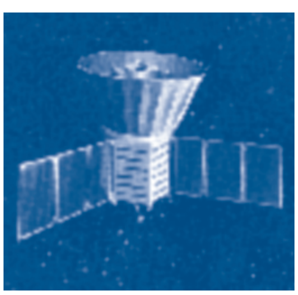
Scientists have a new tool to search for the 'fossil record' of the Big Bang and uncover clues about the evolution of the universe. Launched in 1999, NASA's **Far Ultraviolet Spectroscopic Explorer** (FUSE) observes nearby planets and the farthest reaches of the universe to provide a detailed picture of the immense structure of our own Milky Way galaxy. The FUSE mission's primary scientific focus is the study of hydrogen and deuterium ($^{2}_{1}H$), which were created shortly after the Big Bang. With this information, astronomers in effect will be able to look back in time at the infant universe.

Photo 6.4
The ERS-1 satellite.

By examining these earliest relics of the birth of the universe, astronomers hope to gain a better understanding of the processes that led to the formation and evolution of stars, including our solar system. Ultimately, scientists hope data from FUSE will allow them to make a huge leap of understanding about how the primordial elements were created and have been distributed since the beginning of time.

The **Cosmic Background Explorer** (COBE) spacecraft has been engaged in some of the most exciting work ever done in the study of the universe. It has peered back in time some 10 billion years, very nearly the point of creation. Launched in 1989 and managed by NASA's Goddard Space Flight Center, COBE has uncovered landmark evidence to support the Big Bang theory of an expanding universe. Science researchers continue to analyse data received from the spacecraft.

Photo 6.5
NASA's Cosmic Background Explorer (COBE). What a success!



SR Activity 6.7 UPDATE ON SATELLITE PROGRAMS

The problem in writing about satellite programs in textbooks like this is that progress is so rapid and developments unfolding so fast that information dates very quickly. The only way to keep up is with newspaper and magazine articles or by direct communication with the satellite agencies themselves. Alternatively, you could always see what answers you can get on the Internet (start with NASA's Home Page on the Worldwide Web or join some of the Astronomy newsgroups).

Using the Internet, try to get an update on one of the following programs:

- CSIRO's involvement in **ERS-2** and when its program is likely to finish.
- **Cape York space port** — will it ever launch a satellite?
- How does the **Global positioning system** (GPS) work?
- Sailors often have an **EPIRB** on their boats. What is it and how does it work?

Satellite motion

For any object to orbit the Earth, it must have sufficient velocity to overcome the Earth's gravitational pull. Figure 6.11 shows the path of four objects projected horizontally from a high tower. Path A would happen if gravity did not act; Path B if the speed was low; Path C if it was higher than B but still too slow; and Path D if the speed was just right.

In a **circular orbit**, a satellite always travels at the same speed and stays the same distance from Earth. The earliest measurement of the Moon's period shows that it hasn't changed over the past few thousand years — all we've been able to do is measure it more accurately. It is known to be 27.321 661 days.

The 'right' speed for a satellite is such that the centripetal force needed to keep it in a circular path exactly equals the force of gravity or its weight. This velocity is called its **critical velocity** ($\boldsymbol{v}_{\text{crit}}$)

Centripetal force = satellite's weight

$$\boldsymbol{F}_c = m\boldsymbol{g}$$

$$\frac{m\boldsymbol{v}_{\text{crit}}^2}{r} = m\boldsymbol{g}$$

$$\boldsymbol{v}_{\text{crit}} = \sqrt{\boldsymbol{g}r}$$

If the velocity is less than critical the satellite will fall back towards Earth. If it is more than critical it will rise to a bigger orbit. Note that the value of the critical velocity is independent of mass. It just depends on the radius of the orbit and the acceleration due to gravity ($\boldsymbol{a}_g$ or $\boldsymbol{g}$) at that radius. The value of $\boldsymbol{g}$ is not 9.8 m s^{-2} — it is a lower value further away from the Earth.

Example

Calculate **(a)** the gravitational force; **(b)** the critical velocity; **(c)** the period of the orbit of a 5000 kg satellite moving uniformly in a circular path 400 km above the Earth. The radius of the Earth = 6.38×10^6 m.

Solution

(a) $m_s = 5000$ kg; $m_e = 5.98 \times 10^{24}$ kg.

Radius of orbit, $d = 400 \times 10^3 \text{ m} + 6.38 \times 10^6 \text{ m} = 6.78 \times 10^6$ m

$$\boldsymbol{F} = \frac{Gm_1m_2}{d^2} = \frac{6.67 \times 10^{-11} \times 5000 \times 5.98 \times 10^{24}}{(6.78 \times 10^6)^2} = 43\ 385 \text{ N}$$

(b) The force of attraction (43 385 N) equals the centripetal force.
Using the centripetal force law:

$$\boldsymbol{F}_c = \frac{m\boldsymbol{v}^2}{r};\ \boldsymbol{v}^2 = \frac{\boldsymbol{F}_c r}{m} = \frac{43\ 385 \times 6.78 \times 10^6}{5000} = 5.88 \times 10^7$$

$$\boldsymbol{v} = \sqrt{5.88 \times 10^7} = 7670 \text{ m s}^{-1}$$

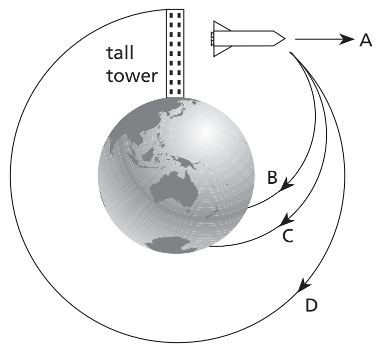
(c)

$$T = \frac{2\pi r}{\boldsymbol{v}} = \frac{2\pi \times 6.78 \times 10^6}{7670} = 5554 \text{ s (1.54 hours)}$$

Escape velocity

If you fire an arrow upward, usually it will slow, stop momentarily, and return to Earth. There is, however, a certain initial speed that will cause it to move upward and escape the Earth's pull. This is called its **escape velocity** and for Earth it is 11.2 km/s. Physicists have derived a formula:

Figure 6.11



NOVEL CHALLENGE

A simple formula for calculating the distance to the horizon is:

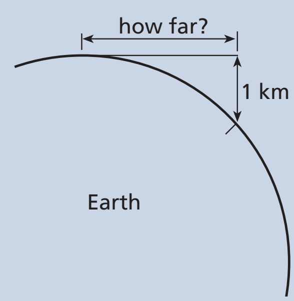
miles to horizon = $\sqrt{\text{eyeheight in feet} \times 1.5}$.

Show that this formula can be converted to:

kilometres = $\sqrt{\dfrac{\text{eyeheight in cm}}{8}}$.

NOVEL CHALLENGE

How far would you have to travel horizontally out from the Earth for your altitude to be 1 km?



$$v_{esc} = \sqrt{\frac{2Gm}{r}}$$

where m is the mass of the planet or moon and r is the radius.

Example

Verify the escape velocity for the Earth as 11.2 km s^{-1}. Ignore the effects of air drag and Earth's rotation. $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

Solution

$$v = \sqrt{\frac{2Gm}{r}} = \sqrt{\frac{2 \times 6.67 \times 10^{-11} \times 5.98 \times 10^{24}}{6.37 \times 10^{6}}}$$

$$= 1.12 \times 10^{4} \text{ m s}^{-1} \text{ (= 11.2 km/s)}$$

NOVEL CHALLENGE

In 1971 Apollo 15 astronauts David Scott and James Irwin drove the 4WD lunar vehicle around for 30 km on the Moon. Would they have used as much fuel as on Earth?

NOVEL CHALLENGE

During a lunar eclipse, the shadow of the Moon on the Earth consists of a black region 100 km wide which travels at 3000 km h^{-1}. Scientists and thrill seekers try to stay in the shadow zone as long as possible to make observations. *What is the maximum time you could stay in the shadow zone if you were in a plane that could travel at 1000 km h^{-1}?*

Questions

Use the data supplied in Table 6.3 on page 146 where necessary.

23 Calculate **(a)** the force between the Earth and an artificial satellite of mass 2500 kg, which is in a 6400 km orbit above the surface of the Earth; **(b)** its velocity; **(c)** its period.

24 A communications satellite is shifted from an orbit of one Earth radius above the surface of the Earth to three Earth radii above the surface. What effect does this have on the satellite's **(a)** gravitational force; **(b)** velocity; **(c)** period?

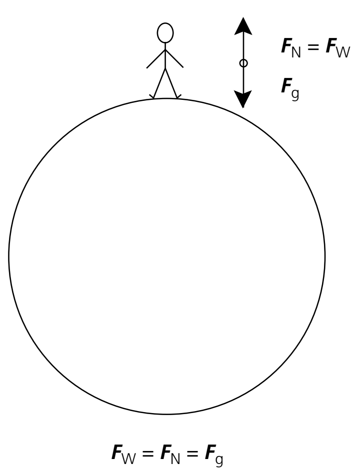
25 Calculate **(a)** the minimum orbiting speed; **(b)** the period of the orbit of a satellite moving uniformly in a circular path 1170 km above the Earth, where g = 7.0 m s^{-2}. The radius of the Earth = 6.4 × 10^6 m.

26 Calculate the escape velocity from **(a)** the Moon; **(b)** Pluto; **(c)** Jupiter; **(d)** the Sun.

WEIGHT, GRAVITY AND CENTRIPETAL FORCE 6.7

The weight of a body is a measure of the force acting on it due to a nearby astronomical object such as a planet. To us on Earth, the astronomical object is the Earth. Your true weight (F_W) is the product of mass and the free-fall acceleration at the surface of the planet (= mg). For us, g = 9.8 m s^{-2} or approximately 10 m s^{-2}. Your apparent weight may change depending on whether the planet rotates or not.

Figure 6.12
A non-rotating planet.



On a non-rotating planet

In such a case (see Figure 6.12), the weight of an object, F_w, is equal to the force due to gravity, F_g. The symbol g is called the acceleration due to gravity but is more correctly referred to as **free-fall** acceleration. It is the net acceleration.

$$F_g = \frac{Gm_1m_2}{d^2}, \text{ which equals } F_w = mg$$

Example

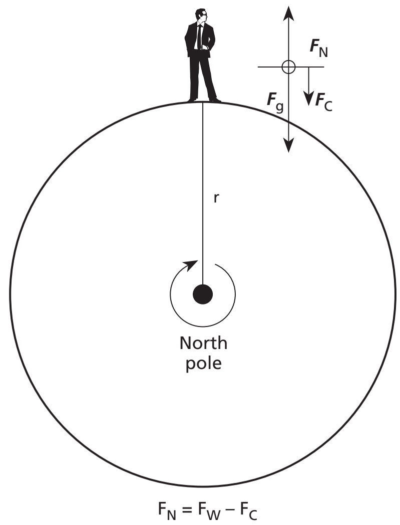
A 1.00 kg block of wood is at rest on the surface of a non-rotating planet of mass 3.0 × 10^{24} kg and radius 3.4 × 10^6 m. Calculate **(a)** its weight; **(b)** acceleration due to gravity (i.e. free-fall).

Solution

(a) $F_g = \dfrac{Gm_1m_2}{d^2} = \dfrac{6.67 \times 10^{-11} \times 3.0 \times 10^{24} \times 1.00}{(3.4 \times 10^6)^2} = 17.3$ N.

(b) $F_g = F_w = mg$, so $g = \dfrac{F_w}{m} = \dfrac{17.3}{1.00} = 17.3$ m s^{-2}.

Figure 6.13
A rotating planet.



On a rotating planet

Consider an object resting on the surface of the Earth at the equator (see Figure 6.13). As in the non-rotating case, the object is acted on by the weight (F_W, which equals F_g) causing the Earth to exert an opposite normal reaction force (F_N) back on the object. If the Earth was not rotating, $F_N = F_W$ and so the resultant force (the difference between the two) is zero; there is no acceleration.

However, the Earth and all objects on its surface rotate. An object will undergo uniform circular motion, which is to say the object is accelerating (centripetal acceleration a_c). As it is accelerating, there must be a resultant force — this is the centripetal force (F_c) acting towards the centre of the Earth. This means that F_N must be smaller than F_W, consistent with Newton's second law.

$$F_C = F_W - F_N$$
$$\text{Hence } F_N = F_W - F_C$$
$$\text{Apparent weight} = \text{weight} - \text{centripetal force}$$

The free-fall acceleration (g) is equal to the acceleration due to the gravitational force (a_g) minus the centripetal acceleration (a_c): $g = a_g - a_c$

Note: you'll find that centripetal acceleration is very small (about 0.3%) compared with free-fall acceleration and can be generally omitted without concern.

Note: for points not on the equator, the centripetal acceleration is not normal to the surface but is normal to the Earth's axis. It is not correct to use the formulas above as the forces are no longer in line. For this reason, the symbol F_N would be better replaced by F_R, the reaction force.

NOVEL CHALLENGE

If the Earth stopped rotating, how much would a 60 kg (590 N) person weigh? What would his bathroom scales read?

Example

A 1.00 kg brick rests on the surface of the Earth at the equator. Given that the radius of the Earth at the equator is 6.4×10^6 m and its mass is 6.0×10^{24} kg, calculate **(a)** the gravitational force on the brick; **(b)** the centripetal force on the brick; **(c)** the weight of the brick; **(d)** acceleration due to gravity; **(e)** centripetal acceleration; **(f)** free-fall acceleration; **(g)** centripetal acceleration as a percentage of free-fall acceleration.

Solution

- T = 24 hours = 86 400 s, m_o (mass of object) = 1 kg.

(a) $F_g = \dfrac{Gm_em_o}{r^2} = \dfrac{6.67 \times 10^{-11} \times 5.98 \times 10^{24} \times 1.00}{(6.38 \times 10^6)^2} = 9.83$ N.

(b) $F_c = \dfrac{4\pi^2 r m_o}{T^2} = \dfrac{4\pi^2 \times 6.4 \times 10^6 \times 1.00}{86\,400^2} = 0.0338$ N.

(c) $F_N = F_W - F_c = 9.83$ N $- 0.0338$ N $= 9.80$ N.

(d) Acceleration due to gravity, $a_g = \dfrac{F_g}{m} = \dfrac{9.83}{1.00} = 9.83$ m s^{-2}.

(e) Centripetal acceleration, $a_c = \dfrac{F_c}{m} = \dfrac{0.0338}{1.00} = 0.0338$ m s^{-2}.

(f) Free-fall or net acceleration, $g = 9.83 - 0.0338 = 9.80$ m s^{-2} (or could use $F_w = mg$).

(g) $\dfrac{a_c \times 100}{g} = \dfrac{0.0338 \times 100}{9.80} = 0.3\%$.

Hence, as centripetal acceleration, a_c, is so small, it is convenient to approximate the net or free-fall acceleration, g, by the acceleration due to gravity, a_g, and still call it g. This is what we did in previous chapters.

Questions

27 A 5.00 kg rock is at rest on the surface of a non-rotating planet of mass 1.5×10^{24} kg and radius 2.4×10^{6} m. Calculate **(a)** its weight; **(b)** acceleration due to gravity (i.e. free-fall).

28 A 750 kg space probe rests on the surface of the Earth near the equator. Given that the radius of the Earth at the equator is 6.4×10^{6} m and its mass is 6.0×10^{24} kg, calculate **(a)** the gravitational force on the vehicle; **(b)** the centripetal force on the vehicle; **(c)** the weight of the vehicle at that point; **(d)** the mass of the vehicle.

THE GRAVITATIONAL FIELD 6.8

All objects are pulled toward Earth by gravity. We could represent the force of gravity by arrows as shown in Figure 6.14. The arrows are called its **gravitational field**. It is because the Earth has mass that it has gravity, so any object with mass could have a gravitational field as represented in Figure 6.15. The more massive an object is, the stronger its gravitational field.

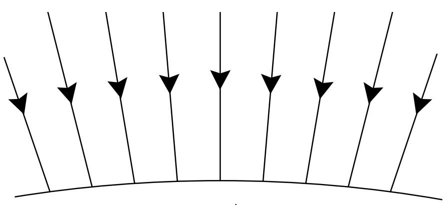


Figure 6.14
The gravitational field is downward, perpendicular to the surface.

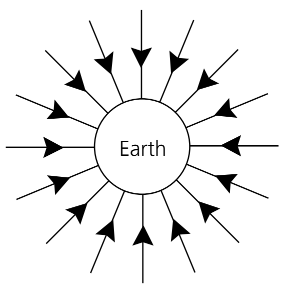


Figure 6.15
The shape of the gravitational field surrounding Earth or any isolated point mass.

Can the arrows point the other way? In other words, can we have **antigravity** — a force that pushes two objects apart? The answer for the moment is no. Unlike electrostatic and magnetic forces (like charges repel, unlike attract), physicists have never observed a repulsive gravitational force, only an attractive one. **Einstein** produced a comprehensive theory of gravity in 1915 — his **general theory of relativity** (see Chapter 30). In it, he argues that gravitational force is different from forces like magnetism and electrostatics even though the mathematical relationships are identical. He said that gravity is not so much something that happens in space but is a distortion or a warp in space itself. His theory encompasses all of Newton's laws but takes them much further and because they have been confirmed by independent research thousands of times, physicists accept his theories as being the best current model for the forces in the universe. Einstein's theories won't be dealt with here but you should have several books in the library on the topic. **Special relativity** is also covered in Chapter 30.

Gravitational field strength

A gravitational field is a region of space where an object experiences a force due to its mass. A measure of the strength of this field is given by the symbol $\boldsymbol{g}$ — the same $\boldsymbol{g}$ you used in acceleration due to gravity. When you think about it, the stronger the field, the faster the acceleration (Newton's second law, $\boldsymbol{F} = m\boldsymbol{a} = m\boldsymbol{g}$):

$$\boldsymbol{g} = \frac{\text{gravitational force}}{\text{mass}} = \frac{\boldsymbol{F}}{m}$$

The Earth has a gravitational field and so does the Moon. Objects placed in the Earth's gravitational field experience a force of attraction given by:

$$\boldsymbol{F} = \frac{Gm_1m_2}{d^2} = m\boldsymbol{g}$$

Hence, the **gravitational field strength** of Earth, $\boldsymbol{g} = \frac{Gm_e}{d^2}$.

Substituting in this formula gives: $\boldsymbol{g} = \frac{6.67 \times 10^{-11} \times 6 \times 10^{24}}{(6.4 \times 10^6)^2} = 9.8 \text{ m s}^{-2}$.

Hence, at the surface of the Earth, $\boldsymbol{g} = 9.8 \text{ m s}^{-2}$, but this value decreases following an inverse square law. At two Earth radii from the *centre* of the Earth, $\boldsymbol{g}$ would be one-quarter of 9.8 m s^{-2} (2.45 m s^{-2}) and at three Earth radii from the centre would be one-ninth of 9.8 m s^{-2} (1.1 m s^{-2}). Further values are shown in Table 6.4.

Novel Challenge

'Zero G' is good for producing genetically engineered crops. Bacteria move more easily in space and it helps the pathogen (bacteria) attach to the surface of the plant and alter it. It is a good way to make soy sauce. What are some other good uses for Zero G?

Table 6.4

ACCELERATION DUE TO GRAVITY (m s^{-2})	DISTANCE FROM EARTH'S SURFACE (km)	DISTANCE FROM EARTH'S CENTRE (km)
9.8	0	6 400
9.0	270	6 670
8.0	670	7 050
7.0	1 160	7 560
5.0	2 540	8 940
3.0	5 150	11 550
2.0	7 740	14 140
1.0	13 590	19 990

Example

Calculate the gravitational field strength in the region of a satellite orbiting 8000 km above the Earth's surface.

Solution

$$m_e = 6 \times 10^{24} \text{ kg}$$

$$d = \text{radius of Earth} + \text{orbiting height} = 6.4 \times 10^6 \text{ m} + 8000 \times 10^3 \text{ m} = 1.44 \times 10^7 \text{ m}$$

$$\boldsymbol{g} = \frac{Gm_e}{d^2} = \frac{6.67 \times 10^{-11} \times 6 \times 10^{24} \text{ kg}}{(1.44 \times 10^7)^2} = 1.93 \text{ m s}^{-2}$$

Questions

29 What is the gravitational field strength at a point whose distance from the Earth's surface is equal to three Earth radii?

30 At what altitude above the Earth's surface must you go for the gravitational field strength to be one-sixteenth the value on Earth?

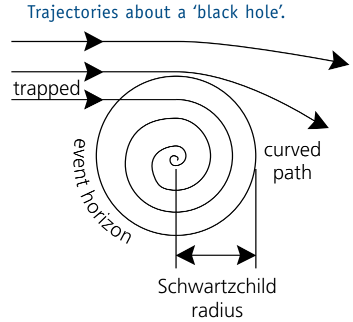
31 Plot a graph of $\boldsymbol{g}$ versus distance from the Earth's centre, using the data from Table 6.4 above. Comment on the relationship. Plot a second graph to prove your prediction.

6.9 BLACK HOLES

A **black hole** is a term used to describe a region of space that contains matter so dense that even light cannot escape its grip. It was coined by John Wheeler of Princeton University (USA) in 1967. A black hole is thought to come about from the **gravitational collapse** of a star. The first tentative identification of a black hole was announced in December 1972 in the binary-star X-ray source Cygnus X-1. The best black hole candidate at the moment is the central star of the triple star system V404, which is 5000 light years (ly) from Earth in the constellation Cygnus. In the case of our galaxy, the Milky Way, there appears to be a two million **solar mass** black hole in the region of Sagittarius A. (One solar mass = mass of the Sun.)

For a black hole with a mass ten times that of our Sun, the point at which light cannot escape (i.e. the point at which it becomes 'black') is within 30 km of the centre. This is called the **Schwartzchild radius**.

Figure 6.12
Trajectories about a 'black hole'.



NOVEL CHALLENGE

Try these Fermi questions:

A How long would a beanstalk (elevated to geosynchronous orbit) have to be? What tensile strength would it require? How does this compare to steel, kelvar, spider silk, the maximum theoretical material strength?

B According to one hypothesis, 20% of the mass of the asteroid that killed the dinosaurs was uniformly deposited over the surface of the Earth at a density of 0.02 g/cm^{-3}. What was the mass of this asteroid?

Falling into a black hole

If a black hole were to exist near some other stellar object such as a **quasar** (quasi-stellar radio source) the gravitational attraction would drag matter from the quasar into the black hole. Atomic particles would accelerate to near the speed of light as they approached. Rather than falling straight in, they would swirl in like a whirlpool, becoming compressed and heated and giving off enormous amounts of energy. It is this radiant energy that astrophysicists detect as they try to identify the location of black holes.

If you tried to travel into a black hole would you survive? It's hardly likely! You'd be stretched, compressed and heated. Hardly the romantic stuff of science fiction novels. If you could avoid getting too close then you wouldn't get sucked in. The minimum distance from which it is still just possible to escape (Schwartzchild radius) marks the boundary called the **event horizon**.

The dilation of time caused by gravity

One of the consequences of Einstein's general theory of relativity is that the passage of time is affected by gravity. To an astronaut falling into a black hole time would pass normally. But to outside observers, for example us on Earth, time would appear to slow down because of the immense gravity near a black hole. It would seem to take ages for the astronaut to disappear into the hole. This is called the **dilation** of time (Latin *dilato* = 'expand').

What's it like inside a black hole?

You can't escape from a black hole — it's sort of like our universe. You can't travel off into space and leave our universe so some scientists have said that our universe is like a black hole in someone else's universe. And their universe is a black hole in some higher universe. Perhaps within black holes in our universe are smaller universes. Who knows? It's all speculation but makes an intriguing thought. Books by Steven Hawking, Carl Sagan, Paul Davies and Kip Thorne examine the possibilities and consequences of such theories. Magazines such as *Scientific American* and *New Scientist* tell of the latest research. There's no room for it here.

What is the fate of a black hole?

Stephen Hawking was the theoretical physicist who showed that black holes eventually evaporate. That's right — evaporate. His technical paper had the unusual title of 'Black holes ain't so black!!' Hawking's calculations confirmed that a spinning black hole loses energy by emitting radiation (the so-called 'Hawking radiation') and as it does it becomes smaller and hotter, eventually becoming so small and hot that it simply evaporates. In fact, today, physicists applying the laws of thermodynamics and quantum gravity require black holes to eventually evaporate in about 10^{67} years. The smaller the black hole, the faster it will evaporate. An interesting theory but who'll be around to see if it's true?

ASTRONOMY VERSUS ASTROLOGY 6.10

Astronomy is the scientific study of heavenly bodies — stars, planets, comets, quasars etc. **Astrology** is a pseudo-science (i.e. non-scientific) that claims to foretell the future by studying the supposed influence of the relative positions of the Moon, Sun and stars on human affairs. Up to the time of Kepler (1600) only astrology existed. Observations and experiments by Galileo, Kepler and Newton produced universal laws and the new science of astronomy began. Astrology slowly became the mumbo-jumbo side of sky watching and was relegated to the irrational, non-scientific and hoaxers club together with pyramid power, clairvoyance, ESP, water divining, flat earth theory, numerology, faxes from the dead, Feng Shui, Tarot Cards, Bermuda Triangle, runes, UFOs, levitation, Philadelphia experiment, faces on the Moon and Elvis sightings, to name just a few.

Astrology has stagnated in the pre-scientific theories of thousands of years ago. It has no testable hypotheses, no statistically reliable evidence of past successes, no research program, no predictive power that can be tested by experiment. Astrologers make many extravagant claims of success but they have never stood up to rigorous scientific scrutiny. In short — astrology is an article of faith, of pseudo-scientific hocus-pocus and rightly belongs in the comic section of the newspaper. There is a group, the *Australian Skeptics*, which examines pseudo-scientific claims and publishes a bi-monthly journal. Visit their Web page at http://www.skeptics.com.au for more information.

SR Activity 6.8 A SIMPLE TEST

Astrologers believe that the positions of planets at the time of birth influence the newborn. Scientists deride this belief and claim that the gravitational force exerted on a baby by the obstetrician or midwife is greater than that exerted by the planets.

1 To check this claim, calculate and compare the gravitational force exerted on a 4 kg baby by **(a)** a 70 kg obstetrician who is 1 m away; **(b)** the massive planet Jupiter ($m = 2 \times 10^{27}$ kg) at its closest approach to Earth ($= 6 \times 10^{11}$ m); **(c)** by Jupiter at its greatest distance from Earth ($= 9 \times 10^{11}$ m). What is your conclusion?

2 Are there any planets that may be lighter but closer that could have some effect? What about the Moon?

3 What new planets have been discovered in the past 2000 years that astrologers conveniently forget about? Where would you look?

6.11 UNANSWERED QUESTIONS

Many of the questions people ask about the structure, history and future of the universe can't be answered with a lot of certainty, but there is consensus among scientists about most of the main theories. There is a lot of debate within the scientific community about many aspects, though. You'll find plenty of books in the library and bookstores in which the authors speculate on the story of the universe. Happy hunting!

Questions you could ask:

- Did the universe create itself?
- Could we possibly know about past cycles of creation if there were any?
- Can we have parallel universes made of antimatter?
- Could astrology work by undiscovered forces?
- Do quantum fluctuations enable the creation of something out of nothing? Yeah right!
- Is there life out there, and what is the SETI project on about?

NOVEL CHALLENGE

If you happened to be a kilometre from the 'Big Bang' when it occurred, what would you hear? We bet you miss the critical problem with this scenario. Good luck!

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

For questions that follow, use the following data:

- $\boldsymbol{g}$ (on Earth) = 10 m s^{-2}.
- $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.
- r_{mo} (radius of Moon's orbit around Earth) = 3.8×10^8 m.
- r_{eo} (mean radius of Earth's orbit around Sun) = 1.5×10^{11} m.

Table 6.5

BODY	MASS (kg)	RADIUS (m)
Earth	$m_e = 6 \times 10^{24}$ kg	$r_e = 6.38 \times 10^6$ m
Moon	$m_m = 7.34 \times 10^{22}$ kg	$r_m = 1.74 \times 10^6$ m
Sun	$m_s = 2.0 \times 10^{30}$ kg	$r_s = 6.96 \times 10^8$ m

Review — applying principles and problem solving

***32** The *Gemini* cluster is 300 Mpc away. Convert this distance to metres.

***33** *Pegasus II* has a radial velocity of 12 800 km/s. What is this in units of 'c'?

***34** The *Coma* cluster is 60 Mpc away from us and has a recession speed of 6600 km/s. Calculate the Hubble constant for this galaxy.

***35** Some early estimates of the Hubble Constant put it at 50 km/s/Mpc.
(a) Would this make the universe older or younger than if a value of 75 km/s/Mpc was used?
(b) What would the age of the universe be if H_0 was 50 km/s/Mpc?

***36** Calculate the wavelength of maximum radiation emitted by the red star Barnard's Star, which has a temperature of 3090 K.

****37** The mean radius of the orbit of planet X in another solar system is 4×10^{12} m. If the average value for r^3/T^2 for this system is 3.5×10^{-18} m^3/s^2, calculate the period of planet Y, which has an orbital radius of 3×10^{10} m.

***38** What is the gravitational force of attraction between:
(a) two apples, of mass 100 g each, placed 30 cm apart on a table
(b) the Earth and the Sun
(c) an electron and a nucleus 1.5 µm apart? The mass of the electron (m_{e-}) is 9.11×10^{-31} kg. Consider the nucleus to be made up of two protons each with a mass (m_{p+}) of 1.67×10^{-27} kg.

***39** How far apart would you have to place two masses each of 1 million kg in order that the force between them was 1.0 N?

***40** The 2270 kg Cosmic Background Explorer (COBE) spacecraft is in a circular polar orbit of radius 900 km and is travelling at 530 m s^{-1}. Calculate
(a) the centripetal acceleration; **(b)** the centripetal force; **(c)** the period.

***41** Ignoring the rotation of the Earth, what is the weight of a 1.5 kg mass
(a) on the surface of the Earth; **(b)** 100 km above the Earth's surface;
(c) in free space?

***42** A satellite of mass 1850 kg is in a orbit 4500 km above the Earth's surface. Calculate **(a)** the gravitational force on the satellite; **(b)** the velocity of the satellite; **(c)** the period of the satellite.

***43** An Apollo spacecraft is orbiting the Earth in a circular orbit 180 km above the surface. If its mass is 3890 kg, calculate **(a)** the force on the satellite; **(b)** the velocity of the satellite; **(c)** the period of the satellite.

***44** Calculate the escape velocity from Ceres, the most massive of the asteroids — mass 1.17×10^{21} kg, radius 3.8×10^{5} m.

***45** **(a)** Calculate the acceleration due to gravity on the surface of the Moon.
(b) What time would it take a spanner to fall (from rest) from a height of 1.5 m to the ground on the Moon?

***46** What is the gravitational field strength at a point **(a)** on the Earth's surface; **(b)** 1.5 Earth radii above the Earth's surface; **(c)** 3.0 Earth radii above the Earth's surface; **(d)** 1000 m above the surface of the Moon; **(e)** on the surface of the Sun?

****47** The constant G can be found by measuring the gravitational force between two spheres of known mass, separated by a known distance. The first person to do this was Henry Cavendish, in 1798, more than a century after Newton proposed his law.

Figure 6.17 shows Cavendish's apparatus. Two small lead spheres, each of mass m, were fastened to the ends of a rod that was suspended from its mid-point by a fine fibre. Large lead balls were brought up close to the small ones. The lead balls attracted each other and caused the fibre to twist. The amount of twist was proportional to the force between the spheres. Cavendish standardised the device beforehand by determining how much force was needed to twist the fibre by certain amounts.

Cavendish used two large lead spheres, each of mass 12.7 kg and two smaller spheres, each of mass 9.85 g. Table 6.6 gives the results for the total force on the fibre with the masses at various distances.

Figure 6.17
Cavendish's apparatus.

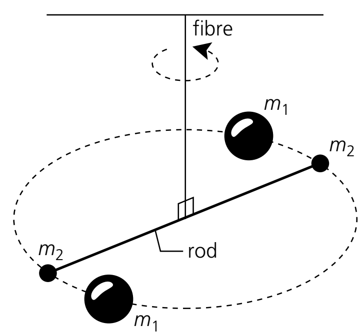


Table 6.6

d (cm)	TOTAL FORCE (N)
5.0	66.8×10^{-10}
8.0	26.6×10^{-10}
10.0	16.6×10^{-10}
12.0	11.6×10^{-10}
13.0	9.9×10^{-10}
15.0	7.4×10^{-10}

(a) Calculate the force between one pair of spheres (divide the total force by 2).
(b) Plot $\boldsymbol{F}$ (y-axis) vs distance. Make sure the distance is in metres.
(c) What relationship does this graph suggest?
(d) Plot another graph to confirm your prediction.
(e) Measure the slope of the graph. This gives the value $\boldsymbol{F}d^2$. Divide it by the product of the masses to get the value for G. How does it compare with the accepted value?
(f) From the graph, determine what force there would be for a distance of 9.5 cm.

PHYSICS FACT

Cavendish didn't actually make the apparatus he used for measuring electrostatic forces (shown above). He inherited it from John Mitchell who died in 1793 before he could try the experiment himself.

**48 The following is based on an excerpt from an article written by Sally Ride. She is a NASA space shuttle astronaut who is Professor of physics and Director of the California Space Institute at the University of California, San Diego. Read the article and answer the questions that follow.

Photo 6.6
Challenger launch.

Adapting to outer space

Astronauts have to adapt to an environment that can't be simulated on Earth. Things in weightlessness seem to be subject to a different set of physical laws. The laws are of course the same but sometimes the implications of those laws are much more apparent. For example, on Earth, frictional effects make it difficult to study Newton's laws of motion. Friction is hard to avoid because of gravity. Gravity holds things in contact with the floor. Newton's laws take some getting used to. A peanut shell set in motion will remain in motion until it hits a wall, a ceiling or somebody's mouth. And a sharp tap on the shoulder can give sufficient impulse to send an astronaut drifting across the room.

When an unanchored astronaut pulls on a drawer the result is frustrating, but predictable. The drawer doesn't open, but the astronaut moves toward the drawer. And if that astronaut uses a screwdriver, the result will be a spinning astronaut, not a turning screw.

Surface tension tends to pull liquids into spheres. On Earth this isn't as obvious: spilled milk lies in a puddle on the floor; in weightlessness, the same milk doesn't splatter on the floor but forms a sphere floating in the middle of the room.

Astronauts eat out of open cartons and use spoons to get the food to their mouths. The trick of course is to use sticky foods. Most of the food is dehydrated and vacuum-packed in plastic cartons with thin plastic tops. It's rehydrated by poking the needle of a water gun through the plastic top and injecting water. Surface tension also causes liquids to creep up drinking straws. Space shuttle straws come with a small clamp to keep the drinks from climbing out. In orbit, a column of liquid has no buoyant effect and no sedimentation. A cork does not bob in water, a bubble would not rise to the surface of a liquid (which means dissolved gases stay in carbonated soft drinks, so they aren't very good to drink), and there would be no layer of chocolate at the bottom of a glass of chocolate milk. As you can imagine, it's a unique living environment.

(from Halliday & Resnick *Fundamentals of Physics*, 3rd edn, 1988)

(a) Which of Newton's laws was the author referring to in her discussion about the drawer?
(b) What would happen to the gases in a soft drink when you drank it?
(c) Milk mightn't splatter but would a tennis ball bounce?
(d) Describe some way astronauts could anchor themselves to the floor.
(e) Friction with the floor is a problem, but does this mean all friction is reduced?

Extension — complex, challenging and novel

***49 How far above the surface of the Earth would you have to be so that a dollar coin (mass 7.4 g) took twice as long to fall 1.5 m when compared with a similar fall on the surface of the Earth?

***50 Describe a planet in terms of its mass and radius that could give **(a)** an acceleration due to gravity half that for $\boldsymbol{g}$ on the surface of Earth; **(b)** a time of flight twice that compared with Earth for a stone dropped from 2.0 m; **(c)** a time of flight twice that on Earth for a stone projected vertically upward at 10 m s^{-1}.

*****51** On a spherical non-rotating planet with a radius of 5.1×10^3 km, what is the value of $\boldsymbol{g}$ for an object when it is at a height of 150 km, expressed as a fraction of its value at the surface?

*****52** In 1610, Galileo used his telescope to discover the four most prominent moons of Jupiter (the Galilean moons). Their mean orbital radii and periods are given in Table 6.7.

Table 6.7

NAME	r ($\times 10^8$ m)	T (DAYS)
Io	4.22	1.77
Europa	6.71	3.55
Ganymede	10.70	7.16
Callisto	18.80	16.70

Plot a graph of r^3 (y-axis) against T^2 (x-axis) and comment on what this graph shows about the relationship between r and T.

*****53** **(a)** At what distance from the Earth would a spacecraft experience zero net gravitational force due to the opposing pulls of the Earth and the Moon?
(b) Express this as a fraction of the total Earth–Moon distance.
(c) Is this the only place you would feel weightless? (See Figure 6.18.)

*****54** **(a)** What is the acceleration due to gravity inside an aeroplane cruising at 10 000 m above the ground? Assume $\boldsymbol{g}$ at ground level is 9.810 m s^{-2}.
(b) How much more time would it take for a coin to fall a metre when inside the plane compared with on the ground?

*****55** Consider a pulsar, a collapsed star of extremely high density, with a mass equal to that of the Sun but with a radius of only 12 km.
(a) Calculate the value of gravitational acceleration.
(b) If the period of rotation is 0.041 s, calculate the value of centripetal acceleration at the equator.

*****56** One clock uses an oscillating spring; a second clock uses a pendulum. Both are taken to Mars. Will they keep the same time there that they kept on Earth? Will they agree with each other? (The mass of Mars is one-tenth that of the Earth and its radius is half that of Earth.)

*****57** Brisbane has a latitude of 28° south, which means that the angle between lines from the centre of the Earth to Brisbane and from the centre to the equator is 28°. (See Figure 6.19.) People in Brisbane still make 1 revolution in 24 hours but travel a smaller distance than someone at the equator.

For a 60 kg person in Brisbane, calculate:
(a) the distance travelled in one day due to the Earth's rotation on its own axis
(b) the linear velocity
(c) the centripetal acceleration
(d) the centripetal force.
(e) In which direction would the following forces be directed:
(i) gravitational force; **(ii)** centripetal force; **(iii)** weight?

Figure 6.18
For question 53**(c)**.

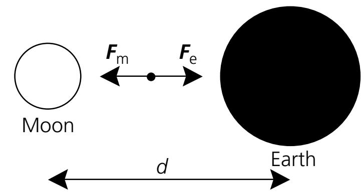
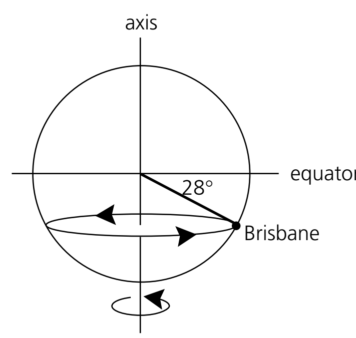


Figure 6.19
For question 57.



✔ TEST YOUR UNDERSTANDING

(Answer true or false)
- The moon is not falling.
- There is no gravity in space.
- Weightlessness means that there is no gravity.
- The Earth's spinning motion produces gravity.
- Planetary orbits are circles.
- Summer is when the Earth is closest to the sun.
- Spacecraft travel in straight lines from one planet to another.
- Black holes are big.
- Light travels in straight lines.

*****58** A spaceship is idling at the fringes of our galaxy, 80 000 light years from the galactic centre. What is the ship's escape velocity from the galaxy? The mass of the galaxy is 1.4×10^{11} solar masses. A light year is the distance light travels in one year at a speed of 3×10^8 m s^{-1}.

*****59** The planet Mars has a satellite, Phobos, which travels in an orbit of radius 9.4×10^6 m with a period of 7 hours 39 minutes. Calculate the mass of Mars from this information.

*****60** The blue stars Procycon and Formalhaut have the same temperature of 7600 K but have radii of 848 000 and 750 600 km respectively.

- **(a)** Determine the wavelength of maximum energy emission of the stars using Wein's law.
- **(b)** Which star would be emitting radiation at the greater rate?
- **(c)** Which star would be emitting the more radiation from an area 1 m^2 in size?

CHAPTER 07

Hydrostatics: The Physics of Fluids

7.1 FLUIDS AT REST?

Fluids play a central role in our daily lives. But what are they? We breathe them, we drink them and they flow through our veins. The sea is a fluid; and the atmosphere; and the core of the Earth. In a car, there are fluids in the tyres, the petrol tank, the radiator, the combustion chambers of the engine, the exhaust pipe, the lubrication system and the hydraulic brakes. Medicine, too, relies on an understanding of fluids in action: the pumping of kidney dialysis machines and the anaesthetist using a heart–lung machine, to name just two. The safe movement of fluids is of vital importance to society. Ruptured oil pipelines and blood vessels are reminders of what can happen if the physics of them is neglected. An understanding of the physics of fluids is essential to society.

You probably already know that fluids include liquids and gases — substances that flow. But they are more than just that. In this chapter we will be considering the science of fluids at rest — **hydrostatics**.

Have you ever wondered about these questions concerning fluids:

- Sugar flows out of a packet when I tip it up. Is sugar a fluid?
- Fresh eggs sink in water but stale eggs float. What is happening?
- In England, sandshoes are called 'plimsolls'. Why is this and who was Plimsoll?
- Why is quicksand so deadly? Why can't you get out?
- Could you walk across a tub of mercury? How far down would you sink?
- Why do some scuba divers die when they come up to the surface too fast?
- Car brake fluid strips the paint off cars if it is spilt. Why don't they use water or oil?
- If a doctor said your blood pressure was '120 over 80' would you care?
- Unwanted pets are sometimes killed by 'decompression'. It sounds cruel — is it?

NOVEL CHALLENGE

Long ago people predicted high and low tides based on observed regularities, but they didn't have a theory of tides. Was this science?

What is a fluid?

A **fluid** is a substance that can flow. Gases and liquids can flow — solids can't. Fluids take on the shape of any container in which they are put. Some materials such as pitch take a long time to flow, but eventually they do, so we call them fluids. Glass in medieval windows is now thicker at the bottom than at the top because it has slumped a bit. Glass too is a fluid, although a very **viscous** (thick) one. 'Viscous' comes from the Latin *viscum* meaning 'birdlime'. This was a sticky gum applied to branches to trap bird pests in orchards. Of liquids, the least viscous is liquid helium (–270°C), which is so mobile that it creeps up and over the sides of its container.

But maybe everything will flow eventually so maybe everything is a fluid! What's the point of distinguishing between fluids and solids? The point is: no, not everything is a fluid. Ice isn't, steel isn't, bricks aren't. A fluid is a system of particles loosely held together by their own attractive forces or by the restraining forces exerted by the walls of the container. A fluid

Photo 7.1
Pitch drop experiment. Professor John Mainstone of the Department of Physics at the University of Queensland, with the current pitch drop about two years after the previous drop.

will flow even if the forces are very weak. A solid will not flow at all unless the applied forces are in excess of some threshold value. Pitch and glass are fluids of high viscosity. Even if the force is small they will flow, although very slowly. (See photo of pitch drop experiment on previous page.) However, in practical and useful terms, we can think of fluids as being substances that can be pumped along pipes.

PRESSURE 7.2

When a fluid is placed into a container it exerts a force on any surface exposed to it. The magnitude of this force divided by the area over which the force acts is called **pressure**.

$$\text{Pressure} = \frac{\text{force}}{\text{area}} \qquad P = \frac{F}{A}$$

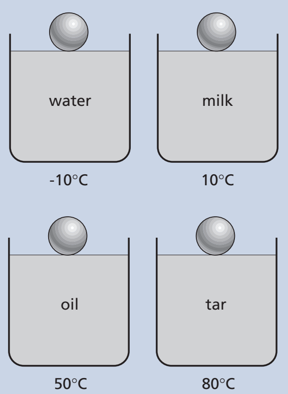
If the force is in newtons (N) and the area in square metres (m^2), the pressure is in **pascals** (Pa). One pascal equals one newton per square metre. Although pressure is defined in terms of a vector quantity, it is not a vector quantity itself. The unit *pascal* was named after the French scientist and mathematician **Blaise Pascal** (1623–62). Pascal suffered from a condition known as a soft fontanelle, in which the cartilage between the bones of the skull never properly hardened. This gave rise to migraine headaches so severe that it halted his scientific thinking. Nevertheless he made huge contributions to science and philosophy during his 39 years on Earth. Although to physicists he is best remembered for his work on pressure, in general he is remembered for his remarkable insights in religious thinking, fragments of which are recorded in his book *Pensées*. A computer language has since been named in his honour.

NOVEL CHALLENGE

In setting the world record for a bed of nails, a 60 kg man lay down on 259 nail points in a 30 × 45 cm board. The contact area for each nail with his skin was 10 mm^2. Then a 268 kg weight was placed on top of him. Calculate the pressure of the nails on his skin.

NOVEL CHALLENGE

Which of the following ball bearings will fall most slowly?



NOVEL CHALLENGE

You can exert a force of 250 N with your incisor teeth and 1220 N with your molars. *Which do you estimate to produce the higher pressure?* Note: *your front incisors are about 8 mm × 0.2 mm and your molars are about 8 mm × 8 mm.*

NEI Activity 7.1 HIGH PRESSURE

The *Guinness Book of Records* lists the highest atmospheric pressure ever recorded on Earth as 108.38 kPa in Siberia in 1968. Where was the lowest recorded; what was it and when?

Table 7.1 SOME PRESSURES

	PRESSURE (Pa)
Centre of the Sun	2×10^{16}
Centre of the Earth	4×10^{11}
Highest sustained laboratory pressure	2×10^{10}
Deepest ocean trench	1×10^{10}
Spike heels on dance floor	1×10^{6}
Car tyre	3×10^{5}
Atmosphere at sea level	1×10^{3}
Blood pressure	2×10^{4}
Loudest tolerable sound	30
Faintest detectable sound	3×10^{-5}
Best laboratory vacuum	1×10^{-12}

Example

A person has a mass of 65 kg. The contact area between his shoes and the floor is 315 cm^2. Calculate the pressure he exerts on the floor.

Solution

Force (weight) = mg = 650 N; area = 315 cm^2 = $\dfrac{315}{100 \times 100}$ m^2 = 0.0315 m^2.

$$P = \frac{F}{A} = \frac{650}{0.0315} = 20\ 600 \text{ Pa}$$

You can have a high pressure without a large force. A chisel (Figure 7.1) has a sharp tip and when a small force is applied it will easily penetrate wood (and your shoe if you drop it). A small force of 1 newton applied to the handle will be transferred to the point and produce a very large pressure. If the point has an area of 10^{-6} m^2, the pressure will be:

$$P = \frac{F}{A} = \frac{1}{10^{-6}} = 10^6 \text{ Pa (1 megapascal)}$$

Because many surfaces cannot stand this pressure, the chisel will penetrate them. Hardwood or pine are good examples.

This concept helps us to understand the action of knives, needles and nails. They work because of the small contact area. On the other hand, army tanks and bulldozers work on the reverse principle — the bigger the surface area of their caterpillar treads, the less likely they are to sink into muddy ground. Four-wheel drive owners will know that this technique can be applied when driving in loose sand on a beach. Tyres can be deflated to half normal pressure to increase the surface area and hence lessen the amount they sink into the sand (Figure 7.2).

NEI

Activity 7.2 ROADS AND TYRES

Roads gradually break up from the constant pressure of passing car and truck tyres. But big trucks are not always the worst offenders as their load is often spread over eighteen or more tyres.

Part A

- Use a ruler to measure the contact area of a bicycle tyre with a flat surface while a person is sitting on the bike. Measure or estimate the total mass of the bike plus rider and calculate the pressure exerted by the tyres on the surface. Assume that half the weight is supported by each tyre.
- Repeat for a car. The mass of the car will be in the owner's handbook. Most cars are between 1000 kg and 2000 kg.

Part B

Rank the following in order of the pressure exerted by the tyres on the road. Include the two results from above.

- A BMX bicycle; mass 20 kg plus rider 45 kg; contact area 10 cm × 5 cm per tyre.
- A Porsche 911 Carerra, mass 1370 kg, on four tyres, each with a contact area of 15 cm × 20 cm.
- A Landcruiser, mass 2960 kg, on four tyres, each with a contact area of 17 cm × 22 cm.
- A fully laden semi-trailer of mass 42 t; 22 tyres, each with a contact area of 20 cm × 20 cm.

Questions

1 A girl with a weight of 500 N stands in snow on a pair of skis. Each ski has a contact area of 1.5 m × 0.13 m. Calculate the pressure on the snow.

2 Calculate the pressure at the bottom of a round swimming pool of diameter 6.0 m filled to a depth of 1.2 m.

NOVEL CHALLENGE

The pressure in an aeroplane's tyre was measured with a pressure gauge at sea level. The plane flew off and landed on a high mountain airstrip. If the temperature was the same as at sea level, how would the pressure gauge reading compare with that at sea level?

Figure 7.1

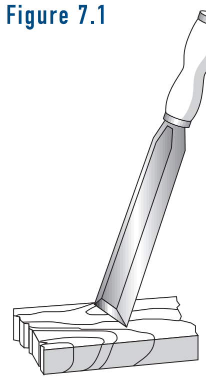
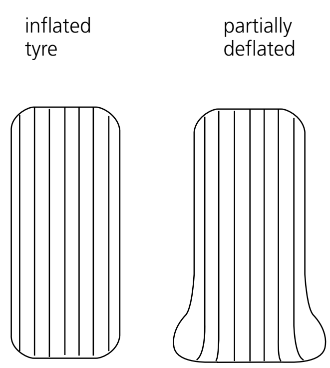


Figure 7.2
In soft sand it often helps to deflate the tyres to about half the normal pressure.



NOVEL CHALLENGE

Four car tyres are inflated to the same pressure. One wheel is jacked up.
How does this change the pressure in the jacked-up tyre and in the other three?

3 In some shops and factories, cheese is cut by a wire. Imagine a steel wire with a diameter of 0.2 mm being pulled through a block of cheese 20 cm wide by the downward force of a 5 kg mass (50 N). Calculate the pressure acting on the cheese.

MEASURING PRESSURE 7.3

Atmospheric pressure

The gas particles of the atmosphere have weight and they exert a pressure on us and the surface of the Earth. It can be calculated that about 150 t of air pushes down on just the floor of a living room. This is the atmospheric pressure and has a value of 101 325 pascal (Pa) or 101.325 kPa. The higher you go up into the atmosphere the smaller the amount of air above you and hence the lower the pressure.

Table 7.2 CHANGES IN ATMOSPHERIC PRESSURE

ALTITUDE (m)	PRESSURE (kPa)
15 000	12.0
11 000	22.5
9 000	31.0
6 000	47.1
3 000	70.0
0 (sea level)	101.3

Concrete floors are able to stand very high pressures. Concrete is extremely resistant to compression forces but relatively weak to stretching forces. Concrete batching plants can supply different mixes of concrete depending on its purpose. Household concrete slabs typically use a 20 megapascal (20 MPa) blend, whereas high-rise columns need 80 MPa concrete.

NOVEL CHALLENGE

A fluorescent light tube is stood upright in a bucket of water and a small hole is cut in the tube underwater with a triangular file. The gas pressure inside the tube is 330 Pa. To what height do you estimate the water will rise? Try it but first think of how you will dispose of the water-filled tube.

NOVEL CHALLENGE

An empty soft-drink can has some water in it and is boiled over a Bunsen burner. It is quickly inverted and stood up in a tray of cool water. *What do you predict will happen? Try it, but use tongs.*

NOVEL CHALLENGE

A piece of burning paper is placed in a conical flask and a boiled egg placed in the top. You can imagine what happens. But how to reverse the process without touching the egg — now that's where the physics is needed.

NEI Activity 7.3 CONCRETE STRENGTH TESTING (CST)

1 Find out the procedure for testing the strength of concrete made by the ready-mix batching plants. They only sample one in 20 truckloads. Why? What do they do with the concrete samples after testing?

2 Find out the specifications (in MPa) for several types of concrete available from the batching plant. What makes them different? Are they sold by the tonne or cubic metre? Does their cost vary?

3 Reinforced concrete and pre-stressed concrete both have steel bars or cables inside them. Why is this?

4 Testing laboratories express the strength of concrete in kilonewtons per square millimetre. How is this related to megapascals?

Measuring air pressure

Three common ways of measuring air pressure are with the **barometer**, the **manometer** and the **Bourdon gauge**.

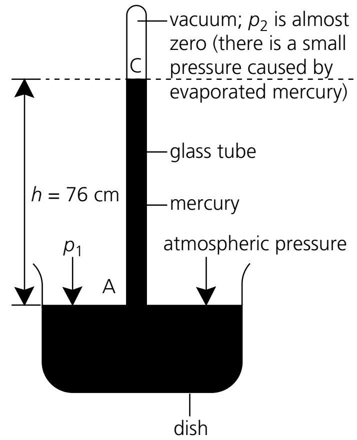
The barometer is used to measure atmospheric pressure. The word is derived from the Greek *baros* meaning 'weight' and hence refers to the weight of the atmosphere. The manometer also measures pressure, not of the atmosphere but of some other gas relative to atmospheric pressure. In Greek, *manos* means 'rare', referring to gases at low pressure where the particles are few and far between.

The barometer

The invention of the mercury barometer (1643) by Evangelista Torricelli arose from his realisation that air has weight. He noted that if the open end of a glass tube filled with mercury is inverted in a bowl of mercury, the atmospheric pressure (p_1) on the bowl of mercury will affect the height of the column of mercury in the glass tube. The greater the air pressure, the longer is the mercury column. Normal atmospheric pressure will support a column of mercury 760 mm high. The symbol for mercury, Hg, is from the Greek *hydro-argentum* meaning 'silvery water'. Hence normal atmospheric pressure can be written as 760 mmHg. (See Figure 7.3.)

Mercury is ideal for a liquid barometer, since its high density permits a short column, whereas a water barometer would be 10 m tall at normal atmospheric pressure.

Figure 7.3
The mercury barometer.



The aneroid barometer (Bourdon gauge)

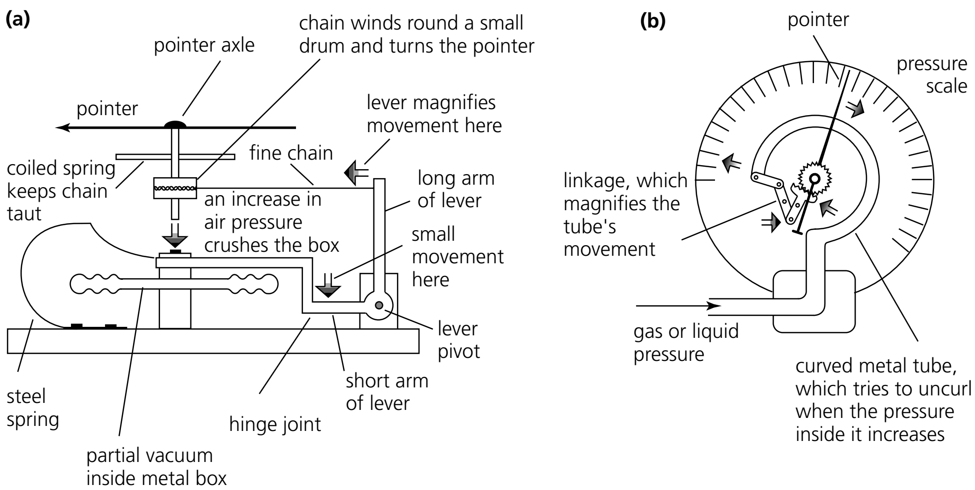
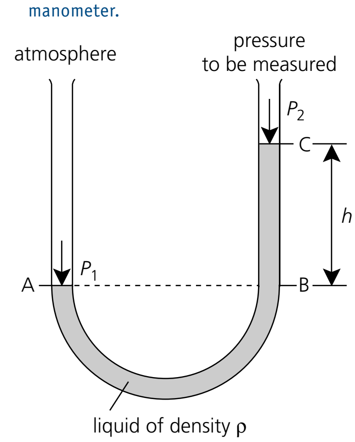


Figure 7.4
When the gas pressure inside the curved metal tube increases, the tube tries to uncurl. The end of the tube is linked to a pointer, which reads pressure on a circular scale.
(a) The aneroid barometer.
(b) The Bourdon gauge.

Most barometers are of the aneroid type and function without liquid. The aneroid barometer, dating from 1843, consists of a small metal box, almost totally evacuated of air. One side is immovable, and the opposite side is connected to a strong spring to keep the box from collapsing. The movable side will expand if the air pressure decreases and will compress if the air pressure increases. The position of the movable side is indicated by a pointer. An aneroid barometer is checked regularly against a mercury barometer for **calibration**.

The aneroid barometer can be easily converted into a barograph, or recording barometer, by adding a pen to the pointer. The ink in the pen describes a trace (**barogram**) on the paper wrapped around a cylinder. The cylinder usually rotates once a day or once a week. Whereas the mercury barometer is used in research laboratories and in important weather stations, aneroid barometers are used in the home, on board ships, and in most weather stations.

Figure 7.5
Measuring pressure with a U-tube manometer.



The manometer

The manometer is the most direct and accurate instrument for measuring liquid and gas pressures of moderate range in the laboratory or in industry. In its common form, known as the U-tube manometer, a tube is partially filled with a liquid such as mercury, oil, or water. With one open end exposed to the atmosphere and the other end to the pressure or vacuum source to be measured, the pressure is determined by noting the difference in level of the liquid in the tube branches. Typically, the reading is in millimetres of mercury (mmHg). A manometer

Photo 7.2
A doctor using a sphygmomanometer.

measures the difference between atmospheric pressure and the pressure in a connected vessel. The reading is not the absolute pressure but the **gauge pressure**. Car tyres, for instance, are inflated to about 200 kPa above atmospheric pressure. Their absolute pressure would then be 300 kPa (100 kPa for the atmosphere and 200 kPa for the extra in the tyres) but a tyre pressure gauge only registers 'gauge pressure' over and above atmospheric pressure. Pressure cookers have gauges that register in the same way.

Blood pressure

Blood pressure is measured with an instrument called a **sphygmomanometer** (Greek *sphygmo* = 'pulse'). It consists of an inflatable cuff, which is wrapped around a patient's arm, and is connected to an open-tube mercury manometer. To take the pressure, a stethoscope is placed over the arteries of the arm just below the cuff and the pulsations of blood in the arteries can be heard. As air is pumped into the cuff, it cuts off the flow of blood, and the sounds stop. Then air is slowly let out of the cuff. When the pressure of the cuff becomes less than the blood pressure, the blood flow returns. The pressure at which the flow of blood starts up again is called the **systolic** pressure, which can be read off the manometer. As more air is let out of the cuff, the sounds become muffled. The pressure at this point is called the **diastolic** pressure.

Blood pressure is expressed as systolic pressure over diastolic pressure. The systolic blood pressure is recorded during the instant that the heart contracts (systole) to force blood into the circulation; it is always higher than the diastolic blood pressure, which is recorded when the heart relaxes between beats (diastole). Thus, blood pressure of 120/80 mmHg means that the systolic (maximum) pressure is 120 and the diastolic (minimum) pressure is 80 mmHg. Readings above 140/90 are usually regarded as high (hypertension). As you get older your normal blood pressure usually gets higher anyway.

Modern electronic sphygmomanometers use a microphone to detect the pulse sound and a pressure sensitive resistor (a strain gauge) to measure the force. However, they need calibration against a mercury sphygmomanometer, which is simple and very reliable.

NOVEL CHALLENGE

The sphygmomanometer was invented by René Laënnec in 1816. He used a rolled-up tube of paper to listen to a very fat patient.
Why isn't the cuff wrapped around the lower part of your arm? Why wouldn't you measure the pressure in that big artery in your neck?

EI Activity 7.4 BLOOD PRESSURE

There is probably a sphygmomanometer in your school, possibly in the biology lab. If you have access to it, take your own blood pressure. If you can borrow a stethoscope, listen to the blood flow and see if you can identify the Korotkoff sound. If you are concerned about your blood pressure you should visit your doctor to have it measured properly.

PASCAL'S PRINCIPLE 7.4

If you squeeze an inflated balloon a bulge pops out somewhere else on the balloon. The increased pressure that you apply is transmitted to the rest of the gas inside the balloon. French scientist Blaise Pascal formulated a principle — **Pascal's principle**:

> Pressure applied at any point to a fluid in a closed vessel is transmitted equally to every other point in the fluid.

Gases can be compressed because the distances between particles are very large compared with the size of the particles themselves. With a liquid, however, the particles are held closely together by a variety of attractive forces including dispersion forces, dipole–dipole forces and hydrogen bonding. Because the particles are in close contact they can't be compressed. If you study chemistry, these forces will be described in detail.

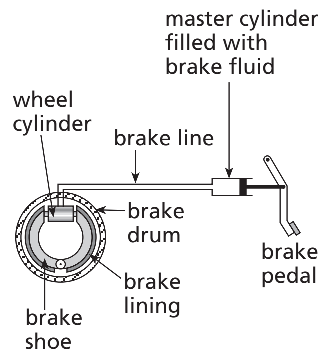
Ch07-Walding 4th 25/8/04 11:10 AM Page 170

Pascal's principle in real life

Many devices in common use rely on Pascal's principle. The hydraulic car jack, the hydraulic hoist in a car workshop, the hydraulic brakes in a car and the hydraulic press used to form sheet metal parts are all examples. The word **hydraulic** is made up of *hydra* meaning 'water' and *aulos* meaning 'a pipe'.

Figure 7.6 shows how hydraulic brakes in a car work. When foot pressure is applied to the brake pedal, a small piston is made to move inside a cylinder (the master cylinder). This piston pushes an oily brake fluid along a steel tube to another cylinder (the slave cylinder), which is located on the wheel axle. The fluid pressure moves a piston, which pushes on the brake callipers, forcing a brake pad to press against a brake disk. With drum brakes, the slave cylinder pushes brake shoes against the drum. In the past, hydraulic fluid was made from castor oil and alcohol, but the demands of modern cars on their braking systems are a lot higher than they used to be. A lot of heat is generated by brake friction so a liquid with a high boiling point is needed. For example, high speed pursuits by the London police generated brake fluid temperatures of 188°C, which was sufficient to boil normal brake fluid so a blend of synthetic polymers (glycol ethers and borate esters), corrosion inhibitors and fluid modifiers, with a boiling point of about 280°C, was developed. Any vapour that forms in the hydraulic brake line will make the brakes feel 'spongy' and render them ineffective. Brake fluid has to be 'bled' out of the line to remove any vapour bubbles if this happens. Brake fluids absorb moisture and if the moisture content reaches 1.5%, the boiling point will be about 155°C, the minimum allowed for safe driving. One drawback of modern brake fluid is that it can make a mess of paintwork but this is a small price to pay for a safe braking system. In other hydraulic systems on trucks, graders and bulldozers, a light oil can be used because no heat is generated.

Figure 7.6
Hydraulic brake system for a car with drum brakes.

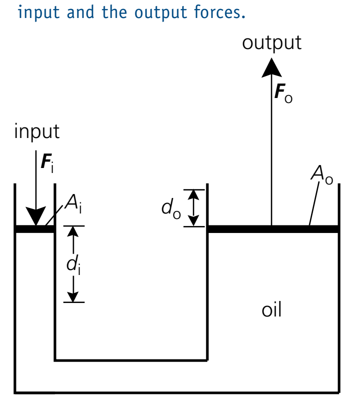


Pascal's principle is also the basis of the hydraulic hoist in a motor garage. An external force (usually a motor) forces a small piston downward. The force is transmitted by an incompressible liquid, in this case water with some soluble oil included as a lubricant, to a larger piston as shown in Figure 7.7. A small input force moving through a large distance equals a large output force moving through a small distance.

In the next chapter you will see that the product of force × distance is called work ($W = \boldsymbol{F}s$). The product of $\boldsymbol{F} \times s$ on input equals the product of $\boldsymbol{F} \times s$ on output. In motor garages, a hydraulic hoist is used to lift a car up high so that it can be worked on from underneath.

In terms of Pascal's principle the relationship can be written:

Figure 7.7
A hydraulic hoist, used to magnify force $\boldsymbol{F}_i$. The work done by this force, however, is not magnified and it is the same for both the input and the output forces.



Pressure in small cylinder = pressure in large cylinder

$$P = \frac{F_1}{A_1} = \frac{F_2}{A_2}$$

Thus if a force of 20 N was applied to a piston of diameter 1 cm, it would produce a force of 20 000 N on a piston 30 cm in diameter. This force would lift 2000 kg — that of a car.

Example

The input piston in a hydraulic hoist has a diameter of 1.50 cm, whereas the output piston is much larger at 25.0 cm. The output piston has to lift a total weight of 30 500 N. Calculate **(a)** the pressure that has to be applied; **(b)** the force needed on the small piston; **(c)** the distance the input piston will move if the output piston moves 1.50 m.

Solution

(a) $P = \frac{\boldsymbol{F}_2}{A_2} = \frac{30\,500}{\pi \times 0.125^2} = 621\,340$ Pa.

(b) $\frac{\boldsymbol{F}_1}{A_1} = \frac{\boldsymbol{F}_2}{A_2}$; $\boldsymbol{F}_1 = \frac{\boldsymbol{F}_2 \times A_1}{A_2} = \frac{30\,500 \times p \times 0.0075^2}{p \times 0.125^2} = 110$ N.

(c) $\boldsymbol{F}_1 \times s_1 = \boldsymbol{F}_2 \times s_2$; $s_1 = \frac{\boldsymbol{F}_2 \times s_2}{\boldsymbol{F}_1} = \frac{30\,500 \times 1.50}{110} = 4.15$ m.

Questions

4 A hydraulic hoist in a truck workshop has to be able to lift 85 000 N. The small piston has a diameter of 5.5 cm, whereas the output piston has a diameter of 45.0 cm. Calculate **(a)** the force needed on the small piston; **(b)** the pressure that has to be applied; **(c)** the distance the input piston will move if the output piston moves 1.80 m.

ARCHIMEDES' PRINCIPLE 7.5

Archimedes' principle is the fundamental natural law of buoyancy, first identified by the Greek mathematician and inventor Archimedes in the third century BC. He was once asked by King Hieron II to work out if a gold crown being given to the King was pure gold or if it contained impurities of silver and copper. The story is told that he was sitting in a bath and noticed that objects underwater seemed lighter in weight than they did in air. On realising this he is said to have run naked around the streets of Syracuse shouting '*Eureka*', which translated from the Greek is 'I have found it'.

Most of the story is fiction but he did discover some facts about floating and sinking. Incidentally, Archimedes worked out a way to tell if the crown was pure gold, as you'll see later. He also did some experiments and found that the volume of water that overflowed from a filled container of water was equal to the volume of the object placed in it. It seems commonsense to say that, but you should remember that experimental science was in its infancy 2000 years ago. Archimedes later showed that the **upthrust** (upward force) on an object equals the weight of water displaced. This is Archimedes' principle. It states:

NOVEL CHALLENGE

I was in the swimming pool holding a steel spanner and I thought the spanner should feel lighter because of the buoyant upthrust. But it felt **heavier**. To my hand it really felt heavier. *By why? It slowly dawned on me!*

> When an object is wholly or partially immersed in a fluid, the upthrust on the object is equal to the weight of the fluid displaced.

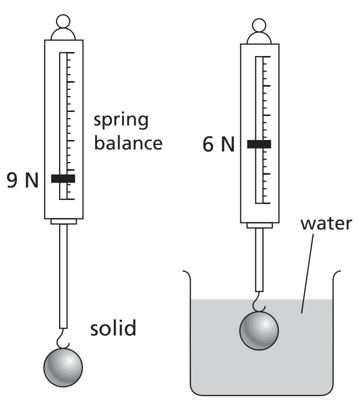
For instance, you probably know that things feel lighter in water than in air. The loss in weight of an object (the upthrust) is equal to the weight of water that it has displaced. If it is fully submerged the volume of water displaced is the same as the volume of the object. The weight of this volume of water is equal to the upthrust.

In general:

> Weight in air – apparent weight (in the fluid) = upthrust

The weight of an object in the fluid is better called the 'apparent' weight (or 'scale reading').

Figure 7.8
When a heavy object is placed in water, it experiences an upthrust. In this case the upthrust is 3 N.



Example

A housebrick has a volume of 1900 cm^3 and a weight in air of 80 N. What is its apparent weight in water? The density of water is 1.00 g cm^{-3}.

Solution

- Volume of water displaced = volume of brick = 1900 cm^3.
- Mass of water displaced = density × volume = 1.00 × 1900 = 1900 g = 1.9 kg.
- Weight of water displaced = mg = 1.9 × 10 = 19 N.
- Weight in water (apparent weight) = weight in air – upthrust = 80 N – 19 N = 61 N.

Note: some people prefer to use the units *kg m*$^{-3}$ for density. While this is the correct usage of SI units, *g cm*$^{-3}$ is commonly used. The conversion is: 1 g cm^{-3} = 1000 kg m^{-3}.

Specific gravity

A litre of water has a mass of 1.00 kg, but the same volume of methylated spirits has a mass of only 0.79 kg. We can say that the density of methylated spirits is 0.79 that of water. This is called its **relative density (RD)** or its **specific gravity (SG)**. The two terms are interchangeable and although relative density is more correct, specific gravity is widely used in industry and the other sciences.

The specific gravity (SG) is defined as the ratio of the mass of an object in air compared with the mass of an equal volume of water.

$$\text{Specific gravity} = \frac{\text{mass of object in air}}{\text{mass of equal volume of water}}$$

But it can also be written:

$$\text{specific gravity} = \frac{\text{weight of object in air}}{\text{weight of equal volume of water}}$$

According to Archimedes' principle, the weight of an equal volume of water equals the upthrust (weight loss) when it is submerged. Thus:

$$\text{specific gravity} = \frac{\text{weight in air}}{\text{weight loss when submerged in water}}$$

or

$$\text{specific gravity} = \frac{\text{weight in air}}{\text{weight in air} - \text{apparent weight in water}}$$

$$\text{SG} = \frac{W_A}{W_A - W_W}$$

In the example above, this equation gives a value of 80/19 = 4.2 g cm^{-3}. This value can be confirmed by dividing the mass of 8 kg (= 8000 g) by a volume of 1900 cm^3, which also equals 4.2 g cm^{-3}.

Example

A crown supposedly made of gold weighs 8.00 N in air. If the SG of gold is 19.3, what should be the crown's apparent weight in water?

Solution

$$\text{SG} = \frac{\text{weight in air}}{\text{weight loss when submerged in water}}$$

$$\text{weight loss} = \frac{\text{weight in air}}{\text{specific gravity}} = \frac{8.00}{19.3} = 0.415$$

The crown's apparent weight should be 8.00 N – 0.415 N = 7.59 N.

EI Activity 7.5 SPECIFIC GRAVITY OF IRREGULAR OBJECTS

1 Tie a piece of cotton thread around a rock and suspend it from a spring balance. Note the scale reading of the rock in air and again when the rock is fully submerged in water. Calculate the SG.

2 Try it with an iron bolt. Did you get an SG of 7.8?

Questions

5 When the brass cannons from the wrecked eighteenth century Dutch trading ship *Batavia* were recovered off the West Australian coast in 1988, they found that the cannons had a mass of 1100 kg when lifted out of the water. If the density of brass is 8400 kg m^{-3}, what would their apparent weight have been in the water?

6 About 800 iron cannonballs were recovered from the *Batavia*. If each ball had a diameter of 13 cm, calculate **(a)** their mass in air; **(b)** their weight in air; **(c)** their weight in water. The density of iron is 7.8 g cm^{-3} (7800 kg m^{-3}).

NOVEL CHALLENGE

Density of some fluids (in g cm^{-3})

Petrol	0.68
Alcohol	0.81
Water	1.00
Sea water	1.03
Sugar syrup 40%	1.15
Mercury	13.6

*1 What would have the greater density: **(a)** low-alcohol beer or normal beer; **(b)** Coke or Diet Coke?*

2 Cans of Coke and Diet Coke were put in a tub of water. Will they both float, sink or one float, one sink? Try it.

NOVEL CHALLENGE

A cube of brass measuring 2 cm along its side was placed in a measuring cylinder containing 50 mL of water.

What volume would the measuring cylinder read now? Another cube with a side twice as big was also added. What would the final volume read?

Ch07-Walding 4th 25/8/04 11:11 AM Page 173

7.6 FLOATING AND SINKING

Fresh eggs sink in water but stale eggs float. Why is that? Fresh eggs are more dense than water — that explains them. But in stale eggs, some of the contents have diffused through the shell and been replaced by air — and they float. They become less dense than water.

So not everything sinks in water. Wood floats, ice floats, polythene floats and so do stale eggs. Archimedes' principle still applies but the volume of water displaced will not be equal to the total volume of the object.

Imagine a 1000 cm^3 block of wood floating in water so that four-fifths of its volume is underwater. Thus, the volume of water displaced equals 800 cm^3 and this has a mass of 800 g or a weight of 8 N. Hence, by Archimedes' principle, the upthrust must also equal 8 N. As the block is floating, the weight of the block must also equal 8 N. As this is equivalent to a mass of 800 g, the density of the wood can be calculated:

$$\text{Density } (\rho) = \frac{m}{V} = \frac{800}{1000} = 0.8 \text{ g cm}^{-3}$$

Note:

The density of a solid floating in a fluid is equal to the density of the fluid times the fraction of the volume submerged.
Density of object = density of fluid × fraction submerged

For regular solids, the fraction submerged can be calculated from measurement of the vertical height under and above the fluid.

Example 1

A cube of polythene floats with seven-eighths of its volume below the water level. Calculate the density of the polythene.

Solution

$$\text{Density of object} = \text{density of fluid} \times \text{fraction submerged} = 1.00 \times \frac{7}{8} = 0.875 \text{ g cm}^{-3}$$

Example 2

A 21 cm long polypropylene drinking straw floats upright in a bottle of water with 2.8 cm of its length above the surface. Calculate the density of polypropylene.

Solution

$$\text{Density of object} = \text{density of fluid} \times \text{fraction submerged}$$
$$= 1.00 \times \frac{21 - 2.8}{21} = 0.87 \text{ g cm}^{-3}$$

Activity 7.6 DENSITY BY FLOTATION

1 Prepare 10 cm lengths of the following. They should be about 1 cm wide.
- The side of a 2 L plastic milk container (high-density polyethylene); SG 0.95–0.97.
- An icecream bucket (low density polythene); SG 0.91–0.94.
- A plastic drinking straw (polypropylene); SG 0.905.
- A wooden paddlepop stick (pine); SG 0.03–0.04.

NOVEL CHALLENGE

The relative density of two types of wood is as follows: ironbark 1.3, balsa 0.24. Balsa is specified as 10 pounds per cubic foot. Is this about right?

NOVEL CHALLENGE

One of our students can float up to her earlobe in a swimming pool when floating upright.
A Estimate her body density.
B How high would she float in seawater (SG 1.030 g cm^{-3})?
C Would she float higher or lower in cold water?
D Would a boy of the same height and mass float the same way?

NOVEL CHALLENGE

The density of salt water is greater than that of fresh water. So, swimming in salt water should be faster as your body floats higher and therefore there is less friction. *True or false, and why?*

2 Float them upright in a filled bottle of water with a narrow neck.
3 Measure the fraction underwater and calculate the density of each.
4 Compare your results with the values given above.
5 How could you estimate what your own density would be in a swimming pool?

Questions

7 A rubber ball floats with 40% of its volume above water. What is its density?
8 A floating piece of wood displaces 80 cm^3 of water. Find the weight and mass of the wood.

The hydrometer

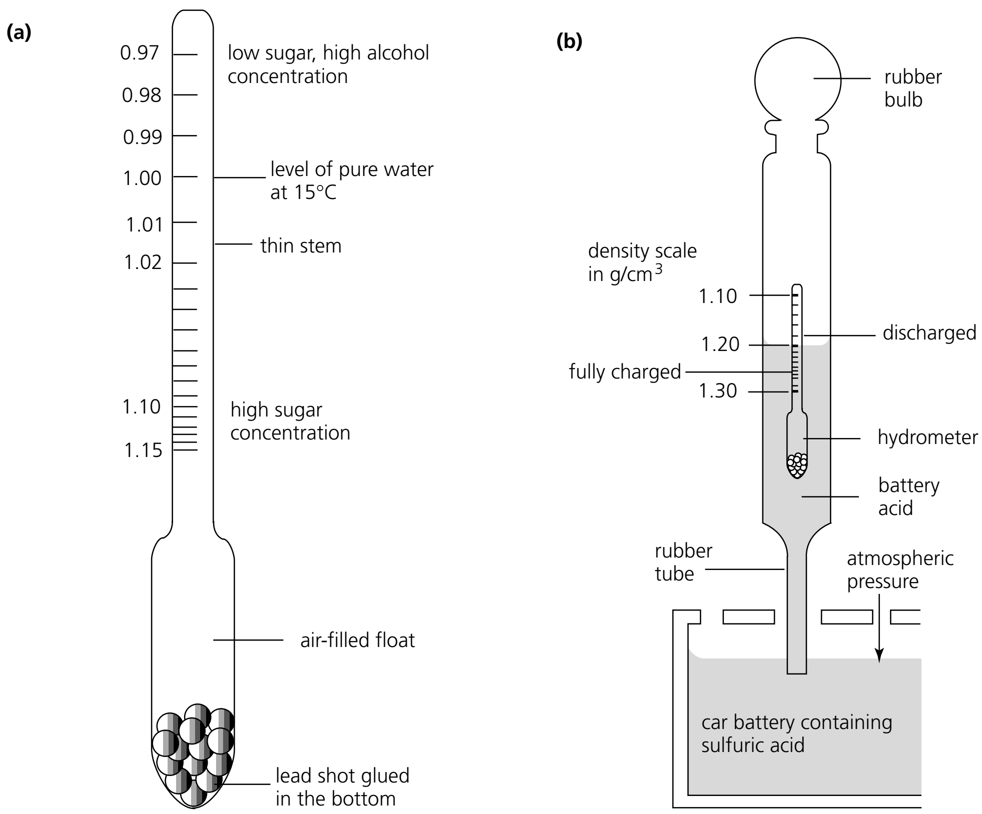


Figure 7.9
(a) A brewer's hydrometer;
(b) A battery acid hydrometer.

The **hydrometer** is a device for measuring the specific gravity of battery acid, antifreeze solutions, milk, alcohol and other liquids. When industry buys alum solution for water purification the concentration is checked simply by measuring the SG with a hydrometer. The depth to which a graduated glass float sinks in the fluid is proportional to the density of the fluid. The less dense the fluid, the deeper the hydrometer float sinks. The hydrometer for car batteries measures the density of the sulfuric acid in the battery and indicates the state of charge of the battery. A low density indicates that the battery needs recharging.

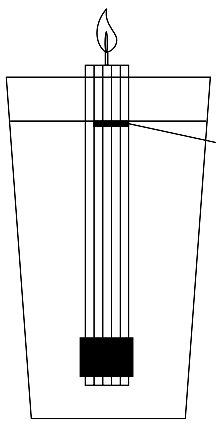
NOVEL CHALLENGE

A boy is 1.8 m tall and can float upright in pool water with only 5 cm above the water. When he breathes in, his body rises so that it is 25 cm above the surface. Can you estimate the change in density of his body?

NOVEL CHALLENGE

Continents float on the liquid mantle of the Earth. Continents have a density of 2800 kg m^{-3} whereas the mantle has a density of 3300 kg m^{-3}. If a continent is 35 km thick, prove that the top of the continent is 5.3 km above the mantle surface.

Figure 7.10
A floating candle — will it rise or fall?



Activity 7.7 FLOATING CANDLE

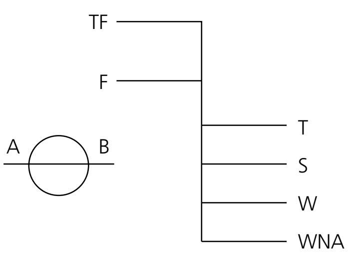
Predict the result of this activity before you attempt it.

1 A candle is weighted at its bottom with some lead or nails taped to it so that it floats upright in water with just a centimetre or two above the surface.
2 Mark the water level on the candle with a pen or a nail (Figure 7.10).
3 Light the candle.
4 As the candle burns, will the mark rise, fall or stay the same in the water?

Quicksand — help!

If you fall into quicksand, people say you get sucked under. This is not true — you will float just as in water. But the underground spring feeding it can wander rapidly and often the spring gets diverted a metre or so by your presence so the sand seems to 'set', making it difficult to get out. You won't sink below your armpits but, with your legs trapped, your life expectancy is short. Does anyone have any suggestions?

Figure 7.11
The Plimsoll line is now called the international load line. The symbols stand for: T = tropical, S = summer, W = winter, TF = tropical freshwater, WNA = winter north Atlantic. The A and B on the circle stand for the agency that assigned the load line to the ship. In this case it was the American Bureau of Shipping (AB).



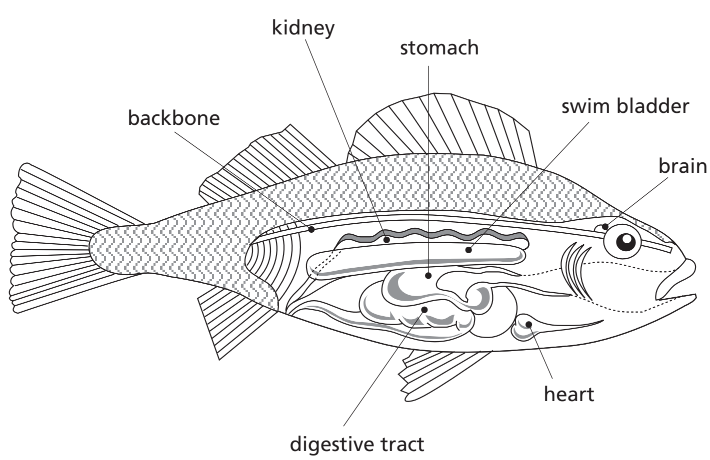
The floating of ships

Women in general float higher in water than do men because women (in general, we stress) have more fat on their bodies than men and fat has a lower density than muscle tissue. But not only is the density of the floating object important, but so too is the fluid in which it floats. Ships will float at different levels depending on how salty and how warm (and hence, how dense) the water is. There are special marks on the sides of ships, called the **Plimsoll line** (Figure 7.11), which indicate how deeply a fully laden ship can safely float in water of different types. It was named after Samuel Plimsoll, a member of the British Parliament who introduced the mark in 1885. Until then, ships could be loaded to any level and were seriously overloaded by unscrupulous owners to cut costs. Plimsoll was known as the 'sailor's friend'.

In England, sandshoes with a green line around their sides are called 'plimsolls' after the Plimsoll line.

Fish

Figure 7.12
The internal organs of a bony fish.



Fish have a baglike organ called a swim bladder just below their backbone (Figure 7.12). This provides buoyancy, which enables the fish to remain at a particular depth in the water.

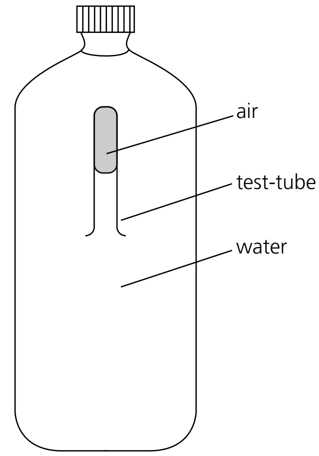
A fish would sink to the bottom if it did not have a way of keeping buoyant. It gains buoyancy by inflating its swim bladder with gases from its blood. But since water pressure increases with depth, a fish's swim bladder would get smaller as it descends and this would reduce its buoyancy. The amount of gas in the bladder must be increased so that the bladder volume is just right to maintain buoyancy. This is done automatically by the fish's nervous system. Sharks and rays do not have a swim bladder. To keep buoyant, these fish must swim constantly. When they rest, they stop swimming and so sink toward the bottom. Many bottom-dwelling bony fish also lack a swim bladder.

Activity 7.8 CARTESIAN DIVER

SR

Fill a 2 L plastic soft drink bottle to about 5 cm from the top with water. Half-fill a test-tube with water and up-end it into the soft drink bottle (Figure 7.13). If the test-tube doesn't float, take it out and remove some of the water. Once you can get it to float upright, screw the cap on the bottle. Squeeze the bottle and watch the 'Cartesian diver'. What happens to the water level inside the test-tube? Can you explain what is going on?

Figure 7.13
Cartesian diver — squeeze the bottle and the diver sinks.



Balloons and blimps

The Montgolfier brothers, Joseph and Jacques, were inventors of the first practical hot-air balloon. They used paper balloons to help English soldiers escape from the Spanish at Gibraltar in 1782, but these caught fire several times. The two Frenchmen were papermakers by trade and discovered in 1782 that smoke from a fire directed into a silk bag made the bag buoyant. In 1783 they gave a public exhibition of their discovery with a balloon that rose to an altitude of about 2 km and stayed aloft for 10 minutes. They later put a sheep, duck, and rooster aboard the balloon to determine the effect of altitude on living creatures.

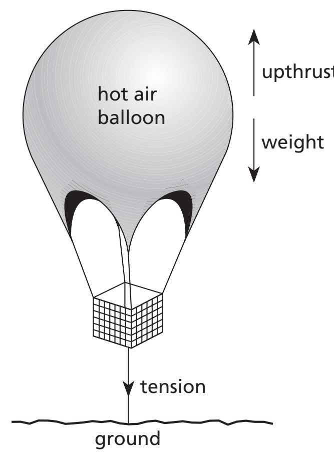
The modern non-rigid blimp has no internal structure to maintain the shape of its hull envelope, which is made of two or three plies of cotton, nylon, or dacron impregnated with rubber for gas tightness (Photo 7.3). Inside the gas space of the hull are two or more air diaphragms called ballonets that are kept under slight pressure, either by blowers or by air that is forced through scoops as a result of the forward motion (ram effect). The ballonets in turn exert pressure on the gas, which fills the envelope, and this pressure in turn serves to stiffen the shape of the envelope and create a smooth flying surface. On take-off the ballonets are almost fully inflated, but as the airship gains altitude and the gas expands, air is bled from the ballonets while a constant pressure is maintained throughout the envelope. When the gas contracts on descent, air is pumped back into the ballonets.

In 1991, Westinghouse Airships launched the 68-m long *Sentinel 1000*, the first in a projected series of blimps to be used by the US Defence Department for a range of surveillance, communications, and patrol duties. The envelope of the *Sentinel 1000* is made of a mix of synthetic fibres that is impervious to weather and almost invisible to radar.

The principles behind the balloon and the blimp are similar to those of objects floating in water. The air is a fluid; it has a density of about 0.0012 g cm^{-3} or 1.2 kg m^{-3}. An object such as a balloon displaces a certain volume of air and so experiences an upthrust equal to the weight of the air displaced. As long as the balloon and its contents are lower in weight than the weight of air displaced, the balloon will rise. To make up for the weight of the balloon fabric, ropes and basket, the balloon has to be filled with a gas lighter than the surrounding air. Hot air, hydrogen and helium are commonly used. Because helium gas is such an expensive and a non-renewable resource, party balloons are filled with 'balloon gas', which is mostly nitrogen but with some helium mixed in.

Photo 7.3
A blimp.

Figure 7.14



Example 1

A large balloon is filled with hot air to a volume of 400 m^3. It has a total weight of 4400 N and is held to the ground by a vertical rope (Figure 7.14). Given that the density of the surrounding air is 1.2 kg/m^3, calculate the tension in the rope.

Solution

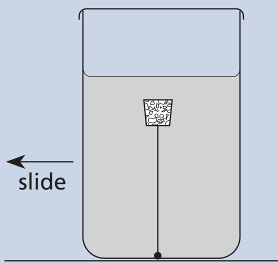
- Mass of air displaced = density of air × volume = 1.2 × 400 = 480 kg.
- Upthrust = weight of air displaced = mg = 480 × 10 = 4800 N.
- Resultant force (tension) = 4800 N (up) − 4400 N (down) = 400 N (up).

PHYSICS FACT

The Wright brothers published their account of the first flight in the journal *Gleaning on Bee Culture*. As manned flight didn't exist at the time there were no aviation journals — so a bee journal was the next-best thing.

NOVEL CHALLENGE

A plastic soft-drink bottle is half-full of water. A small piece of cork is held just under the surface by a piece of string glued to the bottom. The bottle is slid to the right. *Which way does the cork move relative to the bottle — forward, backward, sideways, no movement? You'll be shocked if you try.*



Example 2

A weather balloon has a mass when deflated (empty) of 5 kg. It is inflated to its volume of 8 m^3 with helium, which has a density of 0.178 kg m^{-3}. Find the lifting force on the balloon when the surrounding air has a density of 1.20 kg m^{-3}.

Solution

- Mass of air displaced = density × volume = 1.2 × 8 = 9.6 kg.
- Upthrust = weight of air displaced = mg = 9.6 × 10 = 96 N.
- Total mass of balloon = mass of balloon + mass of helium = 5 + (8 × 0.178) = 6.4 kg.
- Weight of balloon = mg = 6.4 × 10 = 64 N.
- Lifting force = 96 N (up) − 64 N (down) = 32 N.

Questions

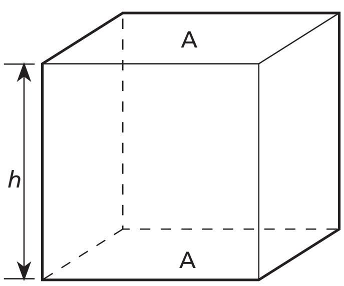
9 A balloon is filled with hot air to a volume of 650 m^3. It has a total weight of 6000 N and is held to the ground by a vertical rope. Given that the density of the surrounding air is 1.18 kg m^{-3}, calculate the tension in the rope.

10 Moving natural gas from the North Sea gas fields in huge dirigibles (blimps) has been proposed, using the gas itself to provide lift. Calculate the force required to tether such an airship to the ground for off-loading when it is fully loaded with 1×10^6 m^3 of natural gas at a density of 0.80 kg m^{-3}. The density of air is 1.18 kg m^{-3}. Neglect the weight of the airship.

11 The Goodyear blimp *Columbia* is cruising slowly at low altitude, filled as usual with helium. Its maximum payload including crew and cargo is 1280 kg. How much more could it carry if the helium was replaced with hydrogen? The volume of the interior space is 5000 m^3; the density of helium is 0.16 kg m^{-3} and the density of hydrogen is 0.08 kg m^{-3}.

PRESSURE AND DEPTH 7.7

Figure 7.15



When you dive to the bottom of a swimming pool you can feel the increased pressure on your eardrums and lungs. When you go up a tall mountain you can feel your ears 'pop' because of the decreased air pressure. **Pressure increases with depth** because there is a greater weight of fluid on top of you.

The pressure exerted by a column of fluid on its base can be calculated by working out the weight of fluid on a given area (Figure 7.15). The base of the column in Figure 7.15 has an area A and a height h. The density of the fluid is given the symbol ρ.

- Volume of fluid in column = Ah
- mass of fluid = $Ah\rho$
- weight of fluid = $Ah\rho\boldsymbol{g}$
- pressure = $\dfrac{\text{weight}}{\text{area}} = \dfrac{Ah\rho\boldsymbol{g}}{A} = \rho\boldsymbol{g}h \quad P = \rho\boldsymbol{g}h$

where P = pressure in Pa, ρ = density in kg m^{-3}, A = area in m^2, h = height in m, $\boldsymbol{g}$ = 10 m s^{-2}.

Note that pressure is independent of the area of the base. The pressure at the bottom of a large dam is no different from that at the bottom of a swimming pool if they are both the same depth.

Example

Determine the pressure due to the water at the bottom of a 12 m deep dam. Fresh water has a density of 1000 kg m^{-3}.

PHYSICS FACT

Could you have an object with a vacuum inside that would float in air? Answer: yes! You could use a titanium sphere 44 m in diameter with a wall thickness of 2 mm. It would have a 29 N upthrust. If it was 310 m in diameter and a wall thickness of 13.9 mm it could lift 1000 kg.

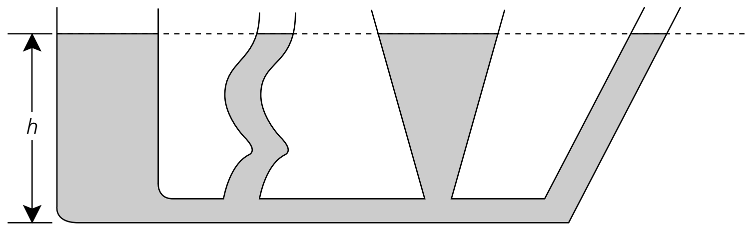


Figure 7.16
Pressure at a point in a fluid at rest is independent of the size or shape of the containing vessel.

Solution

$$P = \rho \boldsymbol{g} h = 1000 \times 10 \times 12 = 120\,000 \text{ Pa } (120 \text{ kPa})$$

Note: this is only the pressure due to the water. The total pressure includes that of the atmosphere on top of the water (+101.3 kPa).

The 'bends'

Scuba divers breathe a mixture of oxygen and the inert gases nitrogen and helium. Under pressure, the inert gases diffuse into the blood and other tissues. If the pressure is relieved too quickly by rising to the surface too fast, bubbles form in the tissues much as they do when a bottle of soft drink is opened. Sudden decompression from a long, deep dive can be fatal; even a slight miscalculation can cause serious injury to the joints or the central nervous system.

This problem is called **decompression sickness** (DCS) or 'the bends'. The most effective treatment for DCS is recompression. The diver is placed into a recompression chamber (RCC) and the pressure is increased according to a specified treatment table. The increased pressure reduces the bubble size, which helps them to diffuse back into the blood. The diver is compressed to an equivalent depth of 18 m of water and then decompressed over a period of 2–5 hours. The diver breathes oxygen from a mask while the rest of the chamber is filled with air, not oxygen, because of the fire risk. Because the attendant sits inside the chamber and breathes chamber air, great care must be taken to monitor his time and pressure profile (dive profile) to avoid the embarrassment of having an attendant emerging from the RCC with DCS.

INVESTIGATING

Did people make paper planes before the Wright brothers flew their aeroplane, Kitty Hawk, in 1903?

NOVEL CHALLENGE

Definitely a tough one! A hollow steel ball with a wall thickness of 2 cm and outside diameter of 12 cm is placed in acid and it sinks. As the steel dissolves evenly from the sphere, suddenly it floats. Prove that this happens when the wall is 0.25 cm thick. The steel has a density of 7.8 g/cm^3 and acid has the same density as water. *Note:* V (sphere) = $\frac{4}{3}\pi r^3$.

Diving barotraumas

As well as the 'bends', divers can suffer other problems when rising to the surface. The lungs of a diver normally contain about 6 L of air. If the diver takes a full breath of air at 20 m depth and rises to the surface, that 6 L volume expands to 18 L; in order to avoid bursting his lungs, the diver must exhale 12 L of air on the way up. Gas must be exhaled about every metre otherwise the pressure in the lungs will be sufficient to rupture them. This is called **lung barotrauma**. It is second only to drowning as a cause of death in recreational scuba divers. As swimming pools are more than 1 m deep, lung barotrauma has occurred in backyard pools.

Other types of barotraumas caused by gas expansion in the body while diving at depth are:

- mask squeeze (gas in the mask is compressed and can cause bleeding of eye tissues)
- gastrointestional barotrauma (gas in the gut expands and can cause cramps, belching and vomiting)
- dental barotrauma (gas pockets in decayed teeth may allow the teeth to implode on descent (going down) or to explode on ascent). It is not common. Diving sounds like fun!

Photo 7.4
Magdeburg Hemispheres. The original hemispheres were devised by Otto von Guericke, were about 30 cm in diameter and, in a famous demonstration in Magdeburg (Germany) in 1654, the air was removed and even two teams of horses (16 of them) couldn't pull the hemispheres apart.

Pressure in the atmosphere

The relationship between pressure and altitude for a gas such as air is more complicated than the relationship between pressure and depth for a liquid because the density of a gas is not constant. It depends on the pressure. The pressure in a column of air decreases as you go up from ground level, but unlike the pressure in a water column, the decrease in air pressure with distance is not linear.

Activity 7.9 ATMOSPHERIC PRESSURE CHANGES

SR

Part A

Plot the following data (Table 7.3), which show the variation in pressure with height above Earth's surface. The pressure halves for each 5.5 km rise in altitude.

Table 7.3

ALTITUDE (m)	PRESSURE (kPa)
16 500	9.1
15 000	12.0
12 000	19.5
11 000	22.5
9 000	31.0
6 000	47.1
5 500	50.6
3 000	70.0
0 (sea level)	101.3

(a) Estimate the pressure at 10 km and at 18 km.
(b) Where would pressure be zero?
(c) What type of relationship is this: inverse, inverse square, exponential or what?
(d) In 1692 Newton said that pressure and density decreased exponentially with altitude. Was he right?

Part B

The pressure in an aeroplane tyre is taken at ground level with a tyre pressure gauge. The plane flies to the top of a high mountain and the tyre pressure is taken again. Assume the air temperature and the mass of the plane are constant. How do the pressures compare? Even engineers argue about this one. Good luck!

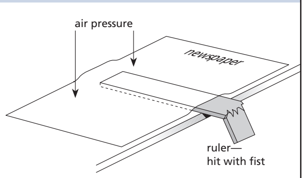
As you may gather from your graph from the above activity, in space the pressure is just about zero for there is approximately only one particle per cubic metre. Without a pressurised space capsule or space suit, animals, including humans, couldn't survive. Not only would their ears pop as the internal body pressure exploded outwards, but eyes and blood vessels would also pop. A cruel death would intervene. Unwanted pets were once put to 'sleep' by decompression at animal pounds but as the process was too distressing for both the animals and the operators the method has generally been discarded.

Examples of decompression

- The *Los Angeles Times* reported that a flight attendant was wearing an inflatable bra when the cabin depressurised during flight. The air expanded according to Boyle's law (Chapter 11) and inflated the bra to size 46, until a woman passenger stabbed her strategically with a hat pin. This sounds like an urban myth to us!
- When a tunnel under London's Thames River had been completed and the two shafts joined, the local politicians celebrated the event at the tunnel's bottom. In the tunnel they found the champagne flat and lifeless. When they returned to the surface, however, 'the wine popped in their stomachs, distended their vests, and all but frothed from their ears'. One dignitary had to be rushed back to the depths to undergo recompression.

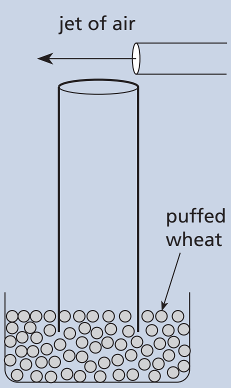
NOVEL CHALLENGE

A 30 cm wooden ruler was placed on a bench with 10 cm overhanging. A page of the *Courier Mail* (58 cm × 40 cm) was placed on top of the ruler on the bench. *Calculate the weight of air on the paper and predict what will happen if the overhang is given a sharp blow with your fist. You won't believe the weight. Quick now — is it more than the weight of three Corollas?*



NOVEL CHALLENGE

A cardboard tube is placed halfway into a container of puffed wheat. *What do you think will happen when you blow air across the top of the tube? But why, and what does Bernoulli have to do with it?*



Questions

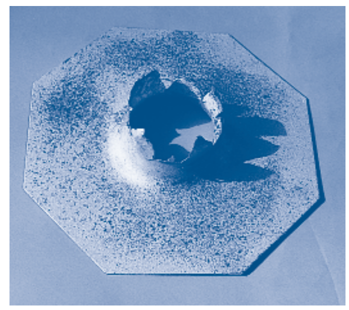
12 Calculate the pressure in kPa at the bottom of a column of mercury 76 cm high. The density of mercury is 13 600 kg m^{-3}.

13 What is the total pressure (water + atmosphere) on a diver 20 m under sea water that has a SG of 1.03? Assume atmospheric pressure is 101.3 kPa.

7.8 IN CLOSING

As humans explore new frontiers, research into extremes of pressure assumes great importance. At the University of Queensland, their 'hypervelocity shock tunnel' is being used in the mechanical engineering department to study the effect of shock waves on various objects such as spacecraft. An enormous piston compresses hydrogen gas to extreme pressures, which then blasts its way through a steel plate (Photo 7.5) to provide the high velocities needed for experiments. But just as important for us is the knowledge that stale eggs float, air in your brakes is bad and high blood pressure is a worry.

Photo 7.5
The blast-hole in the steel plate used in the shock tunnel.



Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***14** A person has a mass of 65 kg. The contact area between his shoes and the floor is 315 cm^2. Calculate the pressure he exerts on the floor.

***15** Calculate the pressure on the ground due to a Ford Falcon, mass 2240 kg, on four tyres, each with a contact area of 15 cm × 17 cm.

***16** A round swimming pool of diameter 4.5 m was filled to a depth of 1.1 m. Calculate the pressure at the bottom if you **(a)** neglect atmospheric pressure; **(b)** include atmospheric pressure (101.3 kPa).

***17** The input piston in a hydraulic hoist has a diameter of 2.50 cm, whereas the output piston is much larger at 29.0 cm. The output piston has to lift a total weight of 26 500 N. Calculate **(a)** the pressure that has to be applied; **(b)** the force needed on the small piston; **(c)** the distance the input piston will move if the output piston moves 2.0 m.

***18** A rock has a volume of 800 cm^3 and a weight in air of 33 N.
(a) What is its weight (scale reading) in water?
(b) Calculate the density of the rock. The density of water is 1.00 g cm^{-3}.

***19** Petrol has a density of 0.8 g cm^{-3}.
(a) What is its specific gravity?
(b) What is its density in kg m^{-3}?

***20** The most dense gas known is radon (Rn) with a density of 0.010 05 g cm^{-3} at room temperature and pressure.
(a) What is the mass of a 15 L balloon full of it under these conditions?
(b) Convert the density to kg m^{-3}.

***21** A mass of small colourless mineral was measured in air using a spring balance and then its apparent mass was measured in water. The spring balance read 16.5 g in air and 10.6 g when the specimen was in water. Calculate the SG.

***22** A pair of Scuba tanks has a volume of 22.4 L and a mass of 24 kg when full. What is their weight (scale reading) in **(a)** air; **(b)** salt water?
(c) What mass would they have to be to have neutral buoyancy in this water? The SG of salt water is 1.02.

NOVEL CHALLENGE

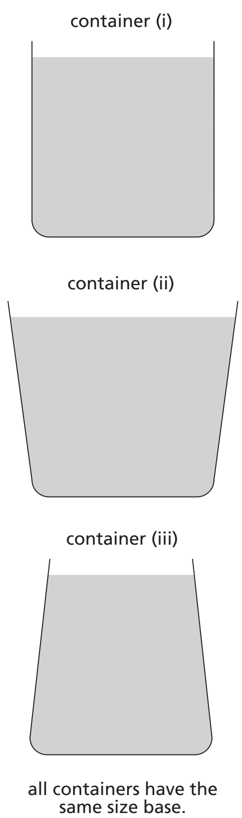
A square pond measures 100 m by 100 m. A block of ice with a mass of 1000 kg is floating freely in the pond. How far will the water level rise when the ice melts? You won't like the answer. (The density of water is 1000 kg m^{-3}; that of ice is 917 kg m^{-3}.)

***23** A float made of polystyrene foam floats with one-fiftieth of its volume below the water level. Calculate the density of the polystyrene.

****24** A 50 cm piece of high density polythene water pipe is made to float upright in a container of water. If the pipe has a density of 0.89 g cm^{-3}, how much of the pipe will be above the surface?

***25** When a cube of ice is placed in tap water, 94% of its volume is submerged.
- **(a)** What is the density of the ice cube?
- **(b)** How much would be submerged if it was floated in salt water of density 1.02 g cm^{-3}?

Figure 7.17
For question 33.



****26** During the Second World War, a damaged freighter that was just able to float in the salty water of the North Sea sank as it came up the Thames estuary toward the London docks. Why?

****27** Beer hydrometers are calibrated at 15°C and if the temperature is different, a correction has to be applied. If a hydrometer placed in a beer 'wort' at 25°C gave a reading of 1042 (1.042 g cm^{-3}) would you expect the real density to be higher or lower than this value? Explain your answer.

****28** A girl has several large rocks in a row boat and she rows out into the middle of a pond. To make some room she throws the rocks overboard.
- **(a)** What happens to the water level on the side of her boat?
- **(b)** What happens to the water level in the pond?
- **(c)** Design your own experiment to test your answer in the classroom. Write down your results. They may surprise you.

***29** Why do you sometimes need to punch two holes in a can of pineapple juice to make it come out evenly?

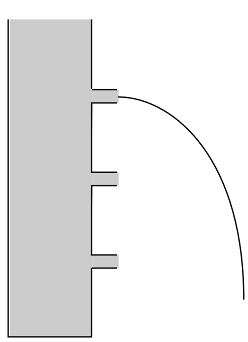
****30** Hydrogen appears to have negative weight as you can't weigh a balloon full of it on a balance. Design an experiment to weigh a litre of hydrogen gas.

Extension — complex, challenging and novel

*****31** A steel ball bearing is placed in a bowl of mercury.
- **(a)** Given that the density of steel is 8.0 g/cm^3 and that of mercury 13.6 g cm^{-3}, calculate what fraction of the volume of the sphere is submerged.
- **(b)** Suppose we pour some water on top of the mercury. Does the steel ball sink more deeply or less deeply in the mercury?

*****32** Crew members attempt to escape from a damaged submarine 100 m below the surface.
- **(a)** What force must they apply to a pop-out hatch, which is 1.2 m by 0.60 m, to push it out?
- **(b)** What mass is this equivalent to lifting? Could a lone sailor manage? Assume the density of sea water is 1025 kg m^{-3}.

Figure 7.18
For question 34.



*****33** Three containers are set up and filled to the same height with water (Figure 7.17).
- **(a)** Which has the greatest volume of water in it?
- **(b)** Which has the greatest weight of water in it?
- **(c)** Which has the greatest mass of water in it?
- **(d)** Which has the greatest pressure at the bottom?

*****34** A can 40 cm tall has small holes in it at 10 cm, 20 cm and 30 cm from the base (see Figure 7.18). It is filled with water. Where do the streams of water strike the ground? To make this calculation you need to use Bernouli's equation, which can be reduced to: pressure = $\rho \boldsymbol{g}h = \frac{1}{2}\rho \boldsymbol{v}^2$. This reduces further to $\boldsymbol{v} = \sqrt{2\boldsymbol{g}h}$, where h equals the height of water above the hole, and $\boldsymbol{v}$, the velocity of the water out of the hole.

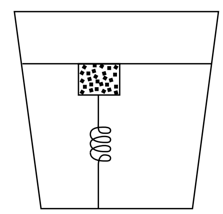
****35** Imagine a glass of water sitting on a table with a few cubes of ice floating in it. The water is level with the top of the glass and the ice projects above the top. As time goes by the ice melts. Will the water overflow the glass?

*****36** Consider two balloons of equal volume (400 000 L) and mass (200 kg), one filled with hydrogen, the other with helium. Helium is twice as dense as hydrogen (hydrogen — 0.09 g/L; helium — 0.18 g/L). The density of air is 1.3 g/L.
- **(a)** How do their lifting abilities compare?
- **(b)** Will the hydrogen balloon be able to lift twice weight as the helium balloon? Explain.

*****37** Two identical buckets are filled to the brim with water but one has a block of wood floating in it. Which bucket is the heavier?

*****38** When a cork is placed in a bucket of water it floats with one-quarter of its volume submerged. Imagine we attach a small spring to the cork and attach it to the inside bottom of the bucket and adjust it so that it is just submerged (the top of the cork is now level with the surface of the water) as in Figure 7.19. The weight of the cork and tension in the spring now equal the upthrust. The bucket is dropped off a tall building. What would happen to the cork during the fall?

Figure 7.19
For question 38.



*****39** The *Guinness Book of Records* lists the greatest ocean descent as that of the US Navy underwater research vessel (a bathyscaphe) the *Trieste*, which reached a depth of 10 916 m on January 23 1960. It says that the pressure at this depth was 1187 kgf/cm^2 and the temperature 3°C. A *kgf* is a non-SI unit known as a *kilogram force*. This is the force of gravity acting on 1 kg. Convert the pressure to pascals and calculate the average density of seawater above the bathyscaphe. Neglect atmospheric pressure.

*****40** Calculate the difference in blood pressure between the brain and the foot of a person of height 1.83 m. The density of blood is 1.06×103 kg m^{-3}.

*****41** The human lungs can operate against a pressure difference of about one-twentieth of an atmosphere. If a diver uses a snorkel for breathing, how far below the surface can he or she swim?

*****42** About one-third of the body of a person swimming in the Dead Sea will be above the water line. Assuming that the density of a human is 0.98 g cm^{-3}, find the density of the water in the Dead Sea.

*****43** Imagine a U-tube containing some mercury. The mercury level is equal on both sides of the tube. An equal volume of water is added to one side. How will their levels now compare? Make a statement about the relative heights of the two columns.

*****44** Prove that hydrogen provides only an extra 8% lift compared to helium for a balloon of mass 2.13 kg, even though helium is twice as dense as hydrogen. See Example 2 on page 177 for density data.

*****45** A 20 cm diameter spherical beach-ball floats with 1 cm submerged. Calculate the mass of the ball. You'll need extra maths formulas.

*****46** A cylindrical log 1 m long and 20 cm diameter floats with 8 cm submerged. Calculate its mass and density. Extra maths formulas required.

Ch08-Walding 4th 25/8/04 11:24 AM Page 183

UNIT 03

Energy & Momentum

CHAPTER 08

Momentum

8.1 EXPLOSIONS, COLLISIONS AND BALLET DANCING

Ballet, bullets, bombs, baseball, boxing and binary stars have something in common. They all involve the combination of mass and velocity. They all involve **momentum**.

These are some questions that a study of momentum can help answer:

- Would you rather be hit by a 1 g ball-bearing travelling at 100 m s^{-1} or by a 100 g ball travelling at 1 m s^{-1}?
- Which would hurt more — being tackled by a lightweight footballer travelling at high speed or by a big fat one travelling at low speed?
- Police bullets are designed to stay inside their targets and not go through them. How?
- You are standing in the middle of a frictionless frozen pond. Someone with a little knowledge of physics says that you can't get to the edge because of the laws of momentum. What can you do to prove him wrong?
- A cat that falls out of a window upside-down can right itself and land on its feet. How can this be if it has nothing to push against while falling? What's the advantage of landing on its feet anyway?
- When you shoot a bullet at a watermelon suspended on a string, the melon moves towards you. How can this be and what has it to do with the assassination of JFK?

8.2 CENTRE OF MASS

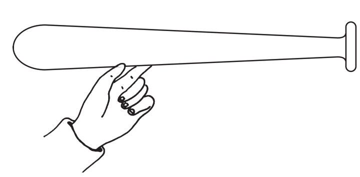
Physicists love to look at something complicated and find in it something simple and familiar. If you toss a cricket bat into the air its motion as it turns is more complicated than that of a cricket ball. A diver who executes a somersault has an even more complicated motion still. Every part of the body moves in a different way from every other part, so you cannot represent the body as a single particle as you can with a ball. However, if you look closely, you will find that there is one special point that moves in a simple path — a parabola — much as the ball does. This point is called the **centre of mass**. It is the point at which the whole mass of an object is considered to be concentrated for the purpose of applying the laws of motion.

EI Activity 8.1 CENTRE OF MASS

You can locate the centre of mass of a bat by finding the point at which it balances on an outstretched finger.

1. If you can get hold of some different bats or racquets, find the balance point of each and mark it with a felt pen. Draw diagrams to show the location.

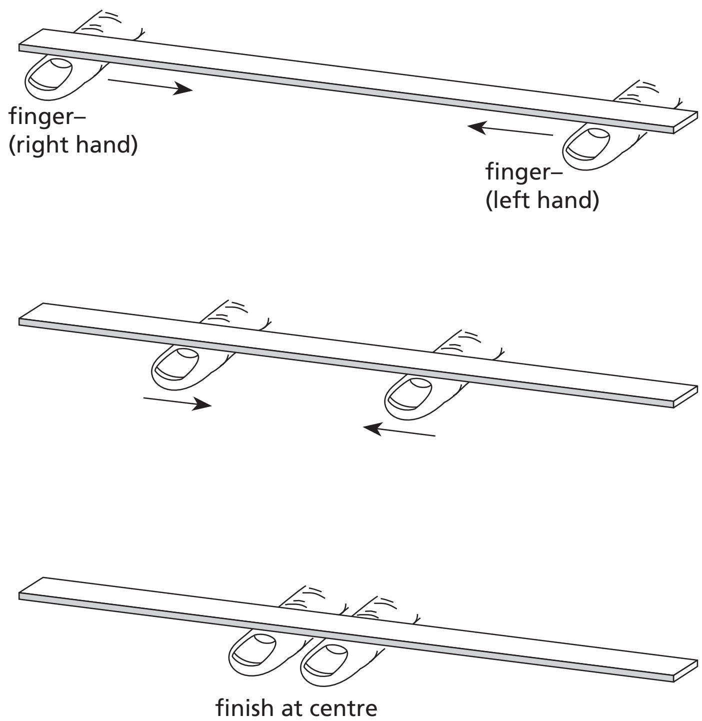
Figure 8.1
The balance point of a baseball bat.



2 How does the centre of mass compare with the centre of percussion as discovered in Chapter 6? (Recall letting the bat swing like a pendulum and finding its effective length.)

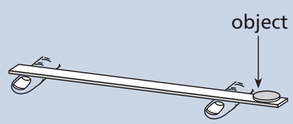
3 Hold a ruler horizontally on the outstretched index fingers of both hands as in Figure 8.2. Slowly bring your fingers in from the ends of the ruler and note where they end up. Does it matter where you start your fingers from? Why?

Figure 8.2
Where do your fingers meet?



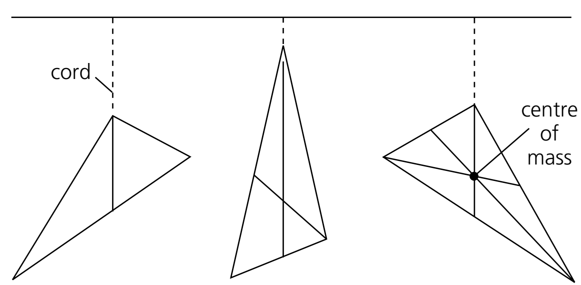
NOVEL CHALLENGE

Imagine you place a finger under each end of a ruler and a coin is placed on one end. You pull away both fingers and the ruler and coin fall together staying in contact. But if you just pull away the finger under the coin something odd happens. *What and why? Try it.*



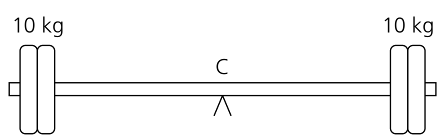
For regular shaped objects like a metre ruler, the centre of mass is at the midpoint. That's why you pick up a plank of wood in the middle. For irregular objects, though, the centre of mass can be found by letting the object hang from a pivot hole or point and drawing a vertical mark on the object. When this is done several times, the point at which the lines cross marks the centre of mass.

Figure 8.3
Locating the centre of mass.



For more regular rigid bodies, some simple principles can be used to find the centre of mass mathematically. Consider a set of weights on the ends of a steel bar (Figure 8.4).

Figure 8.4



Intuitively, the centre of mass is at the centre of the bar. In this case the products of each weight and its distance from the pivot point are equal. When the masses are unequal, obviously the centre of mass is closer to the heavier mass. Children use the ideas of centre of mass when operating a seesaw. The seesaw has a fixed pivot point. If two children of very different mass get on, the heavier child has to sit closer to the pivot point. This positions the centre of mass of the system of two children at the pivot point.

This suggests an inverse relationship between the weights or masses of the children and their distance from the centre of mass:

$$\frac{\boldsymbol{F}_{w1}}{\boldsymbol{F}_{w2}} = \frac{s_2}{s_1}$$

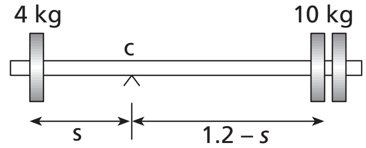
or, in general:

$$\boldsymbol{F}_{w1} \times s_1 = \boldsymbol{F}_{w2} \times s_2$$

Example

Masses of 4 kg and 10 kg are on the ends of a 1.2 m long bar as shown in Figure 8.5. Determine the centre of mass of the system.

Figure 8.5



Solution

Point C is located s metres from the 4 kg mass and (1.2 – s) metres from the 10 kg mass. For the bar to balance:

$$\boldsymbol{F}_1 \times s_1 = \boldsymbol{F}_2 \times s_2$$
$$40 \times s = 100(1.2 - s) \quad \text{using } \boldsymbol{g} = 10 \text{ m s}^{-2}$$
$$s = 0.86 \text{ m}$$

Later in this chapter, the product of $\boldsymbol{F} \times s$ in similar situations will be defined as **torque** and discussed in detail as it applies to rotating bodies.

Motion of the centre of mass

Knowledge of the motion and properties of the centre of mass gives some good insights into everyday phenomena.

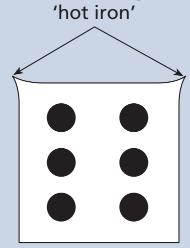
The grand jeté

When you do a long jump, chances are that your body will follow a parabolic path like a baseball thrown in from the outfield. But when a skilled ballet dancer does a split leap across the stage in a grand jeté, the path taken by her head and torso is nearly horizontal during much of the jump. She seems to be floating across the stage. The audience may not know Newton's laws of motion, but they can always sense that something magical has happened. The secret is that she raises her arms and legs as she jumps upward. These actions shift her centre of mass upward through her body. Although the shifting centre of mass faithfully follows a parabolic path across the stage, its movement relative to the body decreases the height that would be attained by her head in a normal jump. The result is that the head and torso follow a nearly horizontal path.

NOVEL CHALLENGE

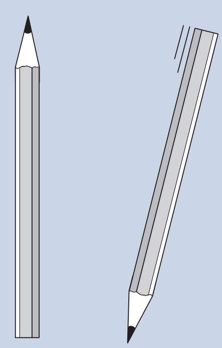
There are several types of 'crooked' dice used by cheats. For each one described, deduce why they are crooked:

1 Green's Load (1880) — two spots drilled out and mercury added.
2 Tapping dice — hollow centre filled with mercury but with a small tube to one corner. Tap to make them crooked.
3 Bevelled — rounded on some edges.
4 Slick — one surface highly polished.
5 Hot iron — a ridge along one edge.
6 Capped — one face capped with rubber.



NOVEL CHALLENGE

It is easy to stand a pencil up on its base but impossible to stand it up on its point. *But why? What if you could put it in a sealed container free of air currents and arranged it so that its centre of mass was exactly over the point — could you do it then? Still no! But what is the physics behind the failure?*



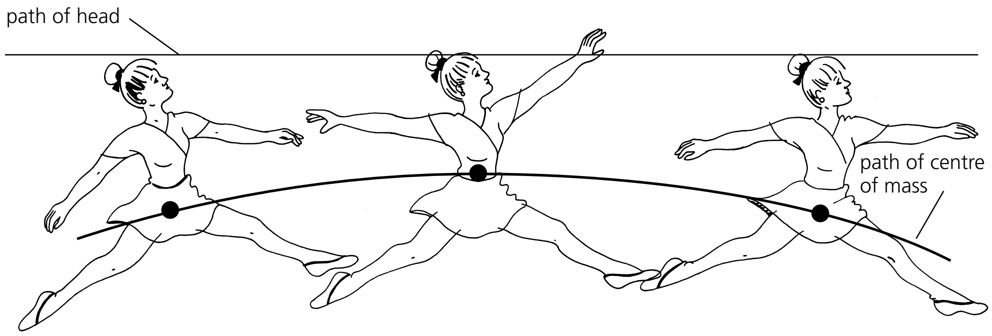


Figure 8.6
A grand jeté.

The Fosbury flop

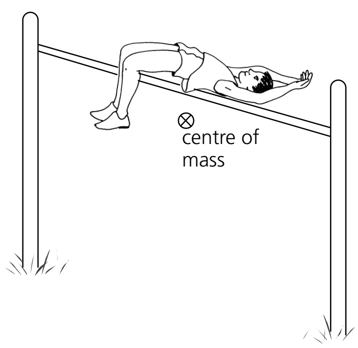


Figure 8.7
In a Fosbury flop, the centre of mass may actually pass under the bar.

The most successful high jumpers are tall, long-legged athletes because their centre of mass is further off the ground and so does not have to be lifted as far as that of a shorter jumper when clearing the bar. The high jumper must try to adopt a body position at take-off that keeps the centre of mass as high as possible. The momentum acquired during the run-up is modified in the last two steps before take-off. The jumper sinks down on the second last step and then comes erect on the take-off step so that the body has an initial upward velocity. The time that the jumper's foot is in contact with the ground on this last step is called the take-off time and is of the order of 0.12 to 0.17 s. The jumper has to also rotate so that the body is horizontal as it goes over the bar. This is the Fosbury flop, a technique popularised by American high jumper Dick Fosbury who developed the style and used it to win the 1968 Olympic gold medal. The technique has the advantage that the centre of mass passes under the bar even though the jumper curls over the top. For example, if the high jump bar was set at 2.07 m it is estimated that the jumper's centre of mass was only lifted from 1.27 m at take-off to 1.95 m when clearing the bar. Jumpers using the older scissor jump or western-roll style would have had to jump at least an extra 12 cm to clear the bar.

Questions

1 A 2.5 kg mass and a 4 kg mass are placed 1.5 m apart. Where is their centre of mass?

2 Where is the centre of mass of the Earth (mass 6×10^{24} kg) and the Moon (mass 7.4×10^{22} kg) when they are 3.8×10^{8} m apart?

MOMENTUM 8.3

'Momentum' is another term like 'velocity' that gets used in newspapers and magazines in strange ways. Journalists write that 'protests against whaling are gaining momentum' or about a truckies' blockade having a momentum of its own. However, what physicists mean by momentum is to do with mass and velocity.

The product of mass and velocity is called **momentum** (Latin *momentum* = 'movement'). It is a useful quantity to describe the motion of an object.

$$\text{Momentum} = \text{mass} \times \text{velocity}$$
$$\boldsymbol{p} = m\boldsymbol{v}$$

Momentum is a vector quantity, being the product of a scalar (mass) and a vector (velocity). The unit of momentum does not have a special name. The unit is kg m s^{-1} and is the same as N s, and this is often used. **The direction of momentum is the same as the direction of velocity.**

Example 1
Calculate the momentum of a 2 kg bowling ball travelling at 8 m s^{-1} south.

Solution

$$\boldsymbol{p} = m\boldsymbol{v} = 2 \times 8 = 16 \text{ kg m s}^{-1} \text{ south}$$

Example 2
A proton of mass 1.67×10^{-27} kg is accelerated from 3×10^4 m s^{-1} north to 3×10^5 m s^{-1} north. Calculate the **change** in momentum.

Solution
- Change in momentum = final momentum − initial momentum.

$$\begin{aligned}\Delta\boldsymbol{p} &= \boldsymbol{p}_\text{f} - \boldsymbol{p}_\text{i}\\ &= m\boldsymbol{v} - m\boldsymbol{u}\end{aligned}$$

- Final momentum = $m\boldsymbol{v} = 1.67 \times 10^{-27} \times 3 \times 10^5 = 5.0 \times 10^{-22}$ kg m s^{-1} north.
- Initial momentum = $m\boldsymbol{u} = 1.67 \times 10^{-27} \times 3 \times 10^4 = 5.0 \times 10^{-23}$ kg m s^{-1} north.
- Change in momentum = 5.0×10^{-22} kg m s^{-1} − 5.0×10^{-23} kg m s^{-1}
= 4.5×10^{-22} kg m s^{-1} north.

Questions

3 Calculate the magnitude of the momentum of the following moving objects:
(a) a 1000 g bowling ball moving at 1.6 m s^{-1}; **(b)** a 2.0 t car moving at 15 m s^{-1};
(c) the Earth in its journey around the Sun. The Earth's mass is 6×10^{24} kg and its radius of orbit is 1.5×10^{11} m.

4 A cricket ball of mass 200 g travelling at 20 m s^{-1} east is struck directly back at 30 m s^{-1} west. Calculate the change in momentum. *Hint:* the change in velocity is *not* 10 m s^{-1}. Can you see why?

8.4 ACTING ON IMPULSE

In 1687 Isaac Newton wrote that the force on an object determined the 'rate of change of the quantity of motion'. He expressed this in his second law of motion, which can be written as $\boldsymbol{F} = m\boldsymbol{a}$. But it can also be expressed in terms of momentum.

The acceleration can be replaced by $\frac{\boldsymbol{v} - \boldsymbol{u}}{t}$ to give:

$$\boldsymbol{F} = \frac{m(\boldsymbol{v} - \boldsymbol{u})}{t} = \frac{m\boldsymbol{v} - m\boldsymbol{u}}{t} = \frac{\text{change in momentum}}{\text{time}}$$

Hence the rate of change of momentum is equal to the external force causing the change. This can be rearranged as:

$$\boldsymbol{F}t = m\boldsymbol{v} - m\boldsymbol{u} \quad \text{or} \quad \boldsymbol{F}t = \Delta\boldsymbol{p}$$

The product $\boldsymbol{F}t$ is called the **impulse** (Latin *pulsus* = 'to beat' or 'drive'). Impulse depends on the size of the force and for how long it is applied. It is also equal to the **change** in momentum. The unit for impulse is Newton second (N s), which is the same as kg m s^{-1}. Some books use the symbol I for impulse but we will leave it as $\boldsymbol{F}t$.

NOVEL CHALLENGE

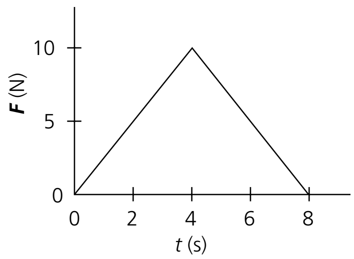
Can you jump off a chair onto the floor while holding a cup full of water without spilling any? Plan how you should land to do this. Hmmm! It sounds good in theory but …

NOVEL CHALLENGE

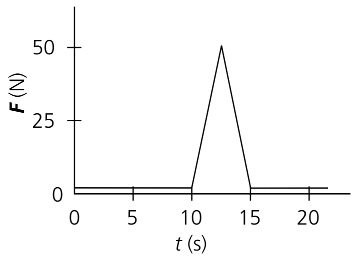
Imagine you are standing on some bathroom scales and you bend your knees quickly. *Predict what will happen to the scale reading. But why?*

Ch08-Walding 4th 25/8/04 11:24 AM Page 189

Figure 8.8
Impulse is the area under an F/t graph.



$$\boldsymbol{F}t = \frac{10 \times 8}{2} = 40 \text{ Ns}$$



$$\boldsymbol{F}t = \frac{50 \times 5}{2} = 125 \text{ Ns}$$

Force–time graphs

Impulse is also equal to the area under the graph (see Figure 8.8).

Most impacts involve forces that do not remain constant.

Example

A graph showing how force varies with time as a stationary 57 g ball is struck by a racquet is shown in Figure 8.9.
Calculate **(a)** the impulse; **(b)** the final velocity of the ball.

Solution

(a) The area under the graph is a measure of $\boldsymbol{F} \times t$, that is, impulse. In this case the impulse is approximated by the dotted triangle:

$$A = \frac{b \times h}{2} = \frac{3.2 \times 10^{-3} \times 2.5 \times 10^{3}}{2} = 4.0 \text{ N s (south)}$$

(b)
$$\boldsymbol{F}t = m(\boldsymbol{v} - \boldsymbol{u})$$
$$4 = 0.057(\boldsymbol{v} - 0)$$
$$\boldsymbol{v} = 70 \text{ m s}^{-1} \text{ south.}$$

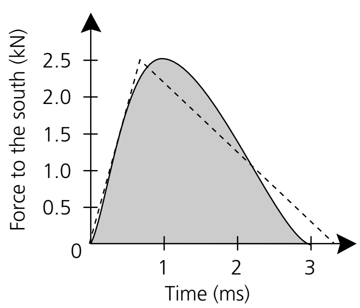
Questions

5 For how long must a frictional force of 5.6 N act in order to bring to rest a mass of 2.4 kg moving at 6.0 m s^{-1} north?

6 A car of mass 1200 kg accelerates at 5.5 m s^{-2} for 12 s. Determine the impulse imparted to the car.

CONSERVATION OF LINEAR MOMENTUM 8.5

Figure 8.9
The force exerted on a tennis ball during a serve can be represented graphically.

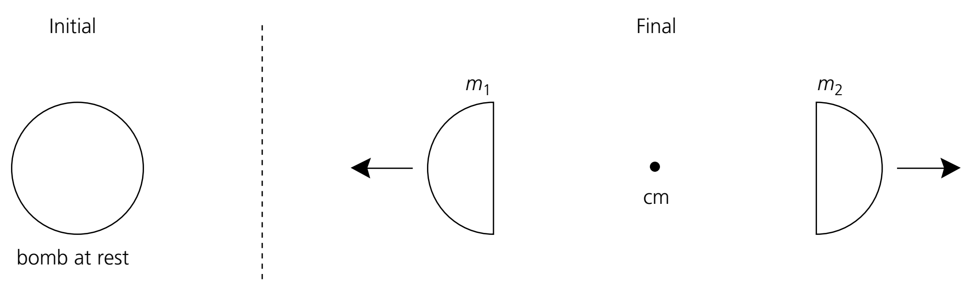


When you throw a ball, shoot a bullet or give someone a push you tend to move backward. Newton's third law of motion explained that action and reaction were equal and opposite forces. A study of momentum can describe the motion of interacting bodies mathematically. The two most common interactions we can study are **explosions** and **collisions**. We'll start with explosions because they are a bit simpler.

Explosions

An explosion can be thought of as a single object separating into two or more fragments. The word 'explode' was first used to mean 'burst with destructive force' in the nineteenth century when a mathematical treatment of explosions became necessary. Prior to that, the Latin verb *explodere* meant 'to drive off the theatre stage with hisses, boos, loud noises and claps'. It came from *ex-* meaning 'out' and *plaudere* meaning 'clap'. Many scientific words started off meaning something else.

Figure 8.10



Consider a 10 kg bomb at rest that explodes into two fragments (Figure 8.10). If a 4 kg piece (m_1) travels west at 15 m s^{-1} ($\boldsymbol{v}_1$), then the 6 kg piece (m_2) would have moved in the opposite direction (at a speed $\boldsymbol{v}_2$). As there was no external unbalanced forces acting on the bomb (all forces were internal), we have a closed system and there would be no change in the total momentum of the system. This is called the **law of conservation of momentum. In a closed system, the change in momentum is zero**. That is,

$$\begin{aligned} \Delta \boldsymbol{p} &= 0 \text{ or } \boldsymbol{p}_{\text{initial}} = \boldsymbol{p}_{\text{final}} \\ (m_1 + m_2)\boldsymbol{u} &= m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2 \\ 10 \times 0 &= 4 \times 15 + 6 \times \boldsymbol{v}_2 \\ \boldsymbol{v}_2 &= -10 \text{ m s}^{-1} \text{ (the negative sign means east)} \end{aligned}$$

Example

A boy on rollerskates is travelling along at 8 m s^{-1}. He has a mass of 60 kg and is carrying his school bag of mass 10 kg. He throws the bag directly forward at 20 m s^{-1} relative to the ground. Calculate the boy's speed after the 'explosion'.

Solution

The boy and the bag have initial velocities in the positive direction. The final velocity of the bag is also positive.

$$\begin{aligned} \boldsymbol{p}_{\text{initial}} &= \boldsymbol{p}_{\text{final}} \\ (m_1 + m_2)\boldsymbol{u} &= m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2 \\ (60 + 10) \times 8 &= 60 \times \boldsymbol{v}_1 + 10 \times 20 \\ 560 &= 60 \times \boldsymbol{v}_1 + 200 \\ \boldsymbol{v}_1 &= 6 \text{ m s}^{-1} \end{aligned}$$

The positive direction means that the boy would continue to move forward.

Relationships such as this can be applied to all sorts of explosions — a cannon or rifle being fired, a bomb exploding, a heart pumping a pulse of blood, a hose squirting water and even a nucleus giving off radioactive particles.

Cases in which the bodies explode in a straight line are not that common, however. Explosions in two dimensions will be dealt with later.

Questions

7 Two children at rest push off from each other in a swimming pool. One with a mass of 50 kg moves east at 1.5 m s^{-1} and the other who has a mass of 45 kg moves to the west. What is the second child's velocity?

8 A girl of mass 50 kg is stationary on an ice rink. She throws a 1.0 kg parcel horizontally at 5.0 m s^{-1}. At what velocity does the girl move?

Collisions

In everyday language, a collision occurs when objects crash into each other. Although we will refine that definition, it conveys the meaning well enough. Some familiar collisions are:

- the Creek meteorite crater in Australia
- a billiard ball being struck by a cue
- a boxer punching a body bag
- hammering a nail into a piece of wood
- gas molecules bouncing off each other.

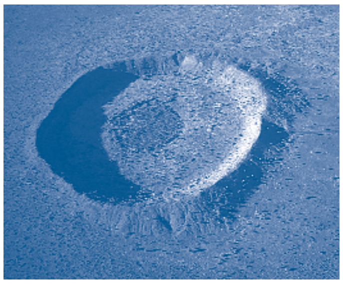
Car collisions

The safety of the passengers in a car during a collision also depends on the time interval in which the moving car is brought to a stop. To reduce the force of the impact, we have to

NOVEL CHALLENGE

A very lightweight boat 24 m long and mass of 30 kg lies still on a quiet pond. A 90 kg man walks from bow (front) to stern. *How far does the boat move relative to the pond? The answer is not 72 m.*

Photo 8.1
Meteorite crater.

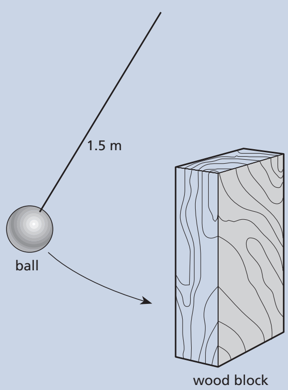


NOVEL CHALLENGE

Shooters who want to reduce the recoil of their rifles use a variety of anti-recoil devices. The simplest is to vent the exhaust gases out sideways instead of leaving them trapped in the barrel. One effective method involves drilling a hole in the rifle butt (the wooden shoulder piece) and inserting a rod of steel about 2 cm in diameter. Better still, inventive shooters use a length of steel water pipe $\frac{3}{4}$ filled with mercury and capped. So how does this help?

NOVEL CHALLENGE

A superball is tied to a 1.5 m string and suspended vertically from a hook. It is pulled back and allowed to strike a wooden block standing on the floor. The experiment is repeated with a lump of plasticine of the same mass as the ball. One knocks the block over, one doesn't. *Which is which and why?*



increase the time it takes. Manufacturers do this by making the front and rear of cars collapsible. These 'crumple zones' must be neither too hard nor too soft. They must progressively collapse so that the time of the collision is made as long as possible. Other safety features in a car are:

- air bags
- safety belts: inertia reel and self-tensioning
- antilocking brakes
- impact-absorbing bumper bars
- a collapsible steering column
- a rigid cabin compartment
- a soft dashboard instead of metal or wood.

Most of these are based on the principle that the longer it takes for your body to come to rest, the smaller the force your body has to stand. While the change in momentum is usually the same no matter how you crash, it is better to suffer a small force for a long time than a large force for a short time.

Does your family own a large four-wheel drive vehicle? It might be interesting to apply your knowledge of physics to the comparative safety of these vehicles in a collision, given the fact that they are very solid and rigid with only limited crumple zones.

Types of collisions

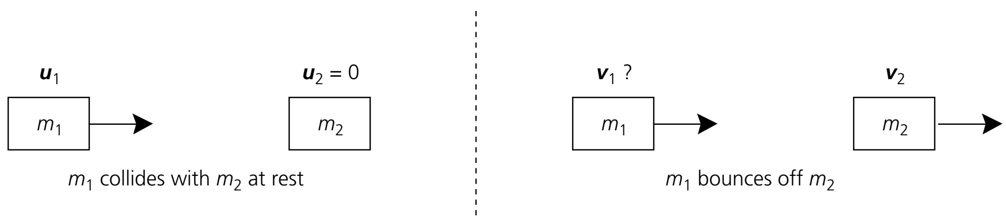
Collisions can be grouped into two types:

- Rebound, where objects bounce off each other (e.g. gas molecules or billiard balls).
- Coupled, where objects remain locked together (e.g. a bullet in a target).

Rebound

Consider a collision between two masses m_1 and m_2 with initial velocities $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ respectively.

Figure 8.11



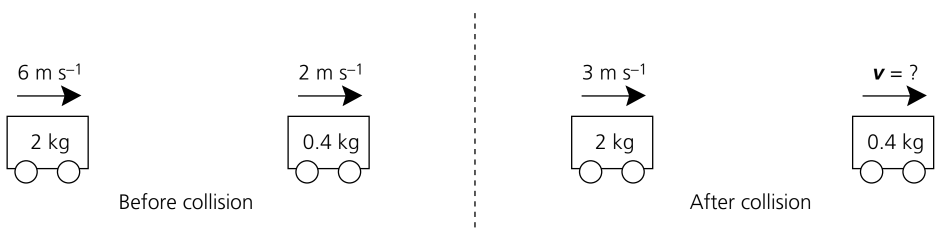
For the law of conservation of momentum to hold, the initial momentum must equal the final momentum:

$$m_1\boldsymbol{u}_1 + m_2\boldsymbol{u}_2 = m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2$$

Example

A cart with a mass of 2 kg travelling at 6 m s^{-1} collides with another cart of mass 0.4 kg travelling in the same direction at 2 m s^{-1}. It bounces off as shown in Figure 8.12. After impact, the 2 kg cart travels at 3 m s^{-1} in the same direction. Calculate the velocity of the 0.4 kg trolley after the collision.

Figure 8.12



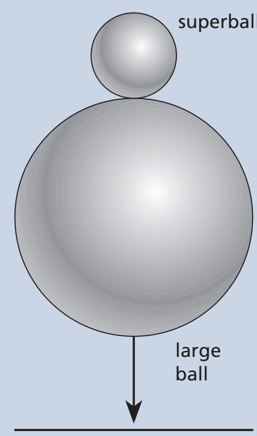
Solution

$$m_1\boldsymbol{u}_1 + m_2\boldsymbol{u}_2 = m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2$$
$$2 \times 6 + 0.4 \times 2 = 2 \times 3 + 0.4 \times \boldsymbol{v}_2$$
$$\boldsymbol{v}_2 = +17 \text{ m s}^{-1}$$
(the positive sign indicates that the trolley is moving in the same direction as before).

NOVEL CHALLENGE

A superball is placed on top of a tennis ball and they are dropped together.

Predict what happened — wow, what a funny rebound — and why?



Coupled or sticking together

When objects stick together or are joined together they are said to be coupled (Latin *copula* = 'to bond'). In a collision where the objects become coupled, the law of conservation of momentum still holds but the mass of the combined body after the collision is equal to the sum of the individual masses of the colliding bodies.

Some examples of coupled collisions are:

- an arrow sticking into its target
- two cars colliding head-on.

Example

A supermarket trolley loaded with shopping has a mass of 60 kg. It rolls across the floor at 4 m s^{-1} and collides with an empty trolley of mass 25 kg, which was stationary. They become fastened together and roll on as one. Calculate the velocity of the two trolleys when locked together.

Solution

$$m_1\boldsymbol{u}_1 + m_2\boldsymbol{u}_2 = m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2$$
$$m_1\boldsymbol{u}_1 + m_2\boldsymbol{u}_2 = (m_1 + m_2)\boldsymbol{v} \quad (\text{as } \boldsymbol{v}_1 = \boldsymbol{v}_2)$$
$$60 \times 4 + 25 \times 0 = 85 \times \boldsymbol{v}$$
$$\boldsymbol{v} = 2.8 \text{ m s}^{-1}$$

Practical use of coupled collisions

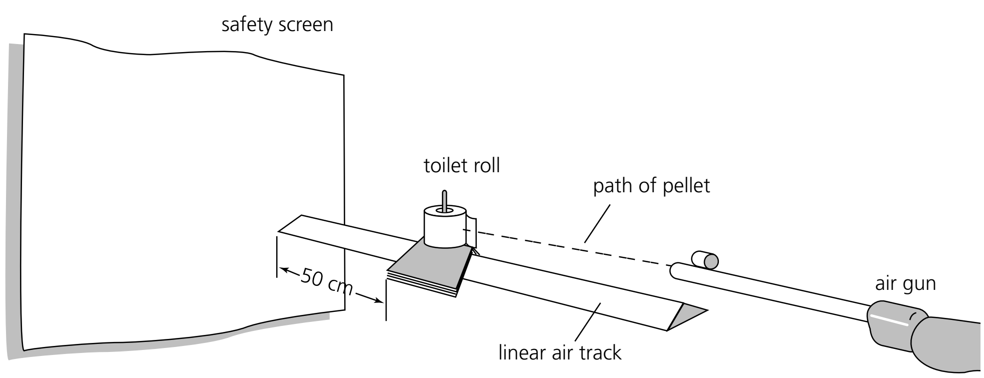


Figure 8.13
Measuring the velocity of an air-rifle pellet in the laboratory.

One way of measuring bullet speeds is to make use of a coupled collision. If an air-rifle pellet is shot into a soft absorbent target (such as a toilet roll) that is attached to a linear air-track glider, the glider moves away under the impact of the pellet. By measuring the time it takes the glider to move say 50 cm, the velocity of the glider can be determined. These data can be used to calculate the velocity of the pellet.

Example

When a 0.45 g air-rifle pellet is fired into a target attached to a glider on a linear air track, the glider moves 50 cm in 3.8 seconds. Calculate the velocity of the pellet. The glider and target have a combined mass of 643 g.

Ch08-Walding 4th 25/8/04 11:24 AM Page 193

NOVEL CHALLENGE

A fly crashes into the front windscreen of a train and reverses its direction. Therefore, at one instant its velocity is zero but as it is squashed onto the window, the window's velocity must also be zero for a short time. *How could a fly stop a speeding locomotive?*

Solution

- Mass of pellet = 0.45 g = 0.000 45 kg.
- Velocity of glider = $\frac{s}{t} = \frac{0.5}{3.8} = 0.13$ m s–1.

$$\boldsymbol{p}_\mathrm{i} = \boldsymbol{p}_\mathrm{f}$$

$$m_\mathrm{pellet} \times \boldsymbol{u}_\mathrm{pellet} = m_\mathrm{target} \times \boldsymbol{v}_\mathrm{target}$$

$$\boldsymbol{u}_\mathrm{pellet} = \frac{m_\mathrm{target} \times \boldsymbol{v}_\mathrm{target}}{m_\mathrm{pellet}} = \frac{0.643 \times 0.13}{0.000\ 45}$$

$$= 186 \text{ m s}^{-1}$$

Note: the final mass of the glider and target should include the mass of the embedded pellet but as it is negligible it can be ignored in this case. Of course, if the mass of the embedded object was large then it would have to be included.

In the next chapter a device called a ballistic pendulum, used for measuring the speed of high-speed bullets, will be described.

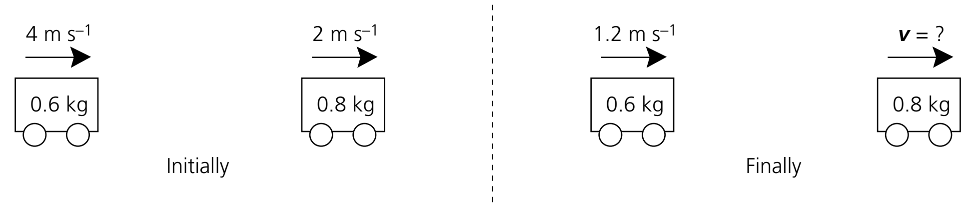
Questions

9 An object of mass 5 kg moving with a velocity of 10 m s^{-1} strikes another of mass 3 kg at rest. The two masses continue in motion together. Find their common velocity.

10 An archer fires an arrow of mass 96 g with a velocity of 120 m s^{-1} at a target of mass 1500 g hanging from a long piece of string from a tall tree. If the arrow becomes embedded in the target, with what velocity does the target move?

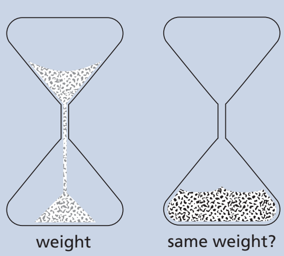
11 Two carts, one of mass 0.6 kg and the other of mass 0.8 kg, are moving north along a smooth horizontal surface with speeds of 4 m s^{-1} and 2 m s^{-1} respectively, as shown in Figure 8.14. After the collision, the 0.6 kg mass continues to travel north but with a speed of 1.2 m s^{-1}. What is the speed of the 0.8 kg mass?

Figure 8.14



NOVEL CHALLENGE

An egg timer is placed on an electronic balance and its mass recorded when all the sand is at the bottom. It is inverted. *As the sand falls, will it weigh the same or less?*



12 A 0.41 g air-rifle pellet is fired into a target made up of a 170 g toilet roll attached to a glider of 350 g. The target slides along a linear air track a distance of 50 cm in 2.8 s (refer back to Figure 8.13).

(a) Calculate the velocity of the pellet.

(b) What additional information would you need to calculate the recoil speed of the air rifle?

REAL-LIFE EXAMPLES OF CONSERVATION OF MOMENTUM 8.6

Examples of the concept and uses of the law of conservation of momentum abound in everyday life but often they need pointing out to become obvious.

Sports

Both collisions and explosions are features of many sports.

- Explosions: firing an arrow, throwing a ball or jumping into the air.
- Collisions: hitting a ball, karate chopping a brick or punching a bag.

Sporting explosions

In many sports the athlete is striving to deliver the maximum momentum to the ball or other projectile such as a discus or javelin. Sports physicists use special platform balances that measure the force being exerted on the ground. Figure 8.15 shows the force–time graph of a shotputter. When the force is in the direction of the ball it is called positive; when it is in the opposite direction it is called negative. Obviously it is the negative force that provides the propulsion force.

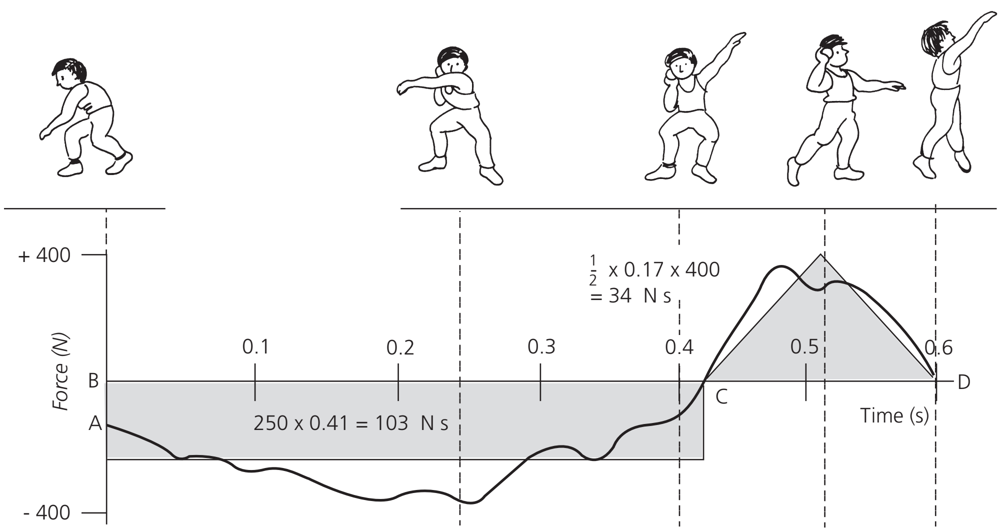


Figure 8.15
Changes in force during a shotput throw. Note how the downward (negative) force changes to an upward force as the ball leaves the hand.

The total impulse is the total area under the curve. The first 0.4 seconds have a negative impulse of −103 N s, whereas the final 0.2 second period is +34 N s. The total impulse is −69 N s.

$$\text{Total impulse } (= \Delta \boldsymbol{p}) = -69 \text{ N s}$$
$$\Delta \boldsymbol{p} = m\Delta \boldsymbol{v}$$

For a 4 kg shot the velocity would be 17.3 m s^{-1}.

Sporting collisions

In some sports, the player has a racquet or bat to strike the ball and momentum is transferred to the ball. In badminton and squash, players flick their wrists to increase the momentum of the light head of the racquet by making it move very fast. In tennis, where the ball's mass is much greater than the shuttle in badminton, this technique is not effective. Players must keep their wrists stiff as they swing at the ball so that the racquet acts as an extension of their body and the effective striking mass is that of the racquet, arm and shoulder.

Table 8.1 shows typical velocities of balls hit in various sports.

Table 8.1 TYPICAL VELOCITIES OF BALLS HIT FROM REST IN A VARIETY OF SPORTS*

BALL	BALLS MASS (kg)	BALL'S VELOCITY ($m\ s^{-1}$) BEFORE	BALL'S VELOCITY ($m\ s^{-1}$) AFTER	STRIKER'S VELOCITY ($m\ s^{-1}$) BEFORE	STRIKER'S VELOCITY ($m\ s^{-1}$) AFTER	IMPACT TIME (s)
Baseball	0.15	0	39	31	27	1.35×10^{-3}
Football (punt)	0.42	0	28	18	12	8.0×10^{-3}
Golf ball (drive)	0.047	0	69	51	35	1.25×10^{-3}
Squash ball (serve)	0.032	0	49	44	34	3.0×10^{-3}
Tennis ball (serve)	0.057	0	51	38	33	4.0×10^{-3}

* Also shown is the impact time of the strike and the velocity of the striking mass before and after impact.

In each case the effective striking mass can be calculated and you can see how much of the player's body is added to the mass of the racquet.

Example

The actual mass of the tennis racquet used in Table 8.1 was 0.40 kg. Calculate its effective mass.

Solution

$$\boldsymbol{p}_{\text{initial}} = \boldsymbol{p}_{\text{final}}$$

$$m_{\text{ball}} \times \boldsymbol{u}_{\text{ball}} + m_{\text{racquet}} \times \boldsymbol{u}_{\text{racquet}} = m_{\text{ball}} \times \boldsymbol{v}_{\text{ball}} + m_{\text{racquet}} \times \boldsymbol{v}_{\text{racquet}}$$

$$0.057 \times 0 + m_{\text{racquet}} \times 38 = 0.057 \times 51 + m_{\text{racquet}} \times 33$$

$$5 \times m_{\text{racquet}} = 2.91$$

$$m_{\text{racquet}} = 0.58 \text{ kg (the racquet's mass was increased from 0.4 kg to 0.58 kg)}$$

Baseball

The 1993 World Championship was determined by an otherwise perfect swing of the bat but it was just 1 mm too high and a less than perfect shot resulted in a run out. There's not much room for error.

What baseball players are looking for is both high bat speed and good control. The problem is, the higher the bat speed the less control the player has over the accuracy of the hit. In major league games, the ball strikes the bat after coming from the pitcher's mound 17 m away in 0.45 s. It collides with a bat just 7 cm wide, which is being swung at 100 km/h so there's not much time for decision-making. The ball is squashed to half its diameter and leaves the bat after 0.001 s contact time. You'd wonder how anyone could have control over the placement of the ball. But they do, although the difference between a foul and a hit over second base is only 0.01 s in timing. What a game!

The role of momentum in a good hit is crucial. Players want to give their bats high momentum and they can do this by increasing the mass of the bat or by swinging it faster. Legendary baseball champion 'Babe' Ruth used heavy bats, often as heavy as 52 ounces (1.5 kg). Today's players use bats of about 850 g, but Ruth had exceptional strength and could whip his bat around at high speed. However, changing from a bat of six times a ball's mass to one of seven times its mass adds little to the transfer of momentum to the ball. What it does is slow down the swing considerably.

But how can bat speed be increased? Watch the lead-off hitters, the small players who must get on base so the power hitters can drive them in. Lead-off hitters need to be able to punch the ball to the opposite field or find a hole in the in-field. They need excellent control and good bat speed. They 'choke-up' on the bat — sliding their hands up higher on the handle, making it easier and faster to swing.

But a short bat is not long enough to reach those fast balls on the outside of the plate. A better solution is to use a lighter bat and, over the past few decades, bat masses have decreased to the current 800–900 g. But as the wood is thinner, the risk of breaking is also increased, so aluminium and composite plastics (graphite, fibre glass) are used in most games except major league, where aluminium is too fast and dangerous.

Squash gets fast too but because the ball heats up during the game. A hot ball has a greater change in momentum than a cold one. The physics of that is interesting to contemplate.

Questions

13 **(a)** Calculate the force imparted to a 145 g ball during a hit as described above. Assume the rebound speed of the ball from the bat is the same as the impact speed of the ball.

(b) Assuming a player can supply the same momentum to bats of different mass, calculate the speed of an 850 g bat if he can swing a 1500 g bat at 30 m s^{-1}.

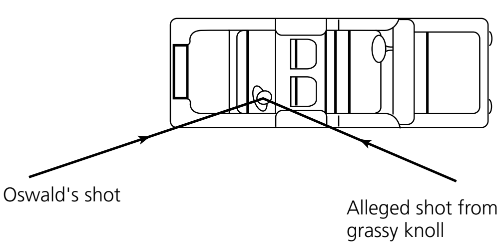
Forensic science

In the course of police investigations into crimes, physicists often play a vital role. Much of the scientific evidence in a forensic investigation (Latin *forum* = 'of the court') is biological or chemical in nature but when car accidents or guns are involved, the physicists who are experts in kinematics or ballistics are called in.

One famous case concerns the assassination of US President John F. Kennedy in 1963. JFK was shot in the head and neck by high powered rifle bullets. Movie film of the event shows that his head tilted forward as he was struck in the back of the neck and then his head moved rapidly back as he suffered a head wound. Assassination buffs have split into two groups, depending on whether they believe in a single 'lone nut' gunman or a conspiracy between two gunmen. The 'lone nutters' believe all wounds were caused by a lone gunman (Lee Harvey Oswald) firing from the sixth floor of a building behind the President's car. The conspiracy theorists believe that Oswald was responsible for the neck wound but another gunman firing from the grassy knoll to the front right of the car was responsible for the fatal head wound. No video of the assassination exists, the only clear film of the events being made by Abraham Zapruder on a hand-held Super-8 movie camera from a distance of about 60 m.

People who claim that JFK was shot from the front say that, because his head moved backward, a second gunman fired from the front (from the 'grassy knoll') (see Figure 8.16). Nobel-Prize-winning physicist Luis Alvarez contradicts this. He has shown that when an object such as a taped-up watermelon (simulating a head) is shot, the melon generally moves towards the gun as chunks are blown out the other side. Dubbed the 'jet effect', Dr Alvarez showed that the matter blown out of the melon carried with it more momentum than was brought in by the bullet. This is similar to the motion of a rocket as jet fuel is ejected and is a good example of conservation of momentum. Other physicists further argued that it was the shot by Lee Harvey Oswald to the back of the neck that caused JFK's arms to fly up under his chin and his body to jerk backward in a nervous reaction known as the 'Thorburn Position'. Either way, physicists agreed that the head shot came from behind JFK and have dismissed the conspiracy theory. Oswald used a $12.50 Italian Carcarno hunting rifle that fired high velocity (670 m s^{-1}) full-metal-jacket bullets, each with a mass of 10.37 g. By the time they reached the President, the bullets had lost momentum and were travelling at 545 m s^{-1}. Imagine such a bullet striking a melon and remaining embedded in it while simultaneously blowing a jet of melon out the other side. The equation becomes:

Figure 8.16
The Presidential limousine: the Kennedys in the rear and the Connallys in the centre.



momentum of bullet = momentum of remains of melon and bullet + momentum of jet

$$m_b v_b = m_r v_r + m_j v_j$$

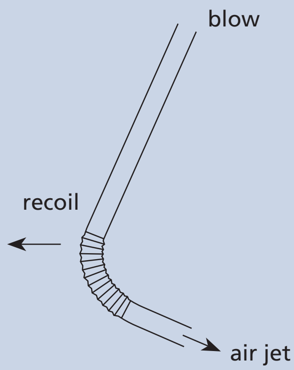
If the momentum of the bullet ($0.010\,37 \times 545 = 5.65$ kg m s^{-1}) was less than the momentum of the jet, then the momentum of the remains of the melon with the embedded bullet would have to be negative, and hence the velocity would also have to be negative. This means the remains of the melon would have moved toward the gunman. The 'lone nutter' theory is supported.

Questions

14 A 10.0 g Carcarno bullet is fired with a muzzle velocity of 545 m s^{-1} at a 3.0 kg watermelon and remains embedded in it.
(a) Calculate the motion of the melon if there is no jet exiting the other side.
(b) Calculate the motion of the melon if a 300 g jet of melon exits the rear of the melon at a speed of 35 m s^{-1}.

NOVEL CHALLENGE

When you blow through a bent drinking straw it recoils away from the jet of air. *But when you suck air in does the reverse happen? I think not! Explain that one in terms of momentum. Try it and see.*



SR Activity 8.2 KENNEDY ON THE INTERNET

1. If you have access to the Internet, try reading the newsgroup alt.assassination.jfk, which deals with the Kennedy investigation. You'll meet lots of cranks but also some physicists who will discuss ballistics and forensic science.
2. If you can't get to sleep, try one of the JFK Web pages. There are dozens of them — some favour the conspiracy theory, the others favour the lone gunman theory. Try www.jfklancer.com for a start. Failing that, try joining a chat room. The chat times are listed on the Web pages. You never know, someone might try to sell you a Carcarno.

SR Activity 8.3 HIRE A VIDEO

Oliver Stone's movie *JFK* starring Kevin Costner is out on video. Stone takes a different line from the one above but still examines the evidence in a scientific way. If you can hire it, look for the discussion on the ballistics evidence. Make notes and compare it to the discussion opposite.

Propulsion of rockets

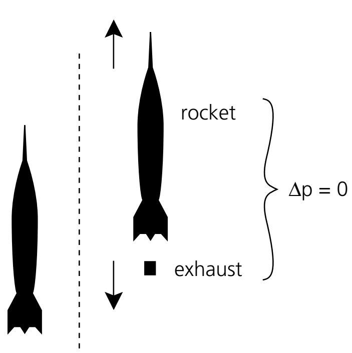
A rocket moves forward because burning gases are ejected at high speed behind it. If an engine supplies a constant force (thrust), the acceleration of the rocket will increase because the total mass of the rocket decreases as fuel and oxygen are burnt. Have you noticed how much faster an inflated balloon goes at the end of its journey than when you first let it go?

Many people think that rockets only work if they have something to push against. But they work in space where there is no air. The momentum of the exhaust gases is equal in magnitude but opposite in direction to the gain in momentum of the rocket.

Example

The German V-2 rockets used to bomb London in the Second World War had a mass of 12 000 kg and produced thrust from exhaust gases that were ejected at the rate of 1500 kg every second and at a speed of 170 m s^{-1}. Calculate **(a)** the initial forward force on the rocket (thrust); **(b)** the net force if the rocket was fired vertically; **(c)** the initial acceleration.

Figure 8.17



Solution

Consider the rocket to be made up of two exploding parts: the rocket itself and the exhaust gases (Figure 8.17). Initially both components are at rest and the momentum of each is zero. When fired, the momentum of each is equal and opposite.

(a)
$$\boldsymbol{p}_{\text{rocket}} = \boldsymbol{p}_{\text{exhaust}}$$
$$\boldsymbol{p}_{\text{rocket}} = m_e \times \boldsymbol{v}_e$$
$$= 1500 \times 170 = 255\,000 \text{ kg m s}^{-1}$$

The interaction time for this 'explosion' is 1 s, so the impulse (***F**t*) equals this change in momentum from zero to 10 000 kg m s^{-1}.

$$F = \frac{255\,000}{1 \text{ second}} = 255\,000 \text{ N (or kg m s}^{-2}\text{)}$$

(b) The net force is the result of the thrust (upward) and the weight (downward):

$$F_{net} = F_{upthrust} - F_w = 255\,000 - 120\,000 = 135\,000 \text{ N (upward)}$$

(c) Initial acceleration: $F_{net} = ma$, hence $a = F/m = 135\,000/12\,000 = 11.25$ m s^{-2}.

As these rockets burnt fuel their mass decreased, and hence their acceleration increased. After 65 s all fuel had been burnt and the rockets were moving at 2 km s^{-1}.

Such rockets were particularly dangerous because there was no warning sound. The whine of the rocket engines came *after* the sound of the explosion on landing because they travelled faster than the speed of sound.

8.7 CONSERVATION OF MOMENTUM IN TWO DIMENSIONS

Rarely is the world as simple as portrayed in the previous discussion. We live in a three-dimensional world and interactions occur in three dimensions. You have been introduced to momentum in one dimension so that the principles can be seen. Now it is time to venture into the two-dimensional world — *Flatland*. Interactions occurring in the 3-D world are beyond the scope of this book. Wait until first-year university physics for that.

It doesn't matter whether it is in one, two or three dimensions — momentum is always conserved. That is, total momentum before the collision equals total momentum after the collision.

Explosions in two dimensions

An object at rest has zero momentum. If it explodes into several pieces, the pieces will still have a zero total momentum. If it is moving when it explodes or separates, then the fragments will have a total momentum equal to the momentum before the explosion.

Example

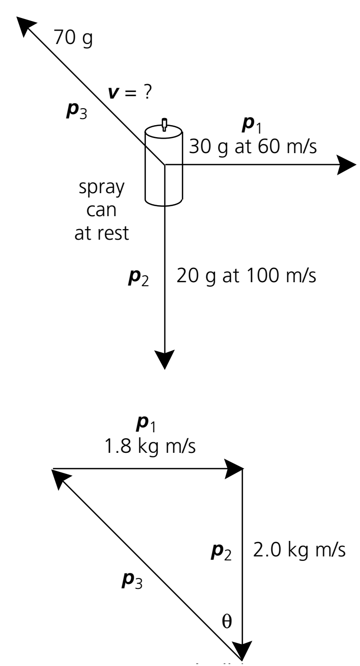
An empty spray can of mass 120 g rests on top of a fire and explodes into three fragments. One 30 g fragment travels east at 60 m s^{-1} and another 20 g fragment goes south at 100 m s^{-1}. (See Figure 8.18.) Calculate the velocity of the third piece.

Solution

The law of conservation of momentum states that the final momentum will be equal to the initial momentum, which in this case is zero. Hence, the sum of the three momentum vectors after the explosion will also be zero. That is, the three vectors will form a closed triangle when added head to tail. All we need do is draw the two known vectors and fill in the remaining gap (p_3) to see the missing vector.

Using Pythagoras' theorem, p_3 equals 2.7 kg m s^{-1}. As the mass is 70 g, the velocity must be 38.4 m s^{-1} in a direction θ of N42°W. This angle is sometimes expressed as '318° True'. Can you see why?

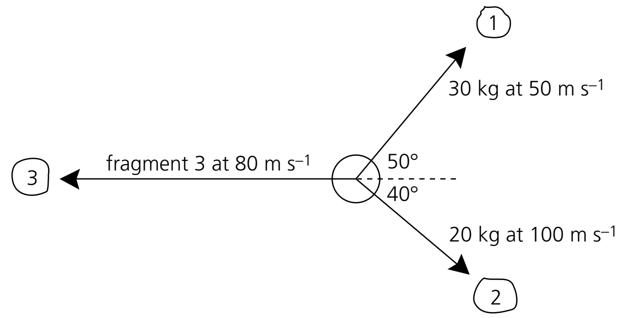
Figure 8.18
Vector diagram showing the momentum of each of the three fragments of the exploding spray can in the example to the left. When placed head-to-tail, the vector arrows p_1, p_2 and p_3 must add to zero.



Questions

15 A bomb, initially at rest, explodes into three fragments as shown in Figure 8.19. Calculate the mass of the third fragment.

Figure 8.19
For question 15.



16 A radioactive nucleus of mass 5×10^{-26} kg is at rest and emits two neutrons, each of mass 1.6×10^{-27} kg, at right angles to each other. If both have speeds of 360 m s^{-1}, calculate the recoil speed of the nucleus.

Collisions in two dimensions

As with collisions in one dimension, collisions in two dimensions can be of the rebound type or the objects can stay coupled together. For example:

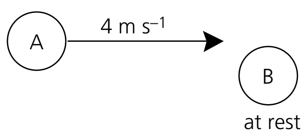
- rebound: billiard balls or cars colliding at an angle
- coupled: cars colliding off-centre and becoming tangled.

Rebound collisions

In a two-dimensional collision the objects approach and rebound obliquely. This means that their paths follow different lines but in the same plane. If you follow these four steps you have a good way of solving problems:

1 Construct a vector diagram showing the total momentum before the collision.
2 Construct another vector diagram showing the total momentum after the collision.
3 Equate these two vectors since $\boldsymbol{p}_f = \boldsymbol{p}_i$.
4 Calculate the unknown quantity by vector analysis.

Figure 8.20
Initial motion of ball A in Example 1.



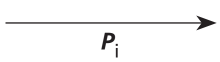
Example 1

A ball A of mass 1.0 kg is moving east at 4 m s^{-1} when it collides with a stationary ball B of mass 2 kg (Figure 8.20). Ball A heads north at 4 m s^{-1} after the collision. Determine the final velocity of ball B.

Solution

Step 1 Initial momentum

Figure 8.21
Total initial momentum of balls A and B in Example 1.

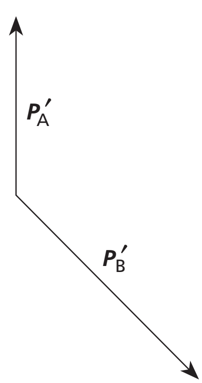


$$\boldsymbol{p}_i = \boldsymbol{p}_A + \boldsymbol{p}_B = 4 \text{ kg m s}^{-1} \text{ E} + 0 \text{ kg m s}^{-1} = 4 \text{ kg m s}^{-1} \text{ E (Figure 8.21)}$$

Step 2 Final momentum

Add $\boldsymbol{p}'_A + \boldsymbol{p}'_B$

Figure 8.22
The final momentum of the balls.



Note: the symbol $\boldsymbol{p}'$ (pronounced p-prime) is used to indicate the *final* momentum.

$$\boldsymbol{p}_f = \boldsymbol{p}'_A + \boldsymbol{p}'_B = m_A\boldsymbol{v}_A + m_B\boldsymbol{v}_B = 1 \text{ kg} \times 4 \text{ m s}^{-1} + 2 \text{ kg} \times \boldsymbol{v}_B \text{ (vectorily)}$$

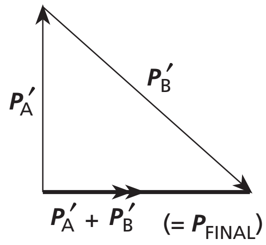


Figure 8.23
The final momentum of ball A and ball B is found by placing the momentum vectors head to tail.

Step 3 Equate $\boldsymbol{p}_f$ and $\boldsymbol{p}_i$

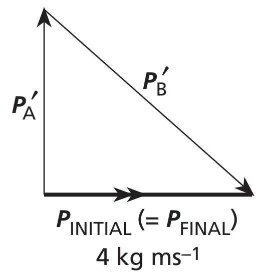


Figure 8.24
The sum of the final momentum of ball A ($\boldsymbol{p}'_A$) and of ball B ($\boldsymbol{p}'_B$) has to be equal to their intial momentum.

Step 4 Vector analysis
Using Pythagoras' theorem: $\boldsymbol{p}'_B = 5.6$ kg m s^{-1}.
As the mass of B is 2 kg, $\boldsymbol{v}_B = 2.8$ m s^{-1}.
Angle θ = S45°E (or SE or 135° True).

Example 2

A 2 kg ball (A) is travelling at 10 m s^{-1} east when it strikes a stationary 2 kg ball (B) a glancing blow. The two balls move away at right angles to each other with ball A travelling 30° to the north of its original path. Calculate the velocity of balls A and B after the collision.

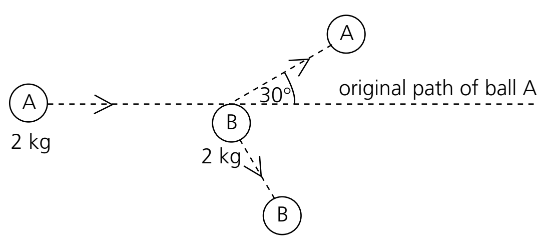


Figure 8.25
Motion of balls in Example 2.

Solution

$$\boldsymbol{p}_i = m_A\boldsymbol{u}_A = 2 \times 10 = 20 \text{ kg m s}^{-1}$$
$$\boldsymbol{p}_f = \boldsymbol{p}'_A + \boldsymbol{p}'_B = m_A\boldsymbol{v}_A + m_B\boldsymbol{v}_B = 2\boldsymbol{v}_A + 2\boldsymbol{v}_B$$

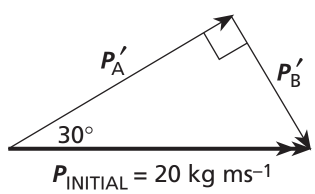


Figure 8.26
The final momentum equals the vector sum of the initial momentums.

- $\sin 30° = \boldsymbol{p}'_B/20$ $\boldsymbol{p}'_B = 10$ kg m s^{-1} $\boldsymbol{v}_B = 5$ m s^{-1}.
- $\cos 30° = \boldsymbol{p}'_A/20$ $\boldsymbol{p}'_A = 20 \times 0.866 = 17.3$ kg m s^{-1} $\boldsymbol{v}_A = 8.65$ m s^{-1}.

Note: as the above example has shown, when two objects of the same mass collide and bounce off, the angle between their paths after collision is a right angle. When the masses are different, the angle may not be a right angle and the solution to the problem is more difficult — the cosine rule is often used. If you try this experiment in class, you might find that the angle is just slightly *less* than 90° because of friction effects.

Coupled collisions

Example

Two skaters, Alfred (A) and Barbara (B), collide and hold each other together after impact. Alfred, whose mass is 83 kg, is originally moving east with a speed of 6.2 m s^{-1}. Barbara, whose mass is 55 kg, is originally moving north with a speed of 7.8 m s^{-1}. What is their velocity after the impact?

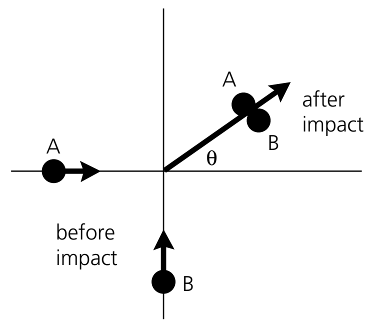


Figure 8.27
The two skaters A and B collide and move off together.

Solution

$$\boldsymbol{p}_A + \boldsymbol{p}_B = \boldsymbol{p}'_{(A+B)}$$
$$m_A\boldsymbol{v}_A + m_B\boldsymbol{v}_B = (m_A + m_B)\boldsymbol{v}$$
$$83 \times 6.2 + 55 \times 7.8 = 138\boldsymbol{v} \text{ (vector addition; do not solve for } \boldsymbol{v} \text{ using algebra)}$$

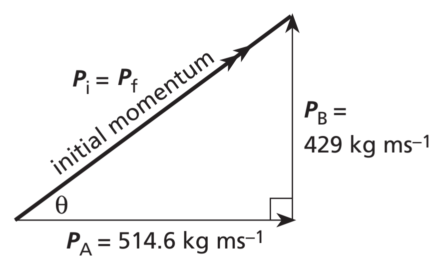


Figure 8.28
Final momentum (thick arrow) is the vector sum (head to tail) of the two initial momentums.

The initial momentum is shown by the hypotenuse. This is also the final momentum because of the law of conservation of momentum.

$$\boldsymbol{p}_{\text{final}} = 670 \text{ kg m s}^{-1}$$
$$\text{hence velocity is } \frac{670}{83 + 55} = 4.9 \text{ m s}^{-1} \text{ at angle } \theta = 40°$$

Questions

17 A stationary 6.0 kg bomb suddenly explodes and the fragments fly off in the directions shown in Figure 8.29. Determine the final velocity of the 1.0 kg piece.

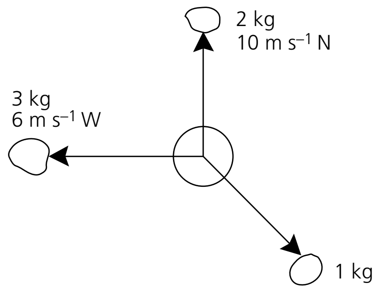


Figure 8.29
For question 17.

18 A ball A with a mass of 1.0 kg and moving at 4 m s^{-1} strikes a glancing blow on a second ball B of mass 2.0 kg that is initially at rest. After the collision, ball A is moving at right angles to its original direction at a speed of 3 m s^{-1} as illustrated in Figure 8.30. Calculate the velocity of B.

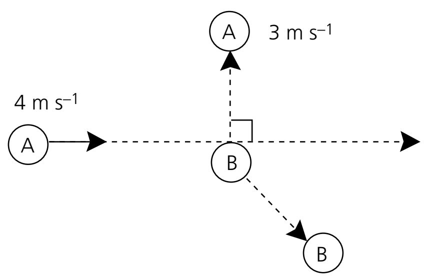
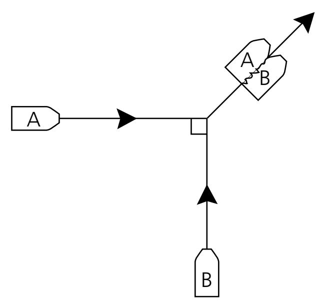


Figure 8.30
For question 18.

Figure 8.31
For question 19.



19 A barge with a mass of 1.5×10^5 kg is proceeding east along a river at 6.2 m s^{-1} in heavy fog when it collides with a barge heading north. The second barge has a mass of 2.78×10^5 kg and was moving at 4.3 m s^{-1}. Upon impact the barges remain coupled together (Figure 8.31). Find their resultant velocity.

8.8 TORQUE: MOVING IN CIRCLES

Not everything that has a force applied to it will move from one place to another. Sometimes it just goes in a circle. When you spin a bicycle wheel, the applied force goes into making the wheel rotate. When you tighten up the wheel nuts when changing a car tyre with a spanner, the force just makes the nuts turn around.

The turning effect is known as the **torque**, the symbol of which is the Greek letter 'tau' (τ). The magnitude of the torque depends on:

- the size of the force applied
- the direction of the force applied
- the distance between the point at which it acts and the pivot.

For example, when you open a door you are applying a torque to twist it about its hinges. You pull on the knob at right angles to the door. If you pulled sideways, the force would be wasted or reduced. As well, the door knob is located as far as possible from the pivot point to make your pull as effective as possible.

If $\boldsymbol{F}$ and r are perpendicular:

$$\text{Torque } \tau = \boldsymbol{F}r$$

If $\boldsymbol{F}$ and r are at an angle θ:

$$\text{Torque } \tau = \boldsymbol{F}r \sin \theta$$

NOVEL CHALLENGE

The torque to remove a lid off a jar of food has been set at 2 N m by food manufacturers. This torque is such that all 20- to 40-year-olds and 97% of 50- to 94-year-olds can remove it. Estimate the force necessary to manage this, assuming the radius of a typical lid is 3.5 cm. Why does it help to put a tight lid under a hot water tap?

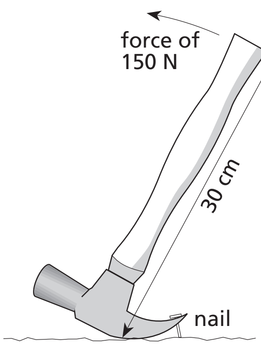


Figure 8.32
Calculating the torque applied to a hammer.

Example

A force of 150 N is applied at right angles to the end of a hammer handle 30.0 cm long to pull a nail from some wood. (See figure 8.32.) What torque is applied to the hammer?

Solution

Torque applied by the hand: $\tau = Fr = 150 \times 0.300 = 45$ N m.

Engine torque

If you hear people talking about four-wheel drives, the subject of torque eventually comes up. Diesel engines have a big reputation for providing a lot of torque at low engine speeds compared with their petrol-engined counterparts. Table 8.2 compares the engine performance of the turbo diesel and petrol Toyota Landcruiser 4WDs.

Table 8.2 ENGINE PERFORMANCES

	PETROL	DIESEL
Capacity (L)	4.5 L	4.2 L
Power (kW)	158 kW @ 4600 rpm	115 kW @ 3600 rpm
Torque (N m)	373 N m @ 3200 rpm	357 N m @ 1800 rpm

NOVEL CHALLENGE

When removing a cork out of a champagne bottle, it is easier if you hold the cork and rotate the base of the bottle rather than holding the base and rotating the cork as most people do. Why is this easier? It seems to defy logic, doesn't it?

Although the petrol engine produces more torque, the diesel engine produces it at a much lower engine speed. This gives it tremendous advantage in climbing sandhills and getting out of bogs. Also, the torque produced by either engine is not constant over the range of engine speeds but peaks at the value shown in the table.

NEI Activity 8.4 DIESEL VS PETROL

Motor enthusiasts seem to either love diesels or hate them.

1. If you know someone with an interest in cars and trucks see if you can get him or her to help you make a comparison of the advantages and disadvantages of diesel versus petrol engines. Use the following criteria as a guide: engine life (wear and tear), cost of engine, fuel price, fuel economy, acceleration ability, availability of fuel, water in cylinders, air pollution.
2. Why do farmers prefer diesels? Is it because of the low-speed torque?
3. What is the difference between a supercharger and a turbocharger?

Questions

20 Calculate the torque on a wheel nut produced by a force of 90 N at right angles on the end of a spanner 40 cm away from the pivot point (the wheel nut).

ANGULAR MOMENTUM 8.9

Now that you've seen how rotation occurs, it's time to look at the laws involved and some other things that rotate. Have you seen the spinning chair at the Sciencentre in Brisbane? Or the spinning chairs at Questacon or just about every other science expo around? They have something in common with a car engine, a springboard diver, a frisbee and the incredible shrinking stars.

Bodies that spin have momentum — **angular momentum**. It is different from linear momentum in that it is the spinning, not the movement from place to place, that is important. Just as you need a force to get a bicycle to move, you have to apply a torque to a bicycle wheel to make it spin. An external torque can change an object's angular momentum.

So far in this chapter you have seen that linear momentum is equal to the product of mass and linear velocity ($\boldsymbol{p} = m\boldsymbol{v}$). Similarly, angular momentum ($\boldsymbol{L}$) is the product of inertia ($\boldsymbol{I}$), which is related to mass, and angular velocity (ω). Hence: $\boldsymbol{L} = \boldsymbol{I}\omega$.

NOVEL CHALLENGE

The Earth gains 100 thousand tonnes each day as interstellar dust settles on this beautiful planet of ours. Calculate how much longer each day will be because of this. You may need to do this long-hand as most calculators won't show an answer unless you know some tricks.

Inertia

For rigid bodies such as a bicycle wheel or a rolling ball, physicists have developed formulas that enable us to calculate their **rotational inertia**. This is different from mass because for a rotating object not all the mass is travelling at the same speed — the outside goes faster than the inside. How the mass is distributed in that object will determine how difficult it is to start or stop the object rotating. Some simple objects are shown in Figure 8.33.

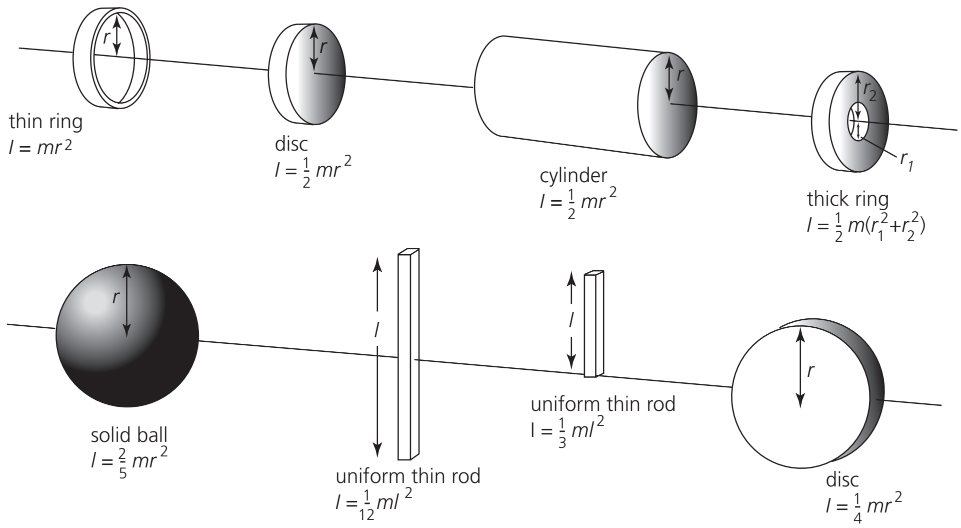


Figure 8.33
Rotational inertia equations for objects rotating about the indicated axes.

Example

Calculate the rotational inertia of a 20 kg snowball of diameter 1.5 m.

Solution

For a solid sphere:

$$\boldsymbol{I} = \tfrac{2}{5}mr^2$$
$$\boldsymbol{I} = \tfrac{2}{5} \times 20 \times 1.5^2$$
$$\boldsymbol{I} = 18 \text{ kg m}^2$$

A rolling snowball will therefore have both rotational momentum *and* linear momentum. To stop it moving you have to stop its translational motion *and* its rotational motion. Not an easy task.

NOVEL CHALLENGE

If everyone faced the same way on Earth and took a step at the same time would the Earth's rotation change? What data would you need to calculate this mathematically?

Angular momentum

The angular momentum $\boldsymbol{L}$ of a rigid body of rotational inertia $\boldsymbol{I}$ rotating at an angular speed ω about an axis is given by $\boldsymbol{L} = \boldsymbol{I}\omega$. The angular speed needs to be expressed in radians per second (rad s^{-1}) as was shown in Chapter 6.

Example

Calculate the angular momentum of the snowball in the previous example if it is rolling at 3 revolutions per second.

Solution

- 1 revolution equals 2π radians, hence 3 rev/s = 6π rad s^{-1}.

$$\begin{aligned} \boldsymbol{L} &= \boldsymbol{I}\omega \\ &= 18 \text{ kg m}^2 \times 6\pi \text{ rad s}^{-1} \\ &= 340 \text{ kg m}^2 \text{ s}^{-1} \end{aligned}$$

Ch08-Walding 4th 25/8/04 11:24 AM Page 205

CONSERVATION OF ANGULAR MOMENTUM 8.10

Just as linear momentum is conserved, so too is angular momentum.

If no net torque acts on a system, the angular momentum $\boldsymbol{L}$ of that system remains constant no matter what changes take place within that system.

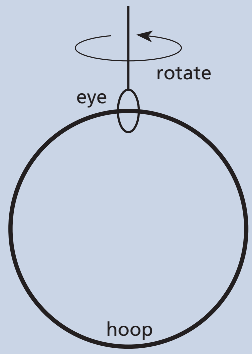
$$\boldsymbol{L} = \boldsymbol{I}\omega = \text{a constant}$$
$$\boldsymbol{I}_i\omega_i = \boldsymbol{I}_f\omega_f$$

In the simplest case, if a solid body is spinning at a particular rate then it will continue to spin at that rate unless an outside torque (twisting force) acts on it. More interestingly though, if the distribution of mass changes within the body, then inertia changes and so the angular speed will have to change to keep the angular momentum constant.

Imagine a student seated on a stool that can rotate freely. The student, who has been set into rotation at a slow initial angular speed ω_i, holds two dumbells in his outstretched hands. The student now pulls his arms in close to his body. This reduces his rotational inerta from its initial value $\boldsymbol{I}_i$ to a smaller value $\boldsymbol{I}_f$ as his mass is closer to the rotational axis (the radius r is smaller). His rate of rotation increases markedly, from ω_i to ω_f. If he wants to slow down all he has to do is extend his arms once more.

NOVEL CHALLENGE

A piece of wire has an eye at one end with a small wire hoop through it. When it is rotated in a drill the hoop lifts up. *Now why is this?*

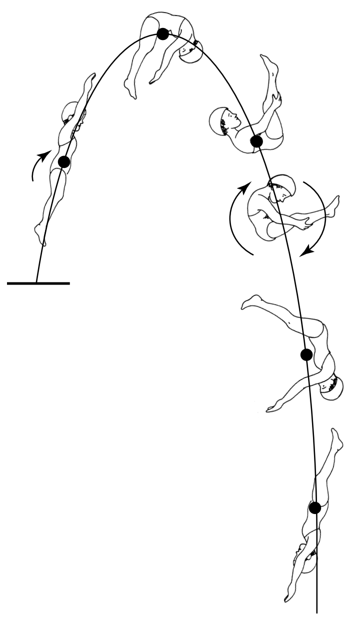


EVERYDAY EXAMPLES OF ANGULAR MOMENTUM 8.11

The springboard diver

Figure 8.34 shows a diver doing a forward one-and-a-half somersault dive. As you would expect, her centre of mass follows a parabolic path. By pulling her arms and legs into the closed tuck position she reduces her rotational inertia and hence increases her angular speed. Pulling out of the tuck position into the open layout position slows her rotation rate.

Figure 8.34 The diver's angular momentum is constant throughout the dive. Her centre of mass follows a parabolic path.



The incredible shrinking star

When a star runs out of nuclear fuel, its temperature decreases and its diameter gets smaller; in fact, it may go from the size of our Sun to just a few kilometres. The star becomes a *neutron star*, so called because the core of the star has been compressed to just an incredibly dense neutron gas. Because stars rotate, the effect of this decrease in radius is an increase in rotational speed. Our Sun rotates once per month; a neutron star may rotate at 800 revolutions per second.

Bullets

When bullets are projected up the barrel of a gun they are guided by spiral grooves inside the barrel. These grooves are called the 'rifling' and give the bullet a high rotational speed by the time it leaves.

One of the most popular firearms among Queensland farmers and sporting shooters is the .257 Weatherby Mark V rifle. It has a barrel 66 cm long and a 'twist' of 30 cm. This means that as the bullet moves up the rifle barrel, it does one complete turn for every 30 cm of barrel length. Seeing that it exits the muzzle with a velocity of 850 m s^{-1}, it means that the bullet is spinning at about 2850 rev/s. You can show this to be correct by dividing the 850 m by 0.30 m to see how many revolutions it does in 1 s. This spinning is designed to keep the bullet travelling point-first so as to reduce air resistance. Without this rotational stabilisation, the bullet would begin to tumble after a short distance and lose its velocity rapidly. It would become useless.

A spinning bullet can be thought of as a cylinder rotating about its long axis. The .257 bullet mentioned above has a diameter of .257 inches (6.5 mm) and a popular type has a mass of 100 grains (6.47 g). For such a bullet to rotate about any other axis, it can be shown that inertia would be greater (by a factor of about 10) and for angular momentum to be conserved, its spin rate would have to be reduced. This is not likely without external forces being applied. So it travels point-first.

Example

Compare **(a)** the linear momentum and **(b)** the angular momentum of a 45 grain .223 Armalite bullet on exiting the muzzle. A .223 bullet has a diameter of .223 inches (5.56 mm) and 45 grains is equal to 2.9 g. The Armalite has a muzzle velocity of 1030 m s^{-1} and a twist of 25 cm, which produces a rotation rate of 4120 rev/s.

Solution

- Linear momentum: $\boldsymbol{p} = m\boldsymbol{v}$
  $= 2.9 \times 10–3 \text{ kg} \times 1030 \text{ m s}^{-1} = 3.0 \text{ kg m s}^{-1}$.
- Angular momentum:
  rotational inertia $\boldsymbol{I} = \frac{1}{2}mr^2$
  $= \frac{1}{2} \times 2.9 \times 10^{-3} \text{ kg} \times (2.78 \times 10^{-3} \text{ m})2 = 1.12 \times 10^{-8} \text{ kg m}^2$
  rotational speed $= 4120 \text{ rev/s} \times 2\pi \text{ rad/rev} = 25\,887 \text{ rad s}^{-1}$
  angular momentum $\boldsymbol{L} = \boldsymbol{I}\omega = 1.12 \times 10^{-8} \text{ kg m}^2 \times 25\,887 \text{ rads}^{-1} = 2.9 \times 10^{-4} \text{ kg m}^2 \text{ s}^{-1}$.

The linear momentum is 10 000 times greater than the angular momentum but each has its own job to do and both are precisely engineered to produce the high impact and good stabilisation effects.

Ballistics experts define 'twist rate' as the number of turns a bullet does per linear metre. Although the rotational speed of the bullet remains constant for most of its journey (due to conservation of angular momentum), the linear speed decreases. Hence the bullet does a lot more spins in a slow metre than it does in a fast metre — so the twist rate increases. It is just a strange way of expressing rotational speed but even keen shooters won't believe you when you tell them that a bullet's twist rate increases as it travels towards the target. You could even win money in a bet. (See Table 8.3.)

NOVEL CHALLENGE

If you spin around a chocolate-covered almond while it is lying flat, an amazing thing happens (usually) — it stands on its end.

Explain this phenomenon in terms of rotational inertia. Before you try it, predict if the fat end or the pointy end will be on the top. What would happen with a Smartie or an M&M?

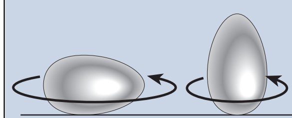


Table 8.3 SOME BALLISTIC DATA*

RANGE (m)	VELOCITY (m s^{-1})	ROTATIONAL SPEED (rev/s)	TWIST RATE (rev/s)	LINEAR MOMENTUM (kg m s^{-1})	ANGULAR MOMENTUM ($\times 10^{-8}$ kg m^2)
0	855	3050	3.6	8.3	7
100	790	3050	3.9	7.7	7
200	730	3050	4.2	7.1	7
300	680	3050	4.5	6.6	7
400	620	3050	4.9	6.0	7
500	570	3050	5.4	5.5	7

* The data are for a Winchester bullet fired from a Remington .308 rifle with a 28 cm twist. The bullet has a diameter of 7.62 mm, a mass of 9.7 g and a muzzle velocity of 855 m s^{-1}.

EI Activity 8.5 BOOK THROWING

Put a rubber band around a hardback book and try throwing it in the air with a rotation about one of its axes as shown in Figure 8.35.

It should be fairly easy to achieve stable rotation about two of the axes, but the third one is very hard. Which of the axes labelled in Figure 8.35 is this difficult one?

The explanation is that rotation about axes that produce maximum or minimum inertia is relatively simple as they are stable against small deviations (wobbles). Intermediate inertia is easier to get to wobble.

Figure 8.35
The three rotational axes of a book.

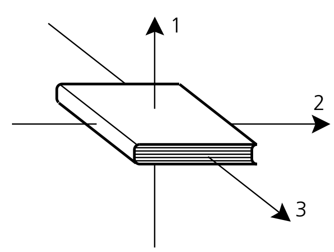


Figure 8.36(a)
The cushion is set at 0.7 times the ball's diameter. This is the same as two-fifths of the radius above the centre of the ball.

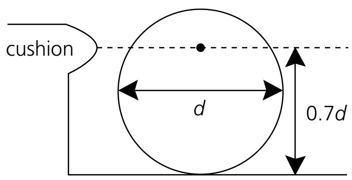
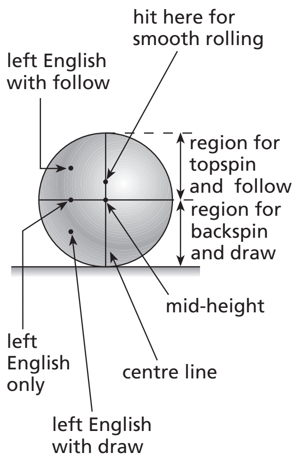


Figure 8.36(b)
Strategic points to hit the cue ball.



Billiards and pool

These games have the feel of physics. Balls collide with each other and bounce off the cushion. But actually the physics of these games is a bit more subtle than that. A skilled player can impart backspin and topspin and other important rotational motions they call 'left and right English'.

When the cue (stick) hits a ball, both linear and rotational motion is imparted. When you strike a billiard ball at mid-height, it will skid away from you and then begin to roll until it collides with another ball or the cushion. Rolling friction is very low. But if it is struck *above* the mid-point it will acquire top spin: the top of the ball moves away from you faster than it otherwise would. Striking the ball *below* the centre results in backspin. You are thus able to control three features of the ball's motion: its linear velocity (by how hard you strike it); the direction of the spin; and how fast it spins.

A spinning ball experiences considerable friction, unlike a rolling ball, which experiences almost none. The direction of the friction is toward you for backspin and so the spin is eliminated quickly; the ball slows down until all spin is lost and then it just continues to roll. With topspin, friction is away from you and slippage tends to speed up the ball. The slippage gradually slows until it exactly matches the forward motion and the ball just rolls. There is a special point on the ball two-fifths of the radius above the centre point, which, when struck, produces a roll with no slippage at all. That is the reason the cushion on the inside edge of the table has a bump at this height — to bounce the ball back without causing it to slip (Figure 8.36**(a)**).

However, players are more interested in collisions with other balls. When the cue ball strikes another ball it transfers its momentum. In a head-on collision, the transfer is complete, leaving the cue ball with no linear motion. In a glancing collision, the cue ball loses only part of its momentum and continues to travel almost at right angles to the motion of the struck ball. In any collision, virtually none of the angular momentum is transferred because of the small amount of friction between the balls. So after a head-on collision, the linear motion stops but the rotation continues. The cue ball will then move back toward you if it has backspin (called 'draw') or away from you if it has topspin (called 'follow') after the collision. If the ball is struck on the left side, it will acquire a clockwise spin (called 'left English') and if struck on the right side will acquire 'right English'. These rotations also affect the result of the collision. See Figure 8.36**(b)**.

Angular momentum (and linear momentum) is conserved in all collisions. A study of momentum transfer is the job of a physicist. To take advantage of the physics is the job of the player.

EI Activity 8.6 BILLIARD PHYSICS

1 If you can get access to a billiard table, see if you can achieve the following (take notes): topspin, smooth rolling, backspin, left English, right English. Note the motion in each case and see if it agrees with the text above. Try hitting the cue ball directly at mid-height.

2 Try hitting the ball with each of the above motions head-on into a stationary ball. Look for topspin and follow, backspin and draw. What happens with left English and follow, left English and draw?

3 Does a glancing blow produce a separation angle of 90° as stated in the above text? What difference does spin have on this angle, if any?

4 What effect on angular momentum transfer does putting chalk on the balls have? This is illegal but it's in the interests of science. Don't do it in a real game — it may give you an unfair advantage.

The falling cat

Don't try this! When a cat falls out of a window upside-down it turns over and lands on its feet. How is this possible? If it starts with no angular momentum, how can the cat acquire it without violating the law of conservation of angular momentum? The answer is that the cat bends itself into a V-shape and by stretching out its front legs while curling up its back legs can change its rotational inertia and turn half its body. It then curls up its front legs and stretches its back legs to change its inertia again and completes the rotation. Pretty clever for a dumb animal! Explore the physics of it for yourself but not with a cat.

NOVEL CHALLENGE

A large ball bearing is placed on a sheet of paper on a desk and the paper is pulled quickly from under the ball.

Does the ball stay in the same place relative to the desk, or what? Please explain!

Questions

21 The 7.62 mm (diameter) bullet has become the NATO standard cartridge for the armed forces. Calculate the angular momentum of a 9.7 g bullet fired from a Russian SKS rifle with a muzzle velocity of 671 m s^{-1} and a rotation rate of 1266 rev/s.

22 Calculate the angular momentum of a smoothly rolling billiard ball of mass 100 g and diameter 8 cm rotating at 10 rev/s.

NEI Activity 8.7 SOME TRICKY QUESTIONS

Here are a few tricky questions on momentum. Before you look at the answers below, try discussing them in class.

1. Why is it hard to stand a bicycle upright but if you give it a push, it will roll along without falling over?
2. Why shouldn't you put your foot on the brakes while you're driving a car through a corner? Racing car drivers only accelerate as they come out of a curve, not while they are in it. Why?
3. A ski turn requires a sinking of the whole body followed by a powerful upthrust and a rotation of the upper part of the body. The lower part of the body rotates the opposite way. Why is this?
4. If you spin a hard-boiled egg and stop it with your finger it stays stopped. A fresh egg will start to spin again. Why?
5. Imagine an egg timer that uses falling sand. If you weigh it while some of the sand is falling in mid-air will it weigh less than when all the sand is at the bottom? After all, some of the sand is in the air and not being supported by the balance.

Answers to Activity 8.7

Answer 1 The spinning wheels acquire angular momentum and for the bike to fall over there has to be a change in this momentum. As well, the bike acquires linear momentum and this also has to be altered.

Answer 2 Sudden braking in a turn throws extra weight on to the front wheels and less on the back wheels resulting in less friction in the rear tyres. This makes it more likely for the car to spin out. Conversely, accelerating puts extra weight on the rear tyres, increasing the friction as drivers come out of a turn.

Answer 3 To conserve angular momentum, when the top part of your body twists one way, the lower part twists in the opposite direction.

Answer 4 The contents of a fresh egg continue to spin when the shell is stopped. When the shell is released the contents make the shell spin again.

Answer 5 They will weigh the same. The loss in weight because some of the sand is in mid-air is compensated for by the impact of the sand when it strikes the bottom. These sand grains transfer their momentum and hence extra force to the balance.

Figure 8.37
For question 27.

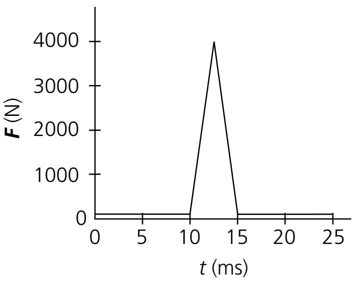


Figure 8.38
For question 28.

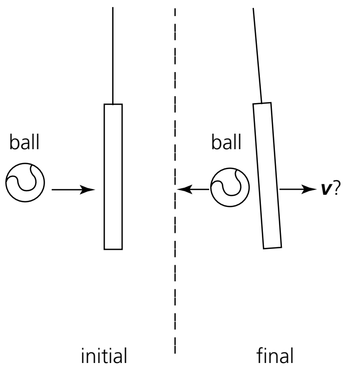
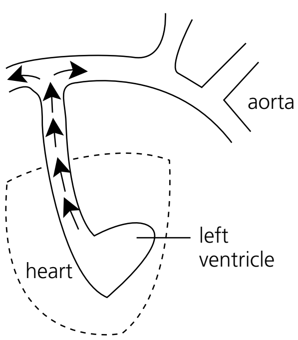


Figure 8.39
Blood from the heart comes up the aorta from the left ventricle. The aorta branches at a 'T'-junction.



Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***23** Masses of 20 kg and 35 kg are on the ends of a 1.4 m long bar. Determine the centre of mass of the system.

***24** What is the momentum of: **(a)** a cricket ball of mass 160 g moving at 12.5 m s^{-1} east; **(b)** a billiard ball of mass 200 g moving at 8.5 m s^{-1} N35°E; **(c)** a 100 kg footballer moving with a velocity of 8 m s^{-1} north?

***25** An alpha particle of mass 7×10^{-27} kg is accelerated from 5×10^{3} m s^{-1} to 2×10^{4} m s^{-1}. Calculate the change in momentum.

***26** A car of mass 2200 kg accelerates from rest at 3 m s^{-2} for 10 s. Determine the impulse imparted to the car.

****27** When a ball of mass 180 g is struck by a bat moving in the opposite direction, the force acting on the ball is as shown in the graph (Figure 8.37). Determine **(a)** the impulse; **(b)** the final velocity of the ball if it was initially moving at 10.0 m s^{-1} south.

****28** A ball of mass 50 g moving horizontally at a speed of 40 cm s^{-1} strikes a suspended plate of mass 1000 g and rebounds from it with a speed of 25 cm s^{-1} as illustrated in Figure 8.38. Find the speed with which the plate begins to move.

****29** Two masses of 4 kg and 3 kg respectively are travelling east along a frictionless surface with respective speeds of 12 m s^{-1} and 5 m s^{-1}. If the 3 kg mass continues to move east with a speed of 9 m s^{-1} after the collision, calculate the speed of the 4 kg mass.

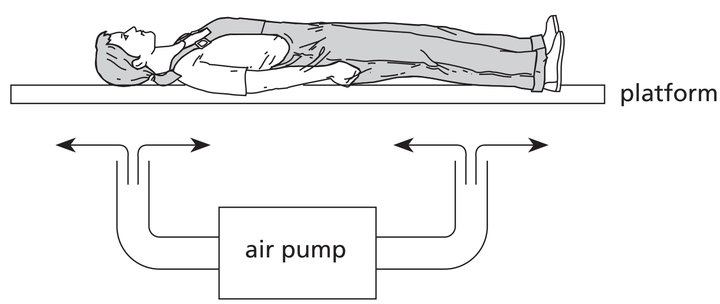
****30** A railway truck of mass 4000 kg moving with a speed of 3.6 m s^{-1} collides with a stationary truck of mass 2400 kg. The two trucks become coupled together. What is their common speed?

****31** What is the angular momentum of the Earth associated with rotation about its own axis? The mass of the Earth is 5.98×10^{24} kg and its radius is 6.37×10^{6} m.

****32** A ball A of mass 6 kg is moving east at 3.5 m s^{-1} when it collides with a stationary ball B of mass 8 kg. Ball A heads north at 5 m s^{-1} after the collision. Determine the final velocity of ball B.

****33** The **ballistocardiograph** (BCG) is an important device in medicine and is designed to employ simple physics to analyse the effectiveness of operations on a patient's heart. With each heartbeat, about 70 g of blood is ejected from the left ventricle of the heart into the aorta (Figure 8.39). The speed of the blood is about 30 cm s^{-1}. Hence the blood from each heartbeat has momentum. The body recoils with each heartbeat due to conservation of momentum. This can be registered on a very sensitive balance attached to the platform on which the body rests. (See Figure 8.40).

Figure 8.40
The air table for the ballistocardiograph.



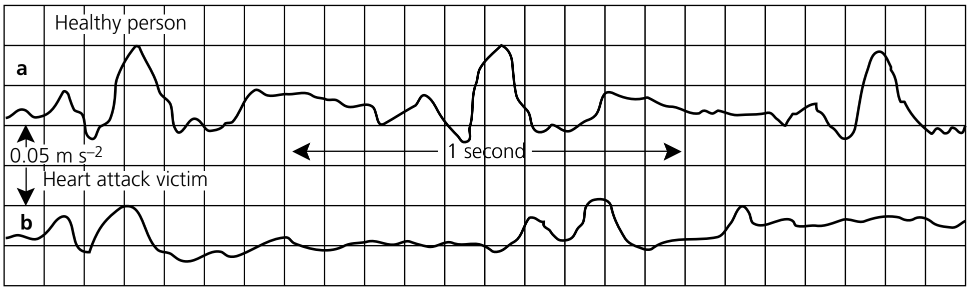


Figure 8.41
A ballistocardiogram of a healthy person and a heart attack victim.

The graph produced shows the acceleration of the body during the different stages of the heartbeat. In Figure 8.41 a ballistocardiogram is shown for both a healthy person and one who has suffered a heart attack. Accelerations can be read to an accuracy of 10^{-5} m s^{-2}.

(a) When the body recoils, the table moves. How has friction been taken into account?
(b) When a pulse of blood travels from ventricle to aorta, which way would the body move?
(c) On the graph, the acceleration goes negative after the main part of the heartbeat. Why is this?
(d) An acceleration of about 0.06 m s^{-2} is considered healthy. What is the value for the heart attack victim?
(e) Extract the information needed to calculate the momentum of a blood pulse and calculate it.

Extension — complex, challenging and novel

*****34** A radioactive Thorium (Th) nucleus decays by emitting an electron and a neutrino at right angles to each other in a horizontal plane. The momentum of the electron is 6×10^{-21} kg m s^{-1} and that of the neutrino 2×10^{-20} kg m s^{-1}. If the mass of the resulting protactinium (Pa) nucleus is 3.905×10^{-25} kg, calculate **(a)** the total momentum of the three particles immediately after the decay; **(b)** the recoil momentum of the nucleus; **(c)** the recoil speed of the nucleus.

*****35** William Tell fires an arrow into a 100 g apple. The arrow has a mass of 100 g and travels at 50 m s^{-1} horizontally. As the arrow hits it, the apple splits into two pieces. One piece weighing 50 g flies vertically up at 20 m s^{-1} while the other piece gets stuck on the arrow and continues on. Calculate the velocity of the arrow and second piece of apple together.

*****36** A body of mass 400 g is moving along a smooth surface at a velocity of 10 m s^{-1} east. It strikes a body of mass 650 g, initially at rest, and then the 400 g body moves at a velocity of 8 m s^{-1} in a direction E35°N. What is the velocity of the 650 g object?

*****37** A rocket of mass 25 000 kg is cruising through space with a constant speed of 1×10^3 m s^{-1} when exhaust gases are expelled for 10.0 seconds at a rate of 500 kg s^{-1} with a speed of 5×10^4 m s^{-1}. Calculate the new speed of the rocket.

✔ TEST YOUR UNDERSTANDING

(Answer true or false)

- Objects moving in straight lines cannot have angular momentum.
- Momentum is not a vector.
- Momentum is the same as a force.

CHAPTER 09

Work and Energy

RUNNING OUT OF ENERGY 9.1

If you leave a torch turned on, its batteries will run out of energy. But has the energy gone forever? Where did it go? These are fundamental questions when it comes to energy. As you probably learnt in earlier science studies, energy is not lost — it just gets transferred from one place to another. This is called the **law of conservation of energy**. The universe seems to have a finite amount of energy that is continually being rearranged.

Think about these questions:

- If you stand still, you are using up energy, but where does it go?
- Why can't a ship extract heat from sea water to power its engines?
- Will there really be an energy crisis soon? Are we running out of energy?

NOVEL CHALLENGE

Why do you lean forward when you get up out of a chair?

NOVEL CHALLENGE

When bodies interact, the energy of one may increase at the expense of another. But we can't intercept the energy and bottle it. So comment on this assertion: 'Energy is not a thing; it is a property of a body.'

NEI Activity 9.1 ENERGY AT HOME

To help you become more familiar with the energy, work and power terms, find out the following:

1. *Electric kettle* If you have an electric kettle, look underneath for its rate of energy consumption, which will be expressed in watts (W). For example, the Kambrook Flash has 2000 W stamped on it. Did anyone in the class get below 1600 W or over 2400 W? Why couldn't the manufacturers make a 20 000 W kettle? It would boil water in a flash!
2. *Food energy* Most foods have their energy content written on the label. A 'Popper' apple juice carton states that the energy is 206 kJ per 100 g. But it also expresses it another way. Look at a food container from your cupboard and note the two ways that energy content is expressed.
3. *Engine power* If your family owns a car, truck or motorbike and you can find the owner's manual, find out the power output of the engine. For example, a GXL Turbo Land Cruiser has a power output of 118 kW. But following this number is a further specification to do with the power. What is it?

ENERGY AT WORK 9.2

A simple definition of energy is that **energy is the capacity to do work.** The word 'energy' stems from the Greek *en* meaning 'in', and *ergos* meaning 'work'. But this doesn't really give us a good understanding of the idea of energy and work. Physicists didn't develop a good understanding of these concepts until 100 years after Newton's death. Today these ideas are considered fundamental to the processes of nature.

Energy transfers

The above definition indicates that energy can be converted into useful work; for example, when the electrical energy in a car's battery is used to start the engine. The reverse is also true — work can be converted into stored energy. For example, we can do work to pump water from a lake to a high reservoir. That stored water has higher energy because of its height and can later be used to drive electric generators and produce electrical energy. When energy is transferred *to* an object we say work is done *on* the object; when energy is transferred *away from* an object we say that work is done *by* the object:

- Energy transferred *to* an object (work done *on* the object): e.g. water pumped up to a reservoir.
- Energy transferred *away from* an object (work done *by* the object): e.g. water flows back down.

Energy losses in transfer

When a torch is turned on, some of the energy stored in the chemical bonds is transferred to the electric charge that flows through the bulb. Some of this energy is transferred into light and some as heat energy to the glass bulb and air.

Energy transfers never achieve 100% **efficiency**, that is, some of the energy is transferred to places you don't intend it to go to. For instance, the energy from the torch that goes to heating up the glass and air is wasted — it is a loss in the sense that it didn't get turned into light. But it is not really lost; energy never is. It just goes to the wrong place. **Efficiency is a measure of the useful energy output compared with the energy input.**

$$\%\ \text{efficiency} = \frac{\text{energy out}}{\text{energy in}} \times 100\%$$

Some energy transfers are listed in Table 9.1.

Table 9.1 ENERGY TRANSFERS AND LOSSES

DEVICE	USEFUL ENERGY TRANSFER (ENERGY IS CONVERTED TO USEFUL WORK)	% OF TOTAL ENERGY TRANSFERRED THAT IS USEFUL (% EFFICIENCY)	NON-USEFUL ENERGY TRANSFERS (ENERGY IS NOT CONVERTED TO USEFUL WORK)
Petrol engine	chemical → mechanical	25	heat, sound
Electric light	electrical → light	5	heat
Fluorescent light	electrical → light	20	non-visible radiation
Solar cell	light → electrical	21	heat; re-emission of light
Battery	chemical → electrical	85	heat
Electric motor	electrical → mechanical	90	heat

Forms of energy

The jumble of terms like light, heat, electricity, sound, mechanical and chemical doesn't provide a systematic way of organising the different forms of energy. Before we can go any further, we need a way of classifying energy.

Bodies that are moving have **kinetic energy** (E_K, KE), e.g. a flying bird, a shooting star, a moving locomotive, a speeding bullet.

Bodies that can do work because of their position have **potential energy** (E_P, PE or U), e.g. water in a reservoir, a compressed spring, a stretched rubber band.

Kinetic and potential energy are said to be forms of **mechanical energy**.

NOVEL CHALLENGE

In 1916, a Dr Taylor observed a man carrying 40 kg 'pigs' of iron 11 m up a 2.4 m high incline to a train carriage. He carried 1156 pigs in 10 hours. The man's mass was 65 kg and he rested for 15% of the time. *What was his average power output for the $8\frac{1}{2}$ hours? On a later occasion and without a rest he could only carry 305 pigs in the 10 hours. By what factor was his power output increased when he had proper rest? Suggest why cyclists use a sprint–coast–sprint sequence.*

INVESTIGATING

What energy transfer occurs for humans? What percentage of the input energy is transferred to non-useful purposes?

Where does this leave chemical, heat and electrical energy? Because they are to do with the random vibrations or motions of electrons, atoms and molecules within an object, they are said to be forms of **internal energy** (E_i). This chapter deals only with mechanical energy. Heat, sound, electricity and nuclear energy are dealt with in later chapters. Chemical energy is mainly left to the other physical science — chemistry.

WORK 9.3

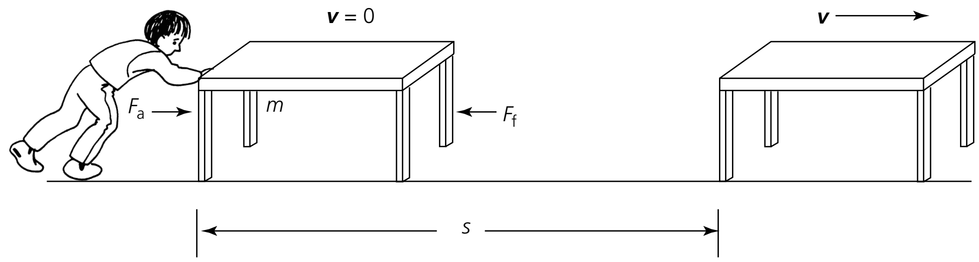


Figure 9.1
The desk with mass *m* is moved from rest a distance *s* across the floor by an applied force ***Fa*** against the frictional force ***F***$_f$. It acquires a velocity v.

NOVEL CHALLENGE

Could you shift a destroyer (a 20 000 tonne ship) moored in a dock with the ropes slack? Let's assume you can apply a force of 500 N. We say 'Yes'; but how long do you think it would take to push it 2 m away from the dock: 400 seconds, 400 hours, or forget it?

If you tried to push a desk across the floor and it didn't move, you might say you did a lot of work on the desk. But to a physicist, if it didn't move then no work was done. If you did move it, then the work done would depend on how hard you pushed and the distance it moved. The word 'work' is often used very loosely; for example, have you done your home*work* tonight? Physicists define work very carefully; **work is defined as the product of the force and the distance moved in the direction of an applied force**. It is a scalar quantity, and yet is the product of two vector quantities.

$$\text{Work} = \text{force} \times \text{displacement} \quad \text{or} \quad W = \boldsymbol{Fs}$$

Since force is measured in newtons and displacement is measured in metres, work has the units **newton metre** or N m. The newton metre is called the **joule (J)** in honour of James Joule (1818–89), an English physicist who studied heat and electrical energy.

Example

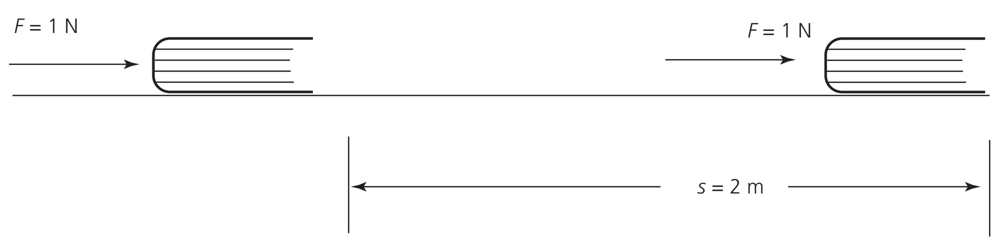


Figure 9.2
Work is done when a force is used to push a book along a desk.

Calculate the work done when a horizontal force of 1 N is used to push a book a distance of 2 m along a desk (Figure 9.2).

Solution

$$W = \boldsymbol{Fs} = 1\ \text{N} \times 2\ \text{m} = 2\ \text{J}$$

Doing no work in class

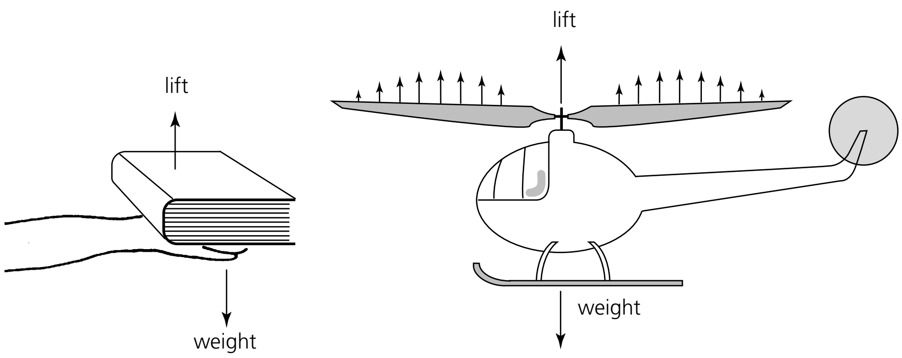


Figure 9.3
No work is done on the book or the helicopter if there is no movement.

Our definition of work leads to the surprising conclusion that, in a scientific sense, you are not doing any work on a book if you hold it in your outstretched arm for a long period of time. Sure, you get tired — but no work is done *on the book*. You will feel tired because your muscles are using energy and burning up fuel. As the muscle fibres relax and contract just keeping your arm still you are using energy. But it is not being transferred to the book. No work is done on the book. But there is a change in the internal energy of your body as microscopic molecular internal motions and reactions go on. The same is true for a helicopter hovering in a stationary position above the ground. These motions never result in any measurable displacement and therefore never do any work in the sense that the word is used in physics.

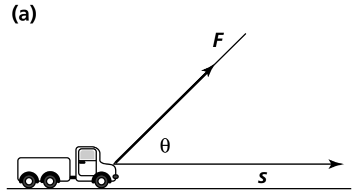
Forces at an angle

When you push a lawnmower or a shopping trolley, the force from your arms is at an angle to the direction of the motion. The same is true if a child pulls a toy by a string at an angle θ along the floor as shown in Figure 9.4 **(a)**. In this case the equation for work done is a little different.

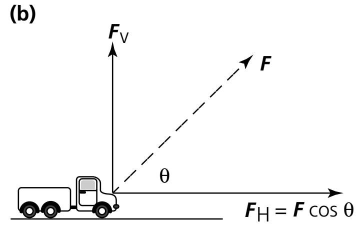
If a force ($\boldsymbol{F}$) is used to pull the toy along a horizontal floor, the useful part of the force is the component in the direction of motion. In Chapter 4 you would have seen that this force ($\boldsymbol{F}_{\text{horizontal}}$ or $\boldsymbol{F}_{\text{H}}$) is equal to $\boldsymbol{F}\cos\theta$. This is shown in Figure 9.4 **(b)**. Some of the force is 'wasted' and tends to lift the toy off the floor. It is not converted to useful work pulling the toy along the floor.

Figure 9.4
Pulling a toy truck with a cord that makes an angle θ with respect to the direction of the truck's motion: $W = \boldsymbol{Fs}\cos\theta$.

(a)



(b)



Example

A force of 5.0 N is applied to a string attached to a toy truck that makes an angle of 40.0° to the floor. How much work is done in dragging the truck a distance of 4.0 m across the floor?

Solution

The horizontal component ($\boldsymbol{F}_{\text{H}}$) = $\boldsymbol{F}\cos 40° = 5.0 \times 0.766 = 3.83$ N.

$$W = \boldsymbol{Fs} = 3.83 \times 4.0 = 15 \text{ J}$$

Alternatively:

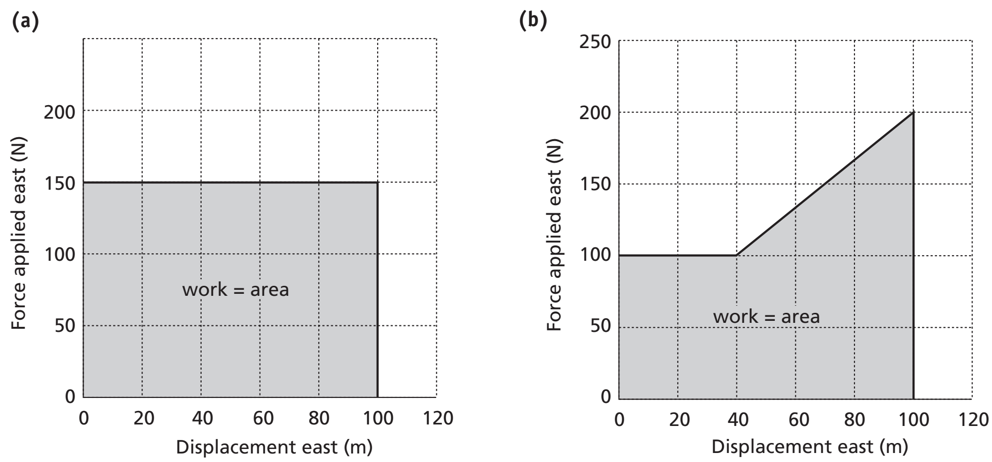
$$W = \boldsymbol{Fs}\cos\theta = 5.0 \times 4.0 \times \cos 40° = 15 \text{ J}$$

Notice that when $\theta = 0°$, the force is in the direction of motion and the formula returns to $W = \boldsymbol{Fs}$ because the cosine of 0° (cos 0°) = 1.

Graphs

Force and displacement can be expressed graphically. In Figure 9.5**(a)**, a constant force of 150 N is applied to an object and shifts it a distance of 100 m. The work done is simply the product of $\boldsymbol{F} \times \boldsymbol{s}$ and so is the shaded area under the line (150 000 J). If the force does not remain constant but varies as in Figure 9.5 **(b)**, the work is still determined by calculating the shaded area (13 000 J).

Figure 9.5
(a) Area = 150 × 100 = 15 000 J.
(b) Area = (100 × 100) + (60 × 100) ÷ 2
= 13 000 J.



Lifting things

When calculating the work done in lifting an object vertically, the force applied will be equal to the object's weight ($\boldsymbol{F}_w = m\boldsymbol{g}$). This is assuming that it is lifted at constant speed. If the speed is varied then Newton's second law formula ($\boldsymbol{F} = m\boldsymbol{a}$) would have to be applied.

Example

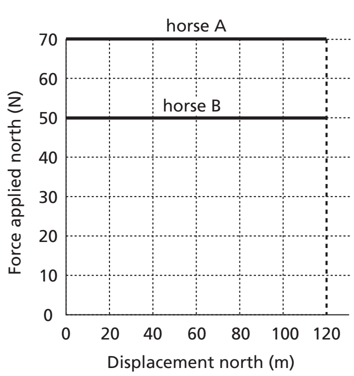
Calculate the work done in lifting a 15 kg schoolbag at constant speed from the floor to a port rack 1.8 m off the ground.

Solution

$$\boldsymbol{F}_w = m\boldsymbol{g} = 15 \times 10 = 150 \text{ N}$$
$$W = \boldsymbol{F}\boldsymbol{s} = 150 \times 1.8 = 270 \text{ J}$$

A *Mars bar* provides you with 1 235 000 J (1235 kJ) of energy. This is equivalent to lifting your bag 4500 times to burn off the energy. It seems hardly worth the effort! But luckily, in lifting a bag, your body uses up a lot more energy than just the amount needed to overcome gravitational forces. Just as well — wouldn't you get really fat!

Figure 9.6
For question 3.



Questions

1 Calculate the amount of work done in:
(a) pulling a bag of dog food 3.5 m along a table by applying a 25 N horizontal force; **(b)** lifting a 20 kg bag of dog food at constant speed on to a table 85 cm off the ground; **(c)** pumping 200 kg of water at a constant flow rate into a tank 25 m high.

2 A 200 kg piano is lowered by a rope out of a third-floor window.
(a) Calculate how much work is done in lowering the piano a distance of 9 m to the ground at constant speed.
(b) Was work done *on* or *by* the piano in this process?

3 A team of two horses is pulling a loaded cart in a northerly direction along a horizontal road at constant speed. A force–displacement graph is shown in Figure 9.6.

(a) Calculate the work done by each horse.
(b) What is the total work done on the cart?

4 Figure 9.7 shows different forces acting on different objects. Calculate the work done in each case.

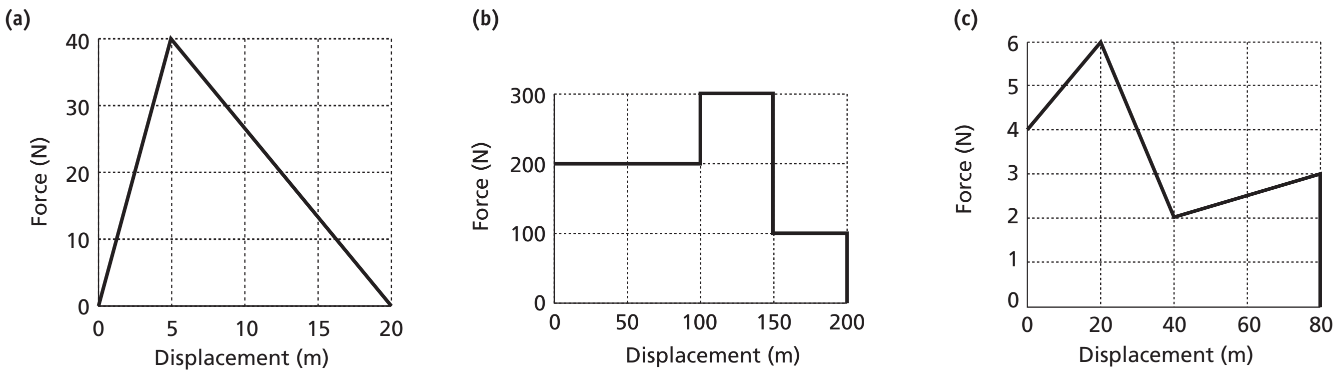


Figure 9.7
For question 4.

9.4 KINETIC ENERGY

A bowling ball resting on the floor has no energy of motion. One that is rolling along a bowling alley does have energy of motion. Energy due to the motion of an object is called **kinetic energy** (Greek *kinema* = 'motion').

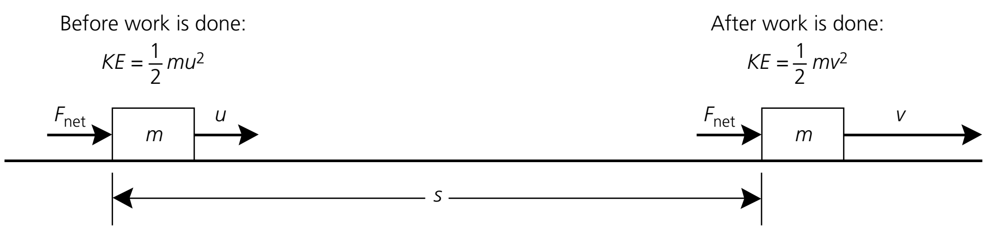


Figure 9.8
Hockey puck on an air table.

To determine an equation for kinetic energy, we will use concepts already developed. Imagine a hockey puck moving on a frictionless surface such as an air table. If an unbalanced force $\boldsymbol{F}_{net}$ is applied to it for a period of time t, the force produces accelerated motion and the object goes from an initial velocity $\boldsymbol{u}$ to a final velocity $\boldsymbol{v}$:

$$\boldsymbol{v}^2 = \boldsymbol{u}^2 + 2\boldsymbol{a}\boldsymbol{s}$$
$$\boldsymbol{a} = \frac{\boldsymbol{v}^2 - \boldsymbol{u}^2}{2\boldsymbol{s}}$$

By letting $F = m\boldsymbol{a}$, the work done in accelerating the puck is given by:

$$W = \boldsymbol{F}\boldsymbol{s} = m\boldsymbol{a}\boldsymbol{s} = m\left(\frac{\boldsymbol{v}^2 - \boldsymbol{u}^2}{2\boldsymbol{s}}\right) \times \boldsymbol{s} = \tfrac{1}{2}m\boldsymbol{v}^2 - \tfrac{1}{2}m\boldsymbol{u}^2$$

This represents a change in the quantity that we call kinetic energy.

Work done equals change in kinetic energy $W = \Delta E_K$

Hence: if an object starts from rest, its final kinetic energy is given by:

$$E_K = \tfrac{1}{2}m\boldsymbol{v}^2$$

This is properly referred to as its **translational** kinetic energy because rolling or rotating objects also have **rotational** kinetic energy. This is not dealt with here.

You or your teacher may prefer to use the symbol *KE* for kinetic energy — it's a matter of choice. Note that kinetic energy is a scalar quantity and hence does not require direction.

Example 1

Calculate the translational kinetic energy of a 6.0 kg bowling ball rolling at 5.0 m s^{-1}.

Solution

$$E_K = \tfrac{1}{2}m\boldsymbol{v}^2 = \tfrac{1}{2} \times 6 \times 5^2 = 75 \text{ J}$$

Example 2

A 520 kg rocket sled at rest is propelled along the ice by an engine developing a constant thrust of 12 000 N. Assuming all of the work goes into motion, calculate its velocity after 40 m.

Solution

- Work done by engine: $W = \boldsymbol{Fs}$ = 480 000 J.
- Work is converted to kinetic energy, hence E_K = 480 000 J.

$$E_K = \tfrac{1}{2}m\boldsymbol{v}^2 \quad \text{or} \quad 480\,000 = \tfrac{1}{2} \times 520\ \boldsymbol{v}^2$$

$$\boldsymbol{v} = \sqrt{\frac{2 \times 480\,000}{520}} = 43 \text{ m s}^{-1}$$

The J-manoeuvre

In aircraft dogfights, it can be difficult to out-manoeuvre planes of similar ability. A technique developed by the US Airforce can tip the balance. It is called the *J-manoeuvre*. When an exhaust nozzle that can turn sideways, up or down is added to the rear of their F-16 fighters, they can radically alter the performance of the planes. The rotating exhaust is similar to that used in the British Harrier Jump Jets (remember Schwarzenegger in the movie *True Lies*!), but is far more flexible. With it, pilots can bring a 2200 km h^{-1} F-16 to a halt in a few seconds.

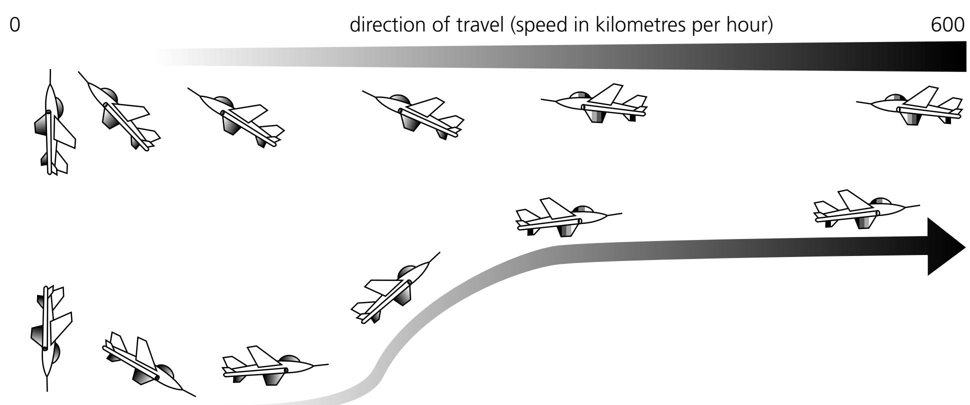


Figure 9.9
The J-manoeuvre: the aerial equivalent of a handbrake turn.

Conventional jet engines thrust the plane forward and the pilot steers using aerodynamic controls — ailerons, elevator and rudder — which alter the air stream and cause the plane to change directions. With the J-manoeuvre, the pilot pulls the nose of the plane through 90° vertically while the plane continues to travel horizontally 'belly-first' (Figure 9.9). This brings the plane to a halt without climbing (the normal way to slow down). The pilot then turns the plane through 180° sideways and points its nose downward and as it picks up speed, the pilot brings it back to level flight. It is much the same as a motorcycle 'wheelie'. So, a 13 tonne jet fighter can stop in about 5 seconds and then reverse directions. This is a 'loss' of 2.4 billion joules of kinetic energy or about 500 megawatts of power being shed without change in potential energy (altitude) — a clever trick. The pilot needs a really strong seatbelt harness system; and a strong stomach.

NOVEL CHALLENGE

In November 1998, forensic police officer Gerard Dutton of the Tasmania Police in Hobart investigated an incident where a man was killed by a piece of fencing wire 27 mm long, 2.4 mm diameter and 0.89 g mass that was hurled up by a roadside slasher mower. The death was curious because international data on 'incapacitation energy' developed by the US Army said that a person would be killed only if the energy of a projectile was between 40 and 236 joules. The fencing wire fragment was thrown from the edge of a rotating blade of radius 54 cm travelling at 2000 revolutions per minute. Was the projectile's energy within the incapacitation energy range?

9.5 ENERGY CHANGES AND COLLISIONS

In the chapter on momentum we saw that in *all* collisions, momentum is conserved. We will now look at conservation of kinetic energy in collisions.

Collisions can be either **elastic** or **inelastic**. The word elastic comes from the Greek *elastikós* meaning 'to drive' or 'propel'.

Elastic collisions

An elastic collision is one in which kinetic energy is conserved.

Conserve is from the Latin for 'to preserve' or 'keep the same'. Jams are often called conserves because they preserve the fruit. In elastic collisions the total amount of kinetic energy before the collision is the same as the total kinetic energy after the collision. Collisions between gas molecules are perfectly elastic. If they weren't, the gas would lose energy and the pressure in a spray can would decrease while it was only sitting on a shelf. Clearly this does not happen. Collisions between steel ball bearings is also approximately elastic.

For an elastic collision (Figure 9.10):

$$E_K(\text{initial}) = E_K(\text{final})$$
$$\tfrac{1}{2}m\boldsymbol{u}_a^2 + \tfrac{1}{2}m\boldsymbol{u}_b^2 = \tfrac{1}{2}m\boldsymbol{v}_a^2 + \tfrac{1}{2}m\boldsymbol{v}_b^2$$

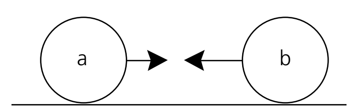


Figure 9.10
Balls colliding head-on.

Example 1

A 3 kg steel ball moving east at 4 m s^{-1} collides with a stationary 1 kg ball. After the collision, the 3 kg mass moves east at 2 m s^{-1} and the 1 kg mass moves east at 6 m s^{-1} (Figure 9.11). Is this collision elastic?

Solution

$$\begin{aligned} E_K(\text{initial}) &= \tfrac{1}{2}m\boldsymbol{u}_a^2 + \tfrac{1}{2}m\boldsymbol{u}_b^2 \\ &= \tfrac{1}{2}\,3 \times 4^2 + \tfrac{1}{2}\,1 \times 0^2 \\ &= 24\text{ J} \\ E_K(\text{final}) &= \tfrac{1}{2}m\boldsymbol{v}_a^2 + \tfrac{1}{2}m\boldsymbol{v}_b^2 \\ &= \tfrac{1}{2}\,3 \times 2^2 + \tfrac{1}{2}\,1 \times 6^2 \\ &= 24\text{ J} \end{aligned}$$

The collision is elastic.
Note: check for yourself that momentum is conserved.

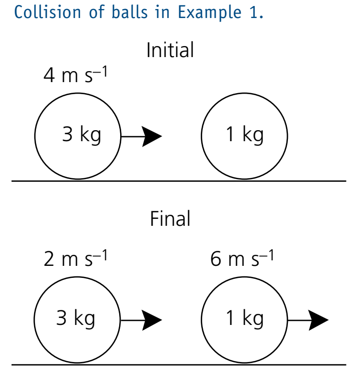
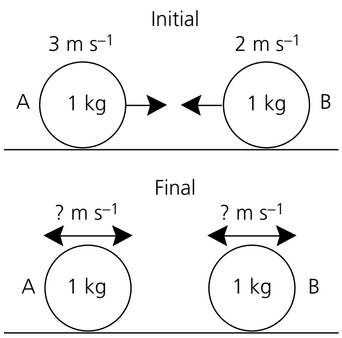


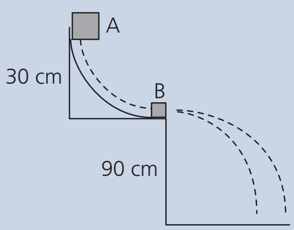
Figure 9.11
Collision of balls in Example 1.

Figure 9.12
Collision of balls in Example 2.



NOVEL CHALLENGE

Blocks A and B are of mass $2m$ and m respectively. Block A is allowed to slide down the curved incline until it hits block B in an elastic collision. Which block will travel the farthest out? How far away from each other will they strike the ground? We say 1.6 m; how about you? We think you'll need conservation of energy *and* momentum to do this one.



INVESTIGATING

To counter the high bounce of the WACA (Perth) cricket ground, many players use bats that have the centre of percussion (the 'meat') higher up the bat than normal. When in Brisbane they can revert to a bat with lower centre of percussion and in India where the wickets are really flat, it's even lower. *How many centimetres variation are there in the bats?*

Example 2

An object A of mass 1.0 kg moving at 3.0 m s^{-1} east collides elastically head-on with an object B of mass 1.0 kg moving at 2 m s^{-1} west. Determine the final velocity of each object. (See Figure 9.12.)

Solution

The solution involves some *complex reasoning* as conservation of both momentum *and* kinetic energy is required.

Let east be the positive direction.

1 *Conservation of momentum:*

$$m_A\boldsymbol{u}_A + m_B\boldsymbol{u}_B = m_A\boldsymbol{v}_A + m_B\boldsymbol{v}_B$$
$$1 \times 3 + 1 \times {}^-2 = 1 \times \boldsymbol{v}_A + 1 \times \boldsymbol{v}_B$$
$$3 - 2 = \boldsymbol{v}_A + \boldsymbol{v}_B$$
$$1 = \boldsymbol{v}_A + \boldsymbol{v}_B$$
$$\boldsymbol{v}_A = 1 - \boldsymbol{v}_B$$

2 *Conservation of kinetic energy:*

$$\tfrac{1}{2}m\boldsymbol{u}_A^2 + \tfrac{1}{2}m\boldsymbol{u}_B^2 = \tfrac{1}{2}m\boldsymbol{v}_A^2 + \tfrac{1}{2}m\boldsymbol{v}_B^2$$
$$\tfrac{1}{2} \times 1 \times 3^2 + \tfrac{1}{2} \times 1 \times -2^2 = \tfrac{1}{2} \times 1 \times \boldsymbol{v}_A^2 + \tfrac{1}{2} \times 1 \times \boldsymbol{v}_B^2$$

Multiply both sides by 2 to cancel out the $\tfrac{1}{2}$

$$9 + 4 = \boldsymbol{v}_A^2 + \boldsymbol{v}_B^2$$
$$13 = \boldsymbol{v}_A^2 + \boldsymbol{v}_B^2$$
$$\boldsymbol{v}_A^2 = 13 - \boldsymbol{v}_B^2$$

3 *Solving both equations simultaneously and eliminating the term $\boldsymbol{v}_A$:*

$$\boldsymbol{v}_A^2 = (1 - \boldsymbol{v}_B)^2 = 1 - 2\boldsymbol{v}_B + \boldsymbol{v}_B^2 \text{ and } \boldsymbol{v}_A^2 = 13 - \boldsymbol{v}_B^2$$

hence:

$$1 - 2\boldsymbol{v}_B + \boldsymbol{v}_B^2 = 13 - \boldsymbol{v}_B^2$$
$$2\boldsymbol{v}_B^2 - 2\boldsymbol{v}_B - 12 = 0$$

using the quadratic formula or factorising into $(\boldsymbol{v}_B - 3)(\boldsymbol{v}_B + 2) = 0$, gives two solutions for $\boldsymbol{v}_B$. They are $\boldsymbol{v}_B = 3$ m s^{-1} and -2 m s^{-1}. The second of these solutions is the case where the velocities are the same as the initial. In other words, no collision took place; they were on parallel but separate tracks.

The first solution is where a collision took place. In this case $\boldsymbol{v}_B = 3$ m s^{-1} and substituting this into the equation $\boldsymbol{v}_A = \boldsymbol{v}_B - 1$ gives a value for $\boldsymbol{v}_A$ of -2 m s^{-1} (west).

This is a long procedure and students find many pitfalls. But if you are careful and meticulous and practise many more examples, success should be yours.

Equal masses colliding elastically

Note that in the above example, the objects swapped their velocities. Objects A and B had speeds of 3 and -2 m s^{-1} respectively; after collision, these had become -2 and 3 m s^{-1} respectively. This is true of linear elastic collisions between objects of **equal mass**. It is sometimes called the 'pool player's result'.

EI Activity 9.2 THE POOL HALL REVISITED

1. Next time you are playing pool or billiards, hit a ball into a stationary one and see if they swap velocities. The cue ball should stop moving and the struck ball should take off. As explained in previous chapters, if you impart spin to the ball other effects come into play. For this experiment you should hit the ball about seven-tenths of the diameter from the base. This produces no spin. Can you explain why it won't spin?

2 Try it again but have one ball moving slowly and the other fast. Do they swap velocities?

3 If you can't get to a pool table, try it with ball bearings rolling from opposite directions along a grooved and curved ruler or try it with Newton's cradle (Newton's balls). (See Figure 9.13**(b)**.)

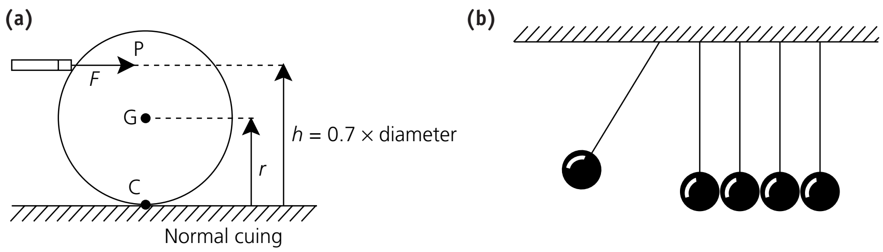


Figure 9.13
(a) The force *F*, applied along the line through point P, causes translation and also rotation about the centre of mass G;
(b) Newton's balls.

Inelastic collisions

An inelastic collision is one in which the total kinetic energy is **not** conserved.

Kinetic energy is lost. The 'missing' kinetic energy is transferred to other types of energy such as thermal (heat) energy and sound. Most collisions in the real world are of this sort. When two objects cling or lock together after impact (i.e. they become coupled), the collision is inelastic.

Examples include:

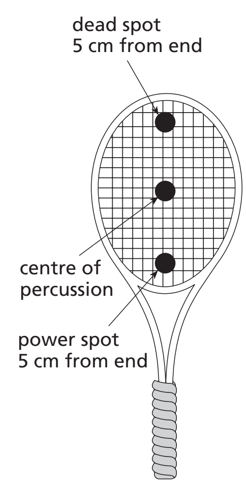
- cars colliding
- bullet hitting a target
- meteorite striking the earth
- tennis ball being struck by a racquet.

When different objects are dropped on to a concrete floor, they bounce to different heights. A perfectly elastic collision would see the ball returning to its original height. A superball is about 90% elastic, a golf ball about 60% and a lump of putty is perfectly *inelastic*.

EI Activity 9.3 ANYONE FOR TENNIS?

What part of the tennis racquet should you use to hit a tennis ball? Well, it all depends on whether you are serving or returning. In this activity you can check and extend the results of Rod Cross, a physicist from the University of Sydney who clamped a tennis racquet to a bench and measured the bounce of a tennis ball from different power points of the racquet head. He found that a point 5 cm from the top end was a dead spot giving no bounce at all. This is good for getting maximum power into the serve as all the kinetic energy of the racquet goes into the ball. At the other end (5 cm in), there is another power spot, which has maximum bounce. This is good for returning a fast ball, but no good for a serve. In the middle is the centre of percussion — no 'ringing' of the hands and gives medium bounce.

Figure 9.14



1 Clamp a racquet to a bench and repeat his tests. Graph the bounce (in cm) for a given drop height versus the distance from the top of the racquet.

2 Does the shape of the graph change when the drop height (i.e. the speed) changes? Of what significance is this?

3 How does the bounce height vary across the racquet? Have you ever seen a three-dimensional graph? How could you show the variation in bounce height across the racquet as well as from the top-to-bottom of the racquet on the one graph?

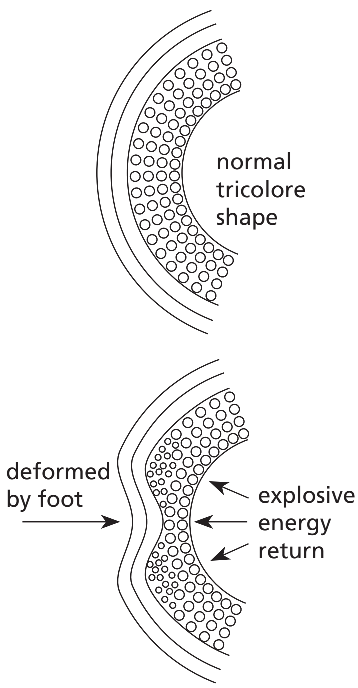
4 Obtain a videotape of a power server and see where the player hits the ball for a serve. Does he or she really use the dead spot?

World Cup players face a whole new ball game

In the mid-1990s, the governing body of soccer (FIFA) decided to change the nature of the regulation soccer ball to add more excitement and goals to matches. Adidas produced a new ball that had 5% extra 'zing', much to the consternation of soccer players, who began over-hitting the ball. The Questra ball has a special polyurethane coating that reduces friction and, on the inside, has a rubbery layer to get 'more bang from the boot'. It achieves this by reducing deformation of the ball and allowing more of the kinetic energy to be transferred from the foot to the ball. In the FIFA bounce test, a ball is dropped from 2.0 m and the time is measured for it to reach its maximum height after bouncing. The Questra ball takes 5% less time than the traditional World Cup ball — the Italian Etrusco Unico.

For the World Cup 'France-98', the Adidas Tricolore was used. It has a mass of 450 g and circumference of 70 cm. The outer layer was a printed polyester film, the second layer a polyurethane matrix of individually closed gas-filled microballoons (see Figure 9.15). The third layer was a poly-cotton mixed fibre backing and the fourth layer was the latex rubber balloon of about 70 kPa pressure.

Figure 9.15



Activity 9.4 BOUNCING BALLS

EI

1 You are going to drop different balls on to a concrete floor from a height of 1.8 m and measure the height to which they bounce. Design a means of measuring the bounce height. Express their bounce height as a percentage of the drop height. This is called **restitution**. Some secondhand data are listed on the next page.

2 What characteristics do the higher bouncing balls have in common and how do they differ from the less bouncy balls?

3 Try inflating a volleyball to different pressures and comparing pressure versus bounce height. Is it linear? If not, why not?

4 Where does the energy go?

NOVEL CHALLENGE

The US baseball manufacturer Rawlings said they hadn't changed the manufacture of balls since 1931; however, it seems odd that balls from the 1970s bounced 157 cm when dropped from 462 cm (15 foot), whereas balls from the 1990s bounced 208 cm. What is intriguing is that in the 1970s players hit 61 home runs per season, whereas in the 1990 they hit 68 home runs. What is the interpretation?

Activity 9.5 DATA-LOGGING BALLS

EI

You can perform the activity illustrated in Figure 9.16 using a TI graphing calculator and a TI 'Ranger'. Other manufacturers (e.g. Casio) have similar equipment.

1 Hold the ball at least 50 cm away from the detector. When you are ready to start collecting data, press ENTER on the TI calculator to start the motion graph.

2 When the motion detector starts clicking, release the ball from rest and allow it to bounce up and down directly below the detector.

3 Analyse your graph to determine the maximum height.

You could also perform this experiment using the CBL and a Bounce program downloaded from the TI web page.

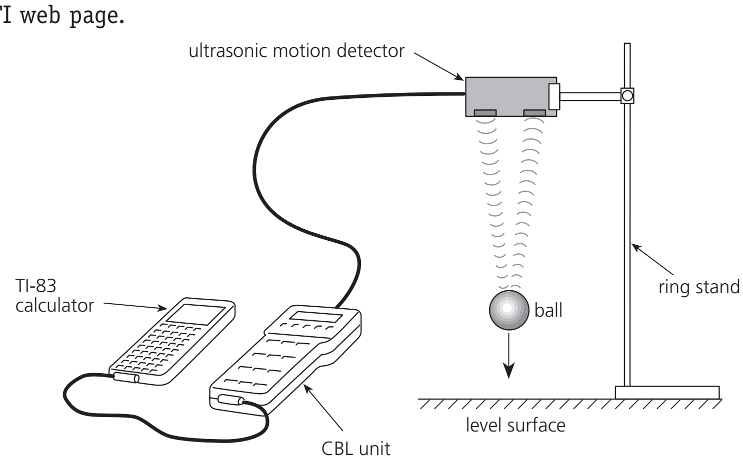


Figure 9.16 A motion detector and data-logger being used to measure bounce heights.

EI Activity 9.6 GETTING WARMER

In this activity you are to compare the bounce heights of a squash ball at different temperatures. Use a 2 m drop height to get a reasonable rebound.

1. Take three squash balls, put one in ice water (0°C), leave one at room temperature (measure), and put one in boiling water (100°C).
2. Compare the bounce height as before and account for your results.

In the 1965 baseball series, the Detroit Tigers accused the Chicago White Sox of illegally refrigerating the ball. The Tigers only scored 17 runs in 5 games using the cold balls whereas the White Sox accused the Tigers of cooking the balls to score 59 runs in the previous 5 games (including 19 home runs).

Although there is always some bounce associated with collisions, in this section we will deal with collisions that are almost totally inelastic for the sake of simplicity.

Example

A body of mass 6 kg travelling east at 4 m s^{-1} strikes a 2 kg mass at rest. After the collision they remain coupled and the mass moves east at 3 m s^{-1}. Is the collision elastic or inelastic?

Solution

$$\begin{aligned} E_K(\text{initial}) &= \tfrac{1}{2} m \boldsymbol{u}_A^2 + \tfrac{1}{2} m \boldsymbol{u}_B^2 \\ &= \tfrac{1}{2}\ 6 \times 4^2 + \tfrac{1}{2}\ 2 \times 0^2 \\ &= 48 \text{ J} \\ E_K(\text{final}) &= \tfrac{1}{2} m \boldsymbol{v}_{A+B}^2 \\ &= \tfrac{1}{2}(6 + 2) \times 3^2 \\ &= 36 \text{ J} \end{aligned}$$

As kinetic energy is lost, the collision is not elastic.

Questions

5 A 2 kg steel ball A travelling west at 5 m s^{-1} collides elastically head-on with a stationary ball B also of mass 2 kg. Without doing any calculations, state the velocities (including directions) of the two balls after collision.

6 A 0.20 kg ball A moving with a speed of 1.75 m s^{-1} approaches a stationary second ball B of mass 0.15 kg, head-on. After the collision, ball B travels at 2.0 m s^{-1} in the same direction. **(a)** Calculate the speed of ball A after the collision. **(b)** Was the collision elastic?

7 A 12 000 kg railway truck travelling along a straight track at a speed of 12 m s^{-1} collides with an identical stationary truck. If the trucks lock together as a result of the collision, calculate **(a)** their common speed; **(b)** the loss of kinetic energy.

9.6 POTENTIAL ENERGY

When you lift something off the floor and put it on a desk, you are applying a force to the object and displacing it — you are doing work on it. If work is done then it gains energy. If it is moved at constant speed then it is not gaining kinetic energy but gaining energy of position. This is called **gravitational potential energy** (Latin *potens* = 'capable'). It has the symbol E_P, *PE* or *U*. When you need to distinguish it from other forms of potential energy it is often written as GPE. In this book we will use E_P in formulas and PE or GPE as an abbreviation.

OUR RESULTS FOR ACTIVITY 9.6

If you need second-hand data or just want to compare, here are some typical results. The drop height was 100 cm.

Tennis ball	48%
Golf ball	60%
Table tennis	26%
Superball	83%

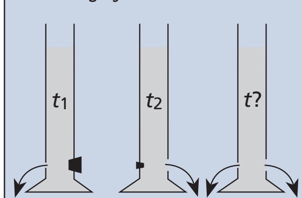
Note: the official standards for a basketball is a rebound height of 125.5 cm to 137 cm from a drop height of 182.9 cm; volleyball — 152.4 to 165 cm from 254 cm. Go and check the PE department's ball.

PHYSICS FACT

The 'restitution' (% energy returned after compression) values for animal bodies are as follows: kangaroo 40%, protein collagen 93%, resilin 97%. Resilin was discovered in 1960 by Danish scientist Dr Torkel Weis-Fogh. It works so well because it doesn't get hot. It is the natural polymer in flying insects; they use it for elastic storage in their wing hinges. This sounds like a good topic for an investigation.

NOVEL CHALLENGE

A plastic measuring cylinder has two stoppered holes near the base, one twice the diameter of the other. The cylinder is filled with water and the small stopper removed. It takes t_1 seconds to empty. When repeated with just the big stopper removed it takes t_2 seconds. *How long will it take with both stoppers removed? Do it algebraically first before you wreck a good measuring cylinder.*



NOVEL CHALLENGE

I live on a hill, and if I let my 2000 kg car roll down the incline, by the time it has travelled 120 m to the bottom it is going at 25 km h^{-1}. The start is 5 m higher than the finish. What average frictional force must be acting?

The work done is a measure of the change in gravitational potential energy. If the object is lifted at constant speed then the force applied equals its weight ($\mathbf{F}_w$) and the vertical distance is given the symbol for height *h*:

$$W = \mathbf{F}s = mgh$$

Hence a 5 kg ball raised 20 m will have work done on it or a change in potential energy equivalent to $5 \times 10 \times 20$ J = 1000 J. We assume that an object on the ground has zero GPE so the GPE of the ball is 1000 J. When the ground is the zero reference we can say that **objects raised gain GPE and objects falling lose GPE**.

Example 1

Calculate the GPE of a 20 kg box of groceries lifted 0.75 m to a bench top.

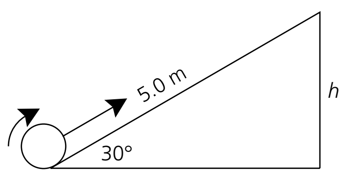
Solution

$$E_P = mgh = 20 \times 10 \times 0.75 = 150 \text{ J}$$

Example 2

A 35 kg beer keg is rolled up a 5 m long plank, which makes a 30° incline to the ground. What is the GPE of the keg at the top (Figure 9.17)?

Figure 9.17
Beer keg being rolled up incline.



Solution

A 30° incline with an hypotenuse of 5 m has a vertical height given by: 5.0 sin 30° = 2.5 m.

$$E_P = mgh = 35 \times 10 \times 2.5 = 875 \text{ J}$$

In summary, gravitational potential energy is defined as the energy associated with the state of separation between bodies that attract each other via the gravitational force. Mathematically, $E_P = mgh$. This of course will only be true over distances where the gravitational force remains constant. When you get too far away from the Earth's surface this relationship does not hold.

Questions

8 A brick of mass 2.5 kg is lifted to a height of 2.5 m above the ground by a bricklayer. Calculate **(a)** the GPE acquired by the brick; **(b)** the work done by the bricklayer in lifting it.

9 Assume that 1 kJ of work is done in lifting a 30 kg steel ball from the ground to the top of a tower. How high is the tower?

10 A 75 kg skier travels down a 40° slope a distance of 100 m. What is his change in potential energy?

NOVEL CHALLENGE

Have you ever tried to compare airconditioner specifications? It's most confusing. We know that 1 horsepower (HP) is equivalent to 0.746 kW, but we saw an airconditioner at school that said '2 HP (4.6 kW)'. What the …! The problem is that HP is the output of the motor, whereas kW is the output of the cooling system after the losses due to conversion of electrical energy to heat energy have been taken into account. The formula is: kW × motor efficiency = HP × 0.746. So what is the percentage efficiency of the school airconditioner?

POWER 9.7

A horse generates about 1 horsepower; a Corolla engine develops about 57 kilowatts of power. Which is the stronger? A horsepower (hp) is the old measure of **power** output where 1 horsepower equals 746 watts, so the car is more powerful. The word power comes from Anglo-Norman *poer* = 'ability to do things'. **Power is a measure of the rate of energy output** — it has the units **joules per second** (J s^{-1}). One J s^{-1} is called 1 **watt** (W) in honour of James Watt (1736–1819), a Scottish physicist who was the inventor of the first practical steam engine. One of his engines, the 26 hp Boulton and Watt, was restored in 1971 and is now in operation pumping water on the Kennet and Avon Canal in the UK.

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

Example 1

What is the power output of a cyclist who transforms 2.7×10^4 J of energy in 3.0 minutes?

Solution

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

$$= \frac{2.7 \times 10^4 \text{ J}}{180 \text{ s}} = 150 \text{ W}$$

Example 2

A 52 kg student runs up a flight of stairs of vertical height 3.0 m in 4.7 s. Calculate the power output.

Solution

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

$$= \frac{mgh}{t}$$

$$= \frac{52 \times 10 \times 3}{4.7}$$

$$= 330 \text{ W (about half a horsepower)}$$

NOVEL CHALLENGE

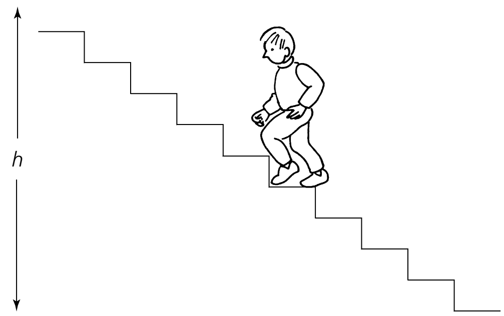
In 1894, bored British house-husband J. C. Ware cycled on a stationary bike for 16 hours and used 42 kJ. What was his average power output?

EI

Activity 9.7 YOUR PERSONAL POWER OUTPUT

1. Measure the vertical height of a flight of stairs. If you have access to a building with several storeys this is even better.
2. Use a stopwatch to time yourself running up the flight of stairs.
3. Measure your mass on some bathroom scales and calculate your power output.
4. If you use a multi-storey building, how does your power output over the first flight compare with the output over the last flight?

Figure 9.18



Typical results for Year 11 boys and girls is about 500 W. The Grand Rialto Stair Trek is a race held annually to see who can run up the steps of one of the tallest buildings in the world. The record is held by a 66 kg man who can run up the 122 steps (vertical height = 247 m) in 6 min 55 s. Show that his power output is 374 watts, or about a half-horsepower. The best result is for a 60 kg woman who did it in 7 min 58 s (304 W).

Power and velocity

The original definition of power was for the unit *horsepower*. It was defined as the rate of energy needed to lift a 550 pound weight at a speed of 1 foot per second upward. As weight is a measure of force, it implies that power is the product of force and velocity ($P = F\boldsymbol{v}$).

This can be shown mathematically:

$$P = \frac{W}{t} = \frac{\boldsymbol{F}\boldsymbol{s}}{t}$$

$$\text{As } \frac{\boldsymbol{s}}{t} = \boldsymbol{v}, \text{ then } P = \boldsymbol{F}\boldsymbol{v}$$

NOVEL CHALLENGE

Postulate 1: knowledge = power
Postulate 2: time = money
As every physicist knows:

$$\frac{\text{work}}{\text{time}} = \text{power}$$

Since knowledge = power, and time = money, we have:

$$\frac{\text{work}}{\text{money}} = \text{knowledge}$$

Solving for money, we get:

$$\frac{\text{work}}{\text{knowledge}} = \text{money}$$

Thus, as knowledge approaches zero, money approaches infinity regardless of the amount of work done.
Evaluate the conclusion:
'The less you know, the more you make'.

Example

An upward force of 6 kN is required to raise a mine cage vertically at a speed of 2.5 m s^{-1}. Calculate the power output of the motor lifting the mine cage.

Solution

$$P = \boldsymbol{Fv} = 6 \times 10^3 \times 2.5 = 15\,000 \text{ W}$$

SR

EI

Activity 9.8 POWER PACKED

1 *Car power* Use the *Guinness Book of Records* to find out the most powerful production car ever made. It seems a long time ago, doesn't it? Obtain a brochure or magazine with your most desirable car in it and note its power output. How many of them would be needed to equal the power of the world's most powerful car?

2 *Microwave power* Put 1 L of tap water in an icecream container, measure its temperature and place it in a microwave oven. Run for 1 minute on 'high'. Stir it, take the temperature again and calculate the power output by using the formula: $P = 70 \times \Delta T$ watts. Take note of the oven's power rating, usually stamped on a tag on the back. Calculate the efficiency of transferring electrical energy to heat energy in the water. Compare notes with others in your class. Which model is the most efficient? Where do the losses occur?

Questions

11 The world's most powerful windmill is the turbine built in Orkney (UK). Its blades are 60 m long and it was turned on in a hurricane in 1987. It now produces enough electricity for 2000 houses at an average of 1.5 kW each. Calculate the power output of the turbine.

12 Calculate the power involved in **(a)** lifting 100 kg on to a 1.2 m high bench in 2 s; **(b)** raising a 2.7 t Land Cruiser up 1.9 m on a hydraulic hoist in 15 s.

13 A 2200 kg Ford Falcon accelerates from 10 m s^{-1} to 15 m s^{-1} in 20 s. Calculate **(a)** the initial and final kinetic energy; **(b)** the work done; **(c)** the power output (in kW) assuming all engine energy goes to changing the kinetic energy of the car.

14 The engine of a jet aeroplane develops 2.5 MW of power when in level flight and travelling at a constant speed of 14 m s^{-1}. Calculate the force being developed by the engine (to overcome air resistance and generate lift).

15 James Watt defined the horsepower (hp) by measuring what weight of coal a work horse could raise up a mineshaft at a standard 1 foot per second. He found that it weighed 550 pounds. Show that a 550 pound weight being lifted vertically at 1 foot per second equals 746 W (1 hp), given: 1 foot = 0.3048 m, 1 pound = 0.4536 kg. Use $\boldsymbol{g}$ = 9.81 m s^{-2}.

CONSERVATION OF MECHANICAL ENERGY 9.8

Mechanical energy was earlier described as including both kinetic and potential energy. The total mechanical energy of a system can be defined as the sum of kinetic and potential energy. Within an isolated system this mechanical energy is conserved.

For example, when a ball is thrown upward in the air, it starts with a high kinetic energy and then at the top of its travel it has none. The kinetic energy has been transferred to gravitational potential energy. As it falls back to earth, it gains kinetic energy at the expense of potential energy. The word potential comes from the Latin *potens* meaning 'able'. In this sense, the object with potential energy is able to do work.

Consider a ball dropped from a high cliff (Figure 9.19). If the ball had a mass of 1 kg and the cliff was 100 m high, we can calculate the KE and GPE at any stage. The sum of the two has to equal the potential energy at the start. Table 9.2 sets this out:

Figure 9.19

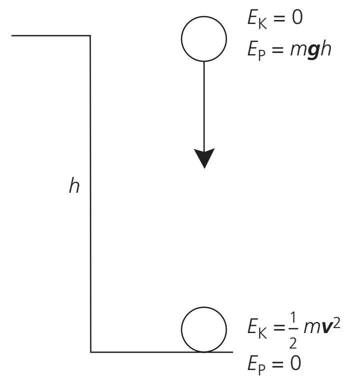


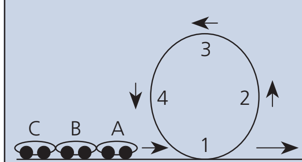
Table 9.2 CONSERVATION OF MECHANICAL ENERGY

HEIGHT (m)	POTENTIAL ENERGY mgh (J)	KINETIC ENERGY $\frac{1}{2}mv^2$ (J)	TOTAL PE + KE (J)	VELOCITY (m s^{-1})
100	1000	0	1000	0
75	750	250	1000	22
50	500	500	1000	32
25	250	750	1000	39
0	0	1000	1000	45

The velocities can be confirmed by using your vertical motion formulas from Chapter 2.

NOVEL CHALLENGE

A roller coaster is made up of three carriages A, B and C, which enter a loop as shown. Which position — 1, 2, 3 or 4 — will give the greatest speed for **(i)** carriage A, **(ii)** carriage B, **(iii)** carriage C?



Applications of energy transfers

The ability of mechanical energy to be converted into useful work has been known for thousands of years. Some examples of transfers are discussed below.

Amusement parks

If you want examples of physics principles and laws being put to use then there is no better place for an excursion than an amusement park. Newton's laws, conservation of momentum, centripetal forces, rotation and weightlessness are all there.

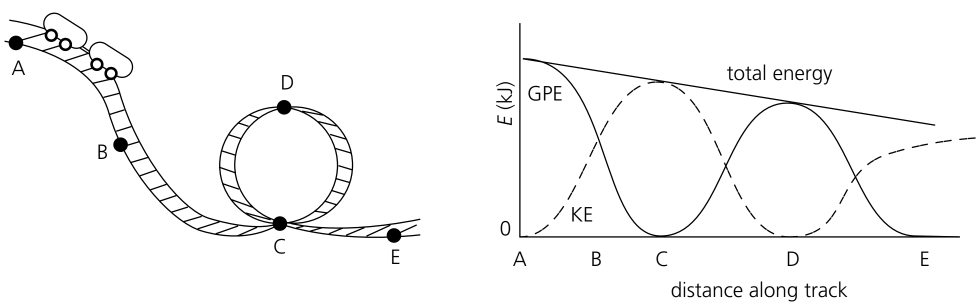


Figure 9.20
Energy changes during a roller coaster ride.

The roller coaster is a good example of conservation of mechanical energy. Electrical energy is used to propel the carriages to great heights, giving them high gravitational potential energy. At the point of release, the kinetic energy is almost zero but as they roll down the tracks, GPE is converted to KE and the carriages accelerate. At the next hill, some of the KE is converted back to GPE but some is transferred to thermal energy (by friction) and sound and can never rise to the previous height.

If there was no transfer to the structure, then mechanical energy would be conserved and the sum of GPE and KE would be constant. But there always is some loss and the designers take this into account. On rainy days, however, the losses become quite small as frictional losses are reduced and high speeds result. Sometimes the speeds are too high and the roller coaster has to be closed down.

Photo 9.1
The feeling of weightlessness on the Tower of Terror ride at Dreamworld.

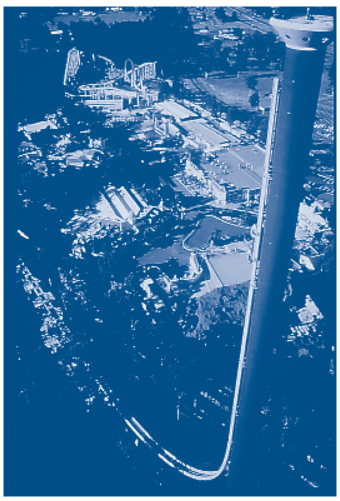


Photo 9.2
The tower at Dreamworld is used for the Tower of Terror ride and the Giant Drop.

Activity 9.9 A DAY AT THE FUN PARK

NEI

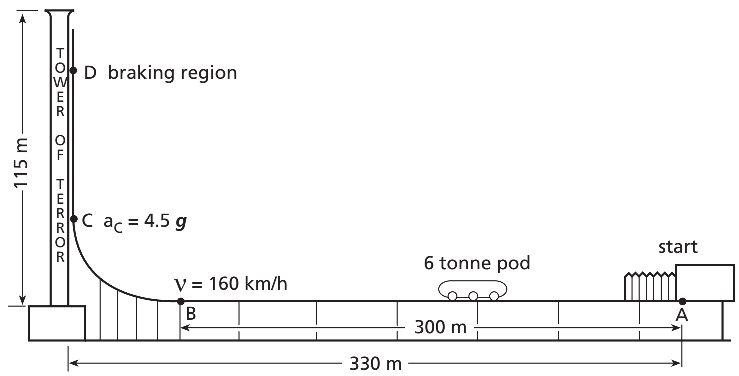
1. Next time you visit an amusement park, make a video of the roller coaster ride from the ground.
2. From the videotape, devise a way of measuring the speed of the carriage as it travels around the circuit.
3. Calculate the KE at various sections of the circuit (top of ramp, bottom of ramp, top of loop etc.). Plot these on a graph to show the changes. Estimate GPE at the same points and add these to your graph. What assumptions did you have to make?

Activity 9.10 DREAMWORLD'S 'TOWER OF TERROR'

SR

The 'Tower of Terror' is a 400 m track that stretches for 300 m horizontally before curving upward for 100 m (Figure 9.21). A 6 tonne pod with 16 people aboard (total mass about 7000 kg) is accelerated from rest (point A) to 160 km/h (at point B) along the horizontal section by electromagnets that draw 2.2 megawatts for 6 s. After this the pod goes unassisted into a vertical curve of radius 100 m, which gradually tightens to a curve of radius 50 m (point C) before travelling vertically for the last part of the trip (Photo 9.1). By this stage 12 s has elapsed (point D). In another 12 s the pod will be back to the start.

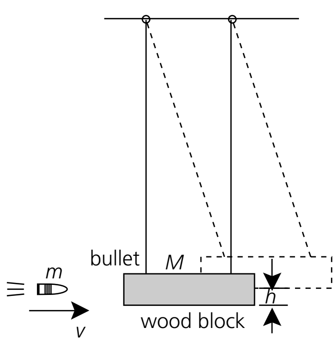
Figure 9.21
Tower of Terror dimensions.



1. If all of the kinetic energy is converted to GPE, to what height will the pod rise?
2. At the top of the curve (point C), the centripetal acceleration is 4.5 $\boldsymbol{g}$. Calculate the velocity at this point.
3. Why does the curve start with a radius of 100 m and then decrease to 50 m. Why not go straight into a 50 m curve?
4. Do the data suggest that the initial acceleration (from A to B) is uniform? Explain.
5. At what stage are you weightless? Explain.

The ballistic pendulum

Figure 9.22
A ballistic pendulum, formerly used to measure the speeds of rifle bullets.



A ballistic pendulum is a device that was used to measure the speeds of bullets before electronic timing was developed. The device consists of a large block of wood hanging by two long cords from the ceiling. When a bullet is fired into the wood, the bullet quickly comes to rest. The block with the embedded bullet swings upward to a maximum vertical height, which is measured. (Figure 9.22.)

We can let m be the mass of the bullet, and M the mass of the block. On collision, momentum must be conserved so if we assign $\boldsymbol{u}$ as the bullet's velocity and $\boldsymbol{v}$ as the velocity of the bullet + block after collision:

$$m\boldsymbol{u} = (m + M)\boldsymbol{v}$$

or

$$\boldsymbol{v} = \frac{m\boldsymbol{u}}{m + M}$$

Since the bullet and block stick together, the collision is perfectly inelastic. But mechanical energy is conserved and all of the kinetic energy is transferred to gravitational potential energy. So the kinetic energy of the block at the bottom of its arc must equal the potential energy at the top of its swing.

We can eliminate $\boldsymbol{v}$ from the equations:

$$\tfrac{1}{2}(m + M)\boldsymbol{v}^2 = (m + M)\boldsymbol{g}h$$

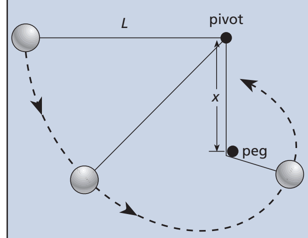
or

$$\boldsymbol{v}^2 = 2\boldsymbol{g}h;\ \boldsymbol{v} = \sqrt{2\boldsymbol{g}h}$$

NOVEL CHALLENGE

A pendulum bob on a string is allowed to fall but the string strikes a peg. If the length of L is 2.5x, the bob has zero velocity when it gets to the suspension point P.

Can you prove this mathematically?



Example

$$\frac{m\boldsymbol{u}}{m + M} = \sqrt{2\boldsymbol{g}h}$$

$$\boldsymbol{u} = \frac{m + M}{m}\sqrt{2\boldsymbol{g}h}$$

A bullet of mass m = 9.5 g was fired into a block of mass M = 5.4 kg and the block rose a vertical distance of 6.3 cm. Calculate the speed of the bullet.

Solution

$$\boldsymbol{u} = \frac{m + M}{m}\sqrt{2\boldsymbol{g}h}$$

$$= \frac{9.5 \times 10^{-3} + 5.4}{9.5 \times 10^{-3}}\sqrt{2 \times 9.8 \times 6.3 \times 10^{-2}}$$

$$= 630 \text{ m s}^{-1}$$

In practice, it is easier to tie a piece of cotton thread to the pendulum, and the sideways motion (the amplitude) is indicated by how far the piece of thread is dragged along the floor. By knowing how long the pendulum is, the vertical displacement h can be calculated.

INVESTIGATING

If you ever go on the Tower of Terror, take a small piece of bark and release it in front of your face as you start to go up.

What do you notice on the way up and down? Surprised huh? Explain that if you can!

Human energy conversions

Do you do any work to carry a bag while walking along a horizontal road? The force needed to carry the bag is vertical (the weight) but the displacement is horizontal (along the road) so no work is done. The bag has no additional potential energy or kinetic energy than it had to start with. But you have used energy and you would feel tired. Where has this energy gone?

A great deal of the energy has gone into internal energy transfers. Your body burns fuel to keep the muscles working. What is overlooked in this description is that the bag and your body move up and down as you walk. Your centre of gravity bobs like a bouncing ball rather than travelling smoothly like riding along on a bike.

As each foot strikes the ground your body drops down and then as you step forward your body rises again. This is shown in Figure 9.23.

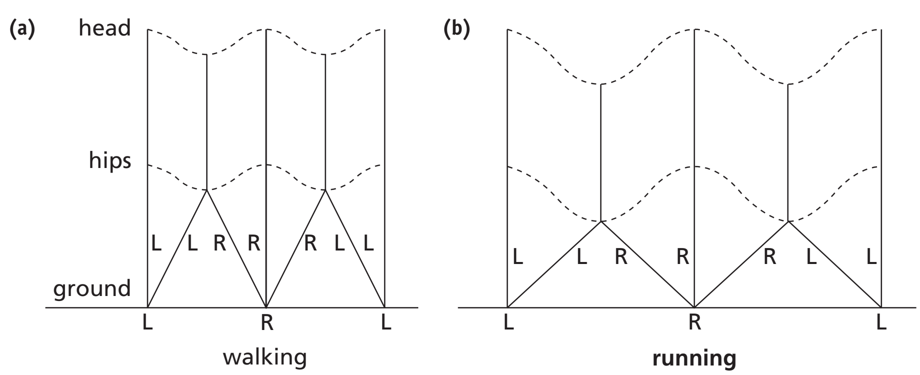


Figure 9.23
Note that the head and hips of a person walking move in a wave motion. They appear to move up and over the support leg and come down again as the foot touches the ground. In figure **(b)** the stride length is increased and the rise and fall of the wave increases.

The up and down motion is a change in potential energy. A typical vertical displacement is 3 cm per step.

Example

A 60 kg student is walking at 1 m s^{-1} with a stride of 50 cm. During each step his centre of mass rises and falls by 3 cm. Calculate **(a)** the work being done against gravity for each step; **(b)** the rate at which work is being done.

Solution

(a)

$$W = mgh$$
$$= 60 \times 10 \times 0.03$$
$$= 18 \text{ J (per step)}$$

(b)
- Each step is 50 cm so he makes two strides in one second.
- Each stride uses 18 J so the work done is 36 J in one second.

$$P = \frac{W}{t} = \frac{36}{1} = 36 \text{ J s}^{-1} = 36 \text{ W}$$

The up and down motion is the main use of energy in running and walking. Athletes try to minimise this vertical motion — just watch hurdlers in action and you will see how little the centre of gravity changes as they go over the hurdles.

Easy rollers — low friction tyres

While energy is expended when you walk or run because your body bobs up and down, you would expect very little energy loss with a rolling tyre. In this case there is no vertical motion of the centre of gravity. Unfortunately as a tyre travels along a road it undergoes two types of deformations — a macro-deformation and a micro-deformation. The first flattens the tyre tread against the road, creating a large 'footprint'. This is responsible for rolling resistance, which is unwanted, so the smaller the better. The second is a micro-deformation in which tiny irregularities in the road surface make imprints in the tread, and is responsible for the traction or friction between the surfaces. This is desirable so the bigger the better. A question you could ponder: do racing car 'slicks' maximise the micro-deformation and minimise the macro-deformation?

At 100 km h^{-1}, a tyre rotates about 20 times a second, with every part of the tyre flexing during each revolution. Most of the energy needed to cause flexing is transferred to heat within the tyre as the polymer chains slip over one another. If you feel a tyre after a trip, you'll note that the tyre is hot. Scientists are trying to develop polymer compounds that spring back into shape without generating heat. Michelin tyre scientists have found that by introducing silicon dioxide into the rubber, reductions of up to 40% in rolling resistance can be achieved. Such tyres are not in commercial production yet, but are expected to be much more expensive when they become available.

Stopping powerful locomotives

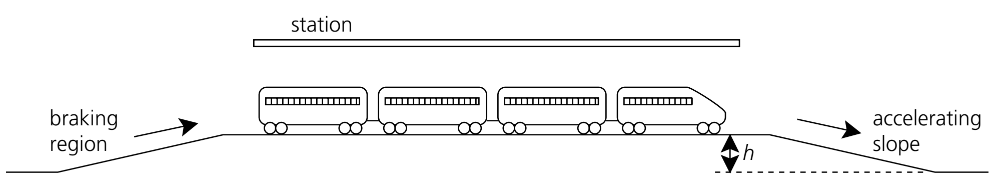


Figure 9.24
London Underground.

When deep sections of the London Underground railway were being constructed, the track at the stations was built at a higher level than the track between the stations (Figure 9.24). Trains approaching a station are decelerated by running up the slope, thus losing kinetic energy and gaining potential energy and at the same time reducing brake wear. Conversely, trains leaving the station are accelerated by running down the slope and saving on fuel. It all seems so logical you'd wonder why it was not thought of earlier. The system is ideal when trains stop at every station and is easy to construct when a new underground line is being built. It is not very practical on surface lines.

Underground trains can accelerate faster, too, not only if they have slope-assisted departures but because of the higher coefficient of friction. Being underground, the lines are protected from rain, dust and dirt so coefficients of friction of about 0.2 are typical, whereas on a good day above ground the coefficient is about 0.1 and can fall to 0.05 in wet weather. Acceleration is limited by the maximum frictional force that can be achieved (recall that the frictional force equals the coefficient of friction times the weight: $F_f = \mu F_N$). Trains used to have a sandbox located just in front of the drive wheels of the locomotive. On steep tracks sand was let out to increase the friction. Pretty ingenious, eh?

NOVEL CHALLENGE

Per kilogram, humans produce 10 000 times the energy produced by the Sun.
But which produces the greatest ***total*** *heat—the Sun or all the people on the Earth? Calculate the energy output of the Sun by estimating the population of the Earth, the average mass of a person and knowing that the average heat output of a person is 12 megajoules per day.*

NEI Activity 9.11 SOME MORE GUINNESS

Using the *Guinness Book of Records*, find where in the world the steepest railway is. The one at Katoomba NSW relies on a cable to pull it up the incline, but where is the steepest that uses wheel friction only?

Questions

16 Using conservation of mechanical energy principles, calculate the impact speed of a 65 kg diver who dives off a platform 8.5 m above the water.

17 Champion weightlifter Leonid Taranenko lifted 266 kg in a snatch and jerk to a height of 2.4 m. Calculate **(a)** how much work he did lifting this; **(b)** the impact speed when he dropped the weight on to the mat from this height.

18 A 0.41 g air-rifle pellet was fired into a 112.3 g toilet roll target suspended from the ceiling by a piece of thread. This ballistic pendulum rose 1.5 cm vertically. Calculate the muzzle velocity of the pellet.

9.9 ELASTIC POTENTIAL ENERGY

When an archer pulls back on the string of a bow, work is done and energy is stored. When the string is released, most of the stored energy is transferred to the arrow as kinetic energy. The bow stores energy as **elastic potential energy** (EPE).

Elastic potential energy is the energy stored in a spring or other elastic body by virtue of its distortion, or change in shape.

Examples include:
- rubber band on a speargun
- springs on a trampoline
- compressed gas in a nailgun
- flexible pole in a pole vault
- bungee rubber rope
- a polymer bumper bar on a car
- vehicle suspension springs or gas shock absorbers.

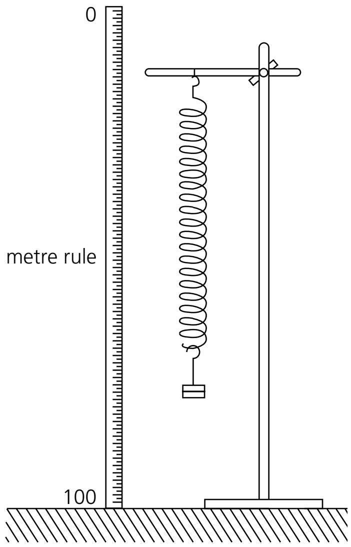
Photo 9.3
Pole vaulter. These photos were taken at evenly spaced time intervals. They show that the elastic potential energy stored in the bent pole is returned to the vaulter, to increase his gravitational potential energy, and his kinetic energy so that he can clear the bar.

Springs

When a spring is stretched (action) there is a restoring force (reaction), which tries to restore the spring back to its original length. The two forces are equal but opposite in direction.

Ch09-Walding 4th 25/8/04 11:38 AM Page 231

Figure 9.25
Hooke's law experiment.



Imagine a spring hanging vertically with a ruler beside it as shown in Figure 9.25. As masses are added the spring stretches under the applied force (the weight of the masses). A graph of force applied versus the extension of the spring is shown in Figure 9.26. The slope of the line (***F****x*) is called the **spring constant** (*k*) and in the graph shown, its value is 172 N m^{-1}. In Chapter 4 this was discussed in detail — you may remember the formula $\boldsymbol{F} = -kx$. The stiffer the spring the greater the spring constant. This relationship is called **Hooke's law** after Robert Hooke, an English scientist of the late 1600s. He didn't publish his findings in a journal as scientists do today. He wrote it in a letter as 'CEIIINOSSSTTUV', which can be rearranged into the Latin *ut tensio sic vis* meaning 'the force is proportional to the displacement'. He wanted to be credited with the law's discovery so he secured it by the time-honoured device of the anagram. That way he could claim to be first without letting anyone know the details.

Figure 9.26

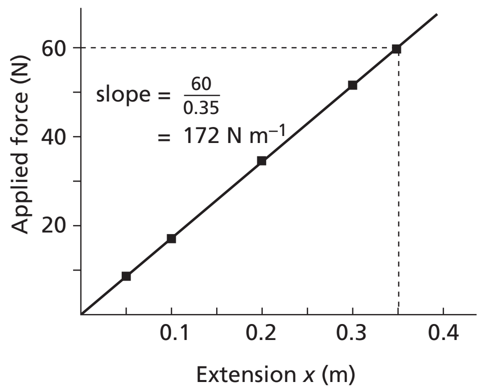
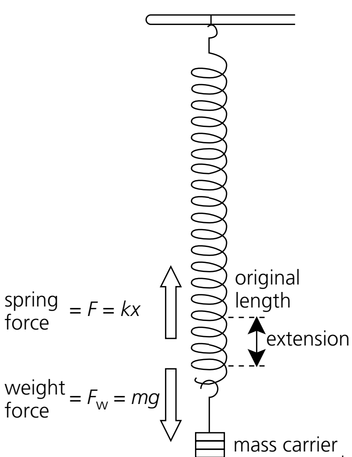


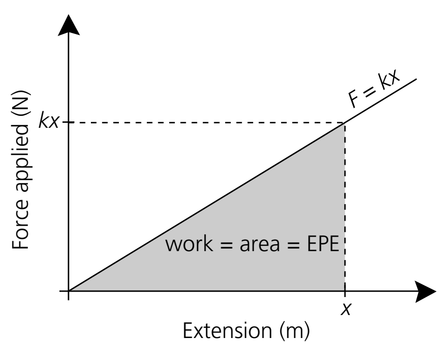
Figure 9.27
The forces acting on a stretched spring.



As the weight on a spring is increased, the spring stretches a little until it comes to equilibrium again. This is due to the fact that the spring exerts an equal but opposite force on the masses. This is shown in Figure 9.27. As the added force acts through a certain distance it implies that work is done. Where did the energy of this work go? It did not go into increased gravitational potential energy, nor did it go into increasing the kinetic energy of the spring. The energy went into stored elastic potential energy (EPE) of the spring.

In the graph shown in Figure 9.28, the area under the line is a measure of the work done, which equals EPE.

Figure 9.28



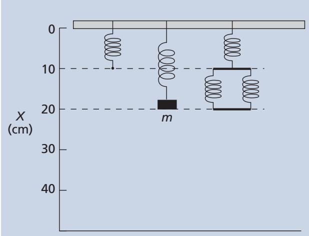
Area of triangle = $\frac{1}{2}$ base times height = $\frac{1}{2}x$ times $\boldsymbol{F}$ = $\frac{1}{2}x$ times kx.

Note: the symbols used here can be confusing. We will be using 'x' for displacement and '×' for multiplication ('times') to avoid confusion.

$$E_P \text{ or EPE} = \tfrac{1}{2}kx^2$$

NOVEL CHALLENGE

A spring has an unladen length of 10 cm. With a mass of 1 kg added it stretches to 20 cm. Three springs identical to the first one are arranged in the pattern shown in the diagram. *How far down will they stretch when 1 kg is added?*



Example 1

A spring with a spring constant of 250 N m^{-1} is stretched to a distance of 15 cm beyond its natural length. How much energy is stored in the spring?

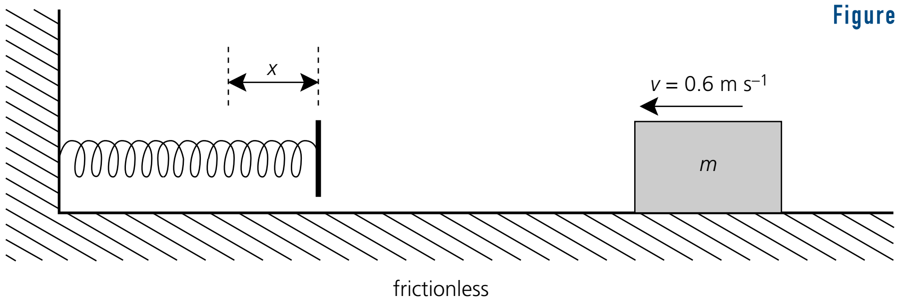
Solution

$$\text{EPE} = \tfrac{1}{2}kx^2 = \tfrac{1}{2}250 \times 0.15^2 = 2.8\ \text{J}$$

Example 2

A 200 g block of wood is travelling horizontally at 0.6 m s^{-1} and strikes a spring that has a spring constant of 15 N m^{-1} (Figure 9.29). Calculate the maximum compression of the spring.

Figure 9.29



Solution

Assume all of the KE is converted to EPE.

$$E_K = \tfrac{1}{2}m\boldsymbol{v}^2$$
$$E_P = \tfrac{1}{2}kx^2$$
Hence:
$$\tfrac{1}{2}m\boldsymbol{v}^2 = \tfrac{1}{2}kx^2$$
$$m\boldsymbol{v}^2 = kx^2$$
$$x = \sqrt{\frac{m\boldsymbol{v}^2}{k}} = 0.069\ \text{m}$$

INVESTIGATING

The spring constant is also called the elastic modulus. In maths, modulus is the number by which two given numbers can be divided and produce the same remainder. These meanings seem to be unrelated. Find out what the Latin *modus* means to explain this.

NOVEL CHALLENGE

A spring 20 cm long has a spring constant (modulus) of 86 N m^{-1}.
If the spring is cut into two 10 cm lengths, what will the modulus of each half be?

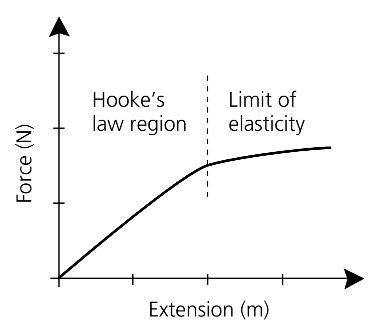
SR Activity 9.12 HOT LIPS

Hold a rubber band against your bottom lip. Stretch it tightly and note the temperature change. Let it go back to normal size and note how its temperature changes again. Can you explain this in terms of changes to the elastic potential energy and work being done?

Elastic limit

Many solids behave as if they are atoms or molecules linked together like a spring. When they are compressed or stretched by an external force they obey Hooke's law. However, every material has a limit to the amount of compressing, stretching or bending it can take. When Hooke's law is no longer obeyed, the object is said to have reached the **elastic limit** (Figure 9.30). A spring that has been stretched beyond the elastic limit will no longer return to its original length and its spring constant will change.

Figure 9.30
The limit of elasticity is eventually reached.



Questions

19 A catapult operates with the aid of a spring that has a spring constant of 265 N m^{-1}. Determine the amount of energy imparted to a rock if the spring is compressed a distance of 34 cm prior to release.

20 The graph in Figure 9.31 shows the extension of a spring when a force is applied to it. Using the graph, determine:
(a) the work done in stretching it to 20 cm; **(b)** the spring constant;
(c) the EPE stored in it when it is stretched 17 cm.

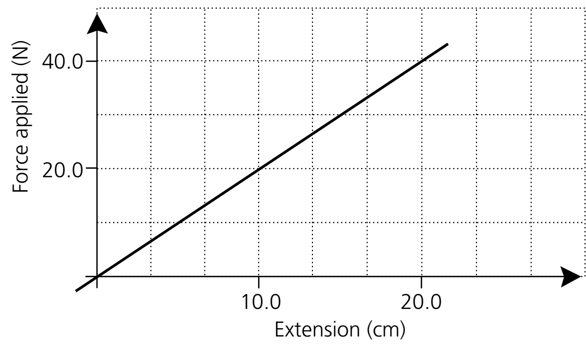


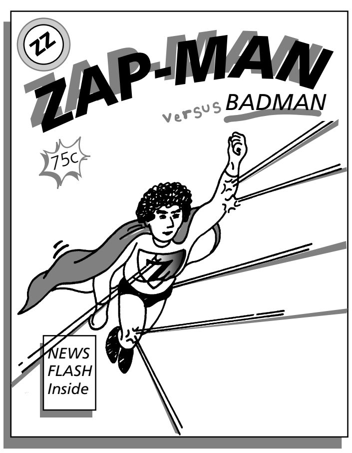
Figure 9.31
For question 20.

NOVEL CHALLENGE

A spring is 10 cm long. When a 0.6 kg mass is attached and allowed to fall the spring stretches by 0.76 cm. *Do you have enough information to calculate the spring constant?*

21 A 336 t locomotive is moving at 5 km h^{-1} when it bumps into a carriage that has a spring bumper. If the spring has a spring constant of 4×10^8 N m^{-1} calculate the maximum compression of the spring.

Figure 9.32
It is well known that bullets and other missiles fired at Superman will simply bounce off his chest.



Human springs

In a previous example it was shown that walking and running use up energy because of the bobbing up and down of the body. But not all of the energy used to raise the body is lost when it falls. Some is stored as elastic potential energy in the muscles, tendons and bones of the foot and leg. Researchers have found that the Achilles tendon in a runner's leg stretches by about 5% on impact and can return more than 90% of the absorbed energy to the muscles of the calf. The foot, by virtue of the arched bones and tendons, can also act like a spring. Deformation of the arch on impact stores the energy elastically and can return about 80% of it. A typical runner loses about 100 J of kinetic and gravitational potential energy at each step but because the foot and leg return a large proportion of this energy back on rebound, the runner effectively only loses about 50 J — the rest is returned.

Professional running tracks can increase this rebound energy transfer. Most are made of a rubber compound that feels quite springy and when a typical runner depresses the track about 7 mm or so, about 90% of this energy is returned. Figures show that a 3% increase in speed can be attained.

FINISHING UP WITH SUPERMAN 9.10

This chapter concludes separate treatment of mechanics and kinematics. However, knowledge, processes and reasoning developed in these nine chapters will be used over and over throughout the rest of your course.

Let's look at Superman to see how our knowledge and reasoning can put some myths to rest.

1 *Energy of the leap* Superman can 'leap tall buildings at a single bound' because he grew up on Krypton where gravity was so strong that the inhabitants needed superstrength to stand up. One early leap by Superman was described as covering an eighth of a mile (200 m). Using projectile formulas we can show that he would have to have a launch velocity of 160 km h^{-1}. That is indeed 'faster than a train', to quote an early Superman description. But as Superman matured, the description became 'faster than a speeding bullet', which as you may recall is about 990 m s^{-1} (3500 km h^{-1}). This corresponds to about 50 million joules of kinetic energy for a man of normal weight. Superstrength doesn't exempt you from the law of conservation of energy, and to gain that energy from food requires more than 20 Big Macs just to do it once. The conclusion: Clark Kent must eat like a pig.

2 *Force and acceleration* How much force must go into Superman's legs to reach 3500 km h^{-1} and thus 'leap a tall building in a single bound'? If he is pushing off against the ground he has to reach this speed before his feet leave the ground, a distance of say 50 cm. This corresponds to an acceleration of about 10^6 m s^{-2} (100 000 '*g*'). This means that his legs have to exert a force of about 10^8 N, that is, 100 000 times his own weight. This is the same as the weight of about 10 000 t of lead. The thrust of the world's most powerful rocket motor (the Russian 'NI' rocket booster) is only 4620 t. Superman could easily stop a speeding locomotive.

3 *Flying* In recent stories, Superman has been shown changing directions in mid-air. With a take-off speed of 3500 km h^{-1}, he could reach an altitude of 50 km in one leap. But Superman can't turn off air resistance. For a body of Superman's size, his terminal velocity is about 200 km h^{-1}. This means even if he survived a take-off blast that started him off at 3500 km h^{-1}, he would slow down and complete his flight at 200 km h^{-1}. Some high-speed skiers do a little better than that; they achieve a speed of 250 km h^{-1} by streamlining their bodies with skintight suits and special aerodynamic headgear.

4 *Vacuum travel* As a boy, Superman carried his earthling father to the Moon without the benefit of a pressurised suit — just a helmet. Even if the father's clothing didn't burn up from air friction at lift-off, it wouldn't stop his dear dad's blood from boiling in the vacuum of space. And as Superman is not rocket-propelled, how can he change directions if he has no air to push against? Surely he doesn't violate the law of conservation of momentum! What provides the centripetal force that enables him to travel in a circle? One possibility for his rocket thrust is his superbreath. He might just blow his superbreath out in front of him and thereby be pushed backward. Some diapraghm he must have. Imagine its spring constant value!

5 *Light speed* Superman's orbital flights pose yet another problem. He is known to circle the Earth at 'seven times per second'. This corresponds to the speed of light. Using centripetal force formulas, it can be shown that to stay in a low earth orbit at this speed he would have to develop about a billion tonnes of thrust (10^{13} N). Chapter 30 (Relativity) has further problems examining travel at speeds close to that of light. You'll soon see that Superman knows little about senior physics. A 'very limited achievement' for him.

Maybe we should look at the Terminator, the Killer Tomatoes, Cinderella, Peter Pan, the Easter Bunny and the Tooth Fairy. No fantasy is safe from a physicist.

Practice questions

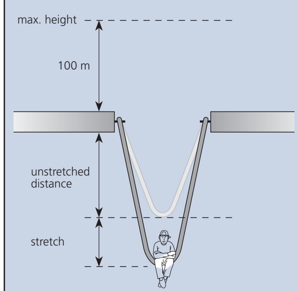
The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

*22 A light bulb consumes 60 J of electrical energy per second but only converts this to 18 J of light energy.
(a) What is the efficiency of the bulb?
(b) Where does the remaining 42 J go — is it lost?

NOVEL CHALLENGE

A 'reverse bungee' is a long elastic cord hanging in a 'V' configuration (as shown in the diagram). It is stretched and held in position by an electromagnet before being released. It rises and goes 100 m above the top support. Prove that the elastic 'spring' constant equals 2×10^3 N m^{-1}.



INVESTIGATION

This is a good extended experimental investigation. What factors influence the range of an arrow shot from a bow (angle of elevation, 'draw')? Experienced archers shoot feathers instead of plastic vanes (fletches) because they argue that feathers result in higher arrow velocities, greater stability, better guidance, higher accuracy and more forgiving flight. Here's their argument. First, feathers are faster because feathers weigh much less than plastic vanes. This means less mass to accelerate and less energy wasted. Feathers typically save 40 grains (2.6 g) over plastic. This is a lot of surplus mass — 30% of a typical 125 grain (8.1 g) steel head. Second, feathers produce less friction as they travel over the arrow rest or other bow parts. Less friction means higher speed. Third, the superior guidance of feathers prevents yawing and fishtailing of the arrows. Yawing and fishtailing add drag and slow arrow speed. What do you make of these arguments?

***23** A force of 90 N is applied to a string attached to a sled that makes an angle of 35.0° to the floor. How much work is done on the sled in dragging it across the lawn a distance of 10.0 m?

***24** A car is pulling a loaded trailer in a easterly direction along a horizontal road at constant speed. A force–displacement graph is shown in Figure 9.33. Calculate the work done by the car.

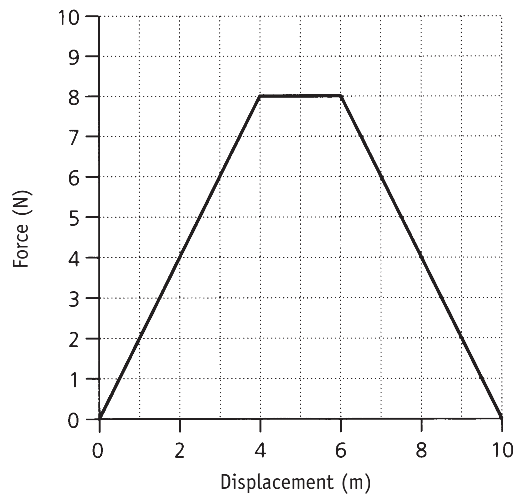


Figure 9.33
For question 24.

***25** Calculate the kinetic energy of a 60 g air hockey puck sliding at 8.0 m s^{-1}.

***26** What amount of energy is consumed by a 2000 W jug that took 1 minute 45 seconds to boil a cup of water?

***27** A body of mass 10 kg travelling east at 4 m s^{-1} strikes a 3 kg mass at rest. After the collision they remain coupled and the mass moves east at 3 m s^{-1}. Is the collision elastic or inelastic?

***28** Calculate the GPE of a 50 kg bag of cement lifted 1.4 m from the ground to a mixer bowl.

***29** A 65 kg student runs up a 12.0 m flight of stairs in 14 s. Calculate his power output.

***30** Using conservation of mechanical energy principles, calculate the impact speed of a 2 kg rock dropped off a cliff 8.5 m above the water.

***31** A spring with a spring constant of 150 N m^{-1} is stretched to a distance of 25 cm beyond its natural length. How much energy is stored in the spring?

***32** The graph in Figure 9.34 shows the extension of a spring when masses are added to it. Using the graph, determine:

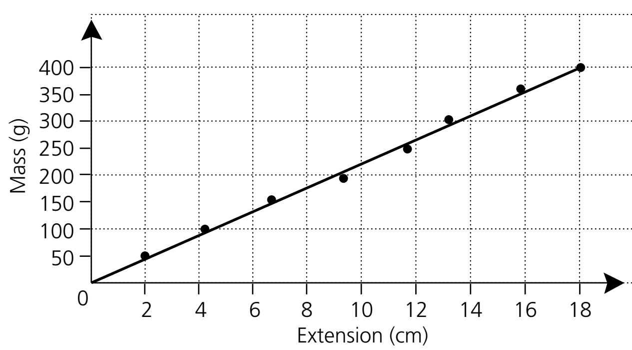


Figure 9.34
For question 32.

(a) the work done in stretching it to 18 cm;
(b) the spring constant;
(c) the EPE stored in it when it is stretched to **(i)** 10 cm; **(ii)** 20 cm.

****33** In Figure 9.35, a sequence of drawings shows a high jumper in action. The changes in mechanical energy during the high jump (Fosbury flop) are shown in Figure 9.36.
(a) Match the numbered drawings with the three phases of the jump.
(b) The total energy at the peak height appears to be greater than the initial energy. If energy is conserved, how can there be energy created as the graph suggests?
(c) Shouldn't potential energy start at zero in the graph? After all, they do start from ground level!
(d) Extrapolate the graphs (on your own paper, not in this book) to show how the curves might look after the jumper clears the bar and lands flat on her back on the landing pad. Give an appropriate name to this phase.
(e) Would the GPE on landing be the same as at take-off? Explain.

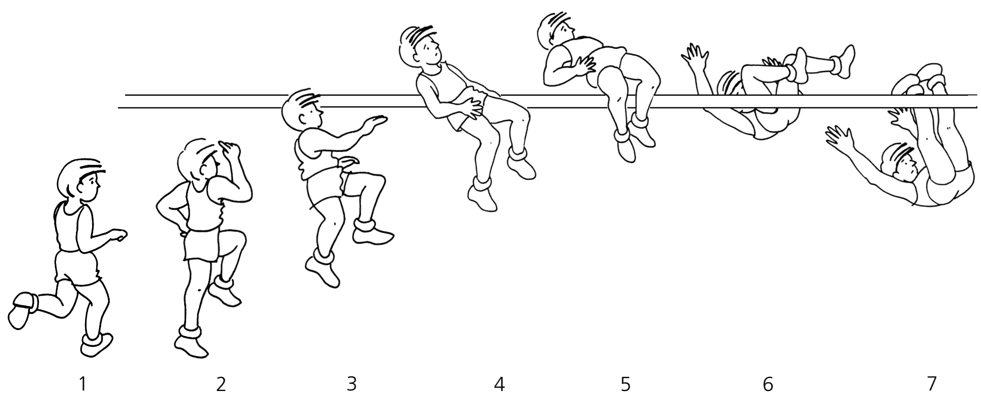


Figure 9.35
Stages of a Fosbury flop.

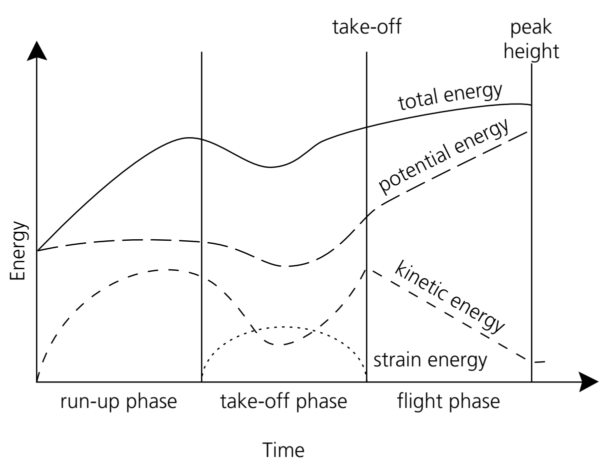


Figure 9.36
Energy changes in a Fosbury flop.

Extension — complex, challenging and novel

*****34** A 1.6 kg ball collides with a 2.4 kg ball as shown in Figure 9.37. After the collision the balls continue to travel, as shown in the diagram.
(a) What is the velocity ***v***?
(b) Is the collision elastic or inelastic? Show your proof.
(c) If the velocity of the 2.4 kg ball was in the opposite direction initially, could the velocity of the 1.6 kg ball after the collision be in the direction shown in the figure?

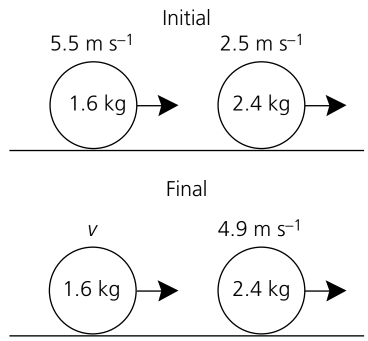


Figure 9.37
For question 34.

*****35** In a ballistic pendulum experiment, a bullet of mass 4.5 g was fired horizontally into a block of mass 3.4 kg suspended by a string 2.0 m long. The block and embedded bullet moved sideways a distance of 53.5 cm. Prove that the vertical displacement is 7.3 cm and then calculate the speed of the bullet.

*****36** Approximately 5.5×10^6 kg of water drops 50 m over Niagara Falls every second. If all of the water's potential energy could be converted to electricity, how much money could the owners get if it was sold at the industrial rate of 2 cents per megajoule?

*****37** A Toyota *Camry* uses about 11.5 L of petrol per 100 km.
- **(a)** If petrol provides 31 MJ per litre, how far could you travel on 1 MJ of energy consumed?
- **(b)** If you are driving at 60 km h^{-1}, at what rate would you be consuming energy (in watts)?

*****38** A 0.63 kg ball is thrown straight up into the air with an initial speed of 14 m s^{-1} and reaches a maximum height of 8.1 m before falling back down again. Assuming that the only forces acting on the ball are the ball's weight and air drag, calculate the work done on the ball during the ascent by the air drag.

Photo 9.4
Jimmy goes bungee jumping.

*****39** American advertising agency McCann decided to bungee jump a GMC Jimmy Sports Utility off a West Virginian bridge to show how well the truck was made. They used a 30 m rubber bungee cord used by the US army for supporting tanks during air drops. It had nine individual cords, each about 12 cm diameter. When released from the top of the bridge, 267 m above the river, the 1587 kg ute stretched the cord to six times its length before returning upward for a few more bounces. Neglecting air resistance, are you able to calculate the spring constant *k* of the bungee cord? If so, what is it? If not, what other information do you need?

*****40** A tough one. A body of mass 2.0 kg makes an elastic collision with another body at rest and continues to move in the same direction but with one-fourth of its original speed. What is the mass of the struck body?

*****41** A London Underground train consisting of a locomotive and 30 carriages with a total mass of 1586 tonnes travelling at 28 km h^{-1} is slowed by travelling up a slight incline to the horizontal platform. If it is raised vertically by 1.2 m over a distance of 300 m, calculate the speed of the train at the top of the incline, assuming no braking takes place.

*****42** The spring of a spring gun is compressed a distance d or 3.2 cm from a relaxed state, and a ball of mass m = 12 g is placed in the barrel. If the spring constant is 292 N m^{-1}, with what speed will the ball leave the barrel once the gun is fired? Assume no friction and a horizontal gun barrel.

*****43** The Greek historian Herodotus said it took 100 000 men 23 years to build the Great Pyramid at Giza but we don't believe him. The total GPE of the pyramid can be found by the formula GPE = $h^2\, d\, s^2\, \boldsymbol{g}/12$ where h = the height (146.7 m), d = density (2700 kg/m^3) and s = the side (230.4 m). But you have to add on the GPE of the stone blocks when they were raised the 19 m from the quarry to the base of the pyramid (GPE = $m\boldsymbol{g}h$). The mass can be found by multiplying the density by the volume ($V = s^2\, h/3$). If an Egyptian man can generate 160 kJ of energy per day, show that the average number of men over the 23 years was only 2845 and not the 100 000 Heroclotus claimed.

*****44** Have you ever been in a high-rise building that sways in the wind? Tall buildings oscillate with periods between 0.5 s and 10 s. To reduce the amplitude of the sway, engineers place 'tuned dynamic dampers' on the roof (see Figure 9.38). These are large blocks of concrete attached by springs to the side of the building, and can slide from side to side on a film of oil. A typical spring has a spring constant of 50 000 N m^{-1}. What mass of concrete would be needed to damp the oscillations of a building with the most sickening period of 5 seconds? Engineers use the formula:

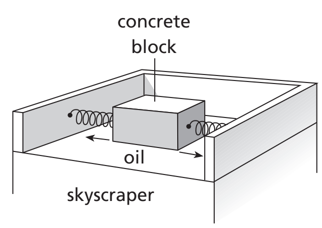
$$T = 2\pi\sqrt{\frac{m}{k}}$$

*****45** Imagine a block of wood held at the top of an inclined plane. At the bottom is a spring, attached to the plane. Write a question based on this and make up any data that may be required for the question. In fact, make up more data than is required. Write out a solution and ask a colleague (or even your teacher) to solve it.

TEST YOUR UNDERSTANDING

(Answer true or false)
- Energy gets used up or runs out.
- Something not moving can't have energy.
- A force acting on an object does work even if the object doesn't move.

Figure 9.38
For question 34.



Ch10-Walding 4th 25/8/04 11:43 AM Page 239

UNIT 04

Thermal Physics

CHAPTER 10

Heat and Temperature

10.1 HEAT AND HUMANS

Heat, and the lack of heat, have been important to humankind since the earliest times. Heat from the Sun and the cold of winter have always affected living conditions. From early childhood, hotness and coldness are two of the first feelings children encounter. Children experience these conditions early in life by sucking an iceblock or walking on a hot road. In fact it is the temperature change that initiates breathing when a child is born. The industrial revolution was based on heat and the generation of mechanical power from heat. Many industries in modern-day life are solely centred around the production of heat or the purposeful removal of heat, for example refrigeration, airconditioners and pot-belly stoves.

NEI Activity 10.1 HOT SPOTS

People have always wondered about the extremes of temperature. Consult the *Guinness Book of Records* or the Internet to find:

- the highest and lowest recorded human body temperature
- the places on earth that recorded the highest and lowest temperature
- the highest and lowest temperature ever achieved on Earth.

But what is heat?

10.2 HEAT AND TEMPERATURE

Up to the eighteenth century heat was regarded as some sort of invisible fluid, a 'caloric fluid' that bodies possessed. Hot bodies, it was believed, contained more of this fluid than cold bodies. When a body was warmed this caloric fluid was added to the body. But this did not explain why two ice cubes melted when they were rubbed together.

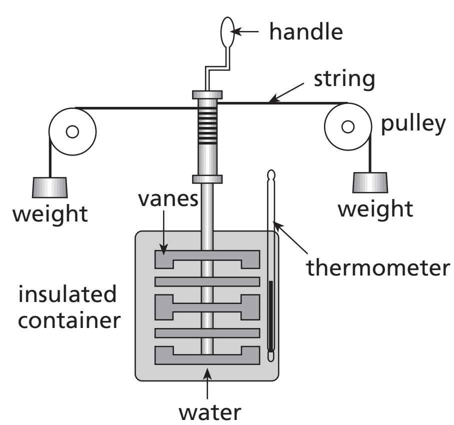


Figure 10.1
A schematic diagram of Joule's apparatus used to investigate the relationship between mechanical and thermal energy.

Figure 10.2
The energy contained in a ball in flight due to its motion and height.

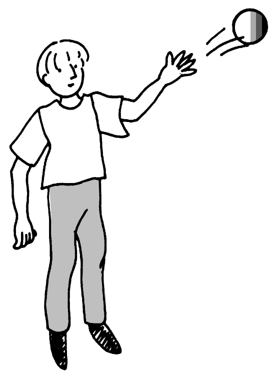


Figure 10.3
The energy contained in a box of bees, including motion of the bees.

English scientist **James Joule** (1818–89) was one of the first scientists to show that heat was a form of energy. He performed an experiment in which falling lead weights turned paddles in water. The work done by these weights caused the water to heat up. He showed that mechanical energy can be converted into heat. Joule concluded that heat is a form of energy, but what form does it take?

Consider a student throwing a ball. (See Figure 10.2.)

What energy does the ball possess?

Some might say it has kinetic energy because of its motion. Others might say it has potential energy due to its height above the ground. Some would say both. Is that all the energy it possesses?

Now consider the same student throwing a box full of bees (Figure 10.3). What is the total energy of the container?

It is easily seen that the container possesses both kinetic energy due to the motion of the box, and potential energy due to its height above the ground. However, it also possesses the kinetic energy the bees themselves might have.

Up to this time we have considered the bulk energy of objects and not the internal energy the particles in the objects might possess. All objects contain particles, atoms and/or molecules, and it is the motion of these particles that makes up the internal energy of the object. This motion can be vibrational kinetic energy, as in solids; or rotational and translational kinetic energy, as in fluids. The particles of matter also possess many forms of potential energy in the bonds that hold particles together, as well as that stored in the nucleus of the atoms. (See Figures 10.4 **(a)**, **(b)** and **(c)**.)

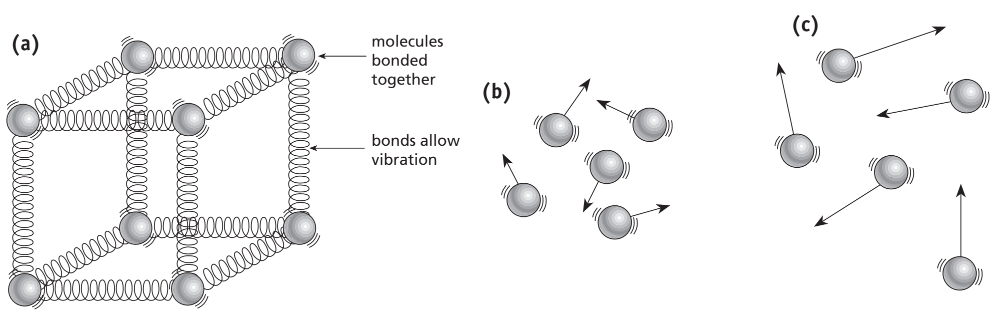


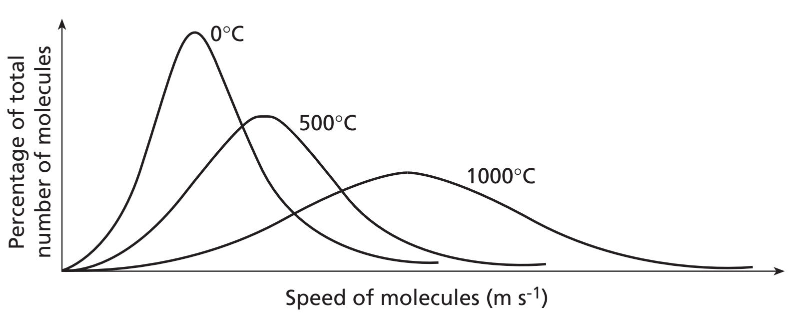
Figure 10.4
(a) A model of the movement of the molecules in a solid. The springs represent the molecular bonds. **(b)** A representation of the molecules of a liquid. They vibrate and are able to flow over one another. **(c)** The molecules of a gas have greater freedom. They have rapid straight line motion.

The sum of the kinetic and potential energies of all the particles is called the **internal** or **thermal energy** of the object. Heating is the term used when some of the thermal energy is transferred from hot objects to cold objects as in the case of a hot spoon being placed in cool water. The term heat is used to describe the internal energy transferred through this heating process. The study of these energy transfers is called **thermodynamics** (from the Greek *thermos* meaning 'heat', and *dynamis* meaning 'powerful').

It would be impossible to measure the motion of all the particles within a substance because of the number of particles and the great variation in speeds of the particles. Figure 10.5 indicates the variation of molecular speeds of a gas at various temperatures.

However, as objects gain heat and become hotter, the particles move faster. Temperature is a measure of the average kinetic energy of the particles of the object. Changing the potential energy of a substance without changing the average kinetic energy of its molecules does not change the temperature of the substance. This occurs when a change of state occurs. That is, when a substance changes from a solid to a liquid, or a liquid to a gas, or vice versa. A common misconception is that heat and temperature are the same, which is not the case.

Figure 10.5
The distribution of speeds of the molecules of a gas at various temperatures.



Ch10-Walding 4th 25/8/04 11:43 AM Page 242

10.3 THERMAL EQUILIBRIUM

Everyone has observed that when a cool spoon is placed in a hot cup of coffee it eventually becomes hot — as hot as the coffee. Or when you use a thermometer to measure a person's temperature the thermometer becomes as warm as the person whose temperature is being measured. How does this occur?

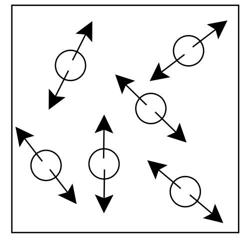
Consider a closed system (a system where there are no energy losses to the environment) in which a hot object A is in contact with a cooler one B. (See Figure 10.6.)

Because A is hot it contains more thermal energy than B, and its molecules have more potential and kinetic energy. The molecules move faster in object A than they do in object B. When A and B are placed in contact the molecules of A collide with the molecules of B, transferring kinetic energy to them. This causes the molecules of B to vibrate further apart, thus increasing object B's potential energy. Object B's thermal energy has increased. At the same time the molecules of A have slowed down and vibrate closer together, thus decreasing A's kinetic and potential energies. Object A has lost thermal energy. Thermal equilibrium is reached when the energy given to B equals the energy B is giving back to A. As the law of conservation of energy is true for all forms of energy:

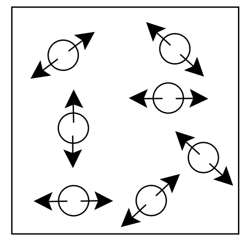
heat lost by object A = heat gained by object B

So when you use a thermometer to measure a child's temperature, molecules of the child in contact with the thermometer jostle the molecules of the glass, which in turn jostle the molecules of the mercury in the thermometer. The mercury expands, indicating the temperature on an appropriate scale of temperature.

Figure 10.6
The flow of thermal energy from the molecules of a hot to those of a cold body.



A



B

Activity 10.2 FREEZING

EI

NEI

1 Find out which freezes first — a cup of hot water or a cup of cool water placed in the freezer.

2 Do pets have the same body temperature as humans?

Questions

1 Which has more thermal energy: a cup of water at 100°C or a bath full of water at 40°C? Why?

2 Steam at 100°C will give you much more severe burns than water at 100°C.
(a) In which one are the molecules moving the faster?
(b) In which one do the molecules have greater potential energy?
(c) Why are steam burns more severe?

3 **(a)** What is the name given to the internal energy of a substance?
(b) What form(s) of energy does this involve?

4 If James Joule did 100 J of work on several quantities of water — 100 mL, 300 mL and 500 mL:
(a) which sample would gain the most thermal energy? Why?
(b) which sample's temperature would increase the most? Why?

NOVEL CHALLENGE

During the Second World War, Nazi scientists threw many prisoners overboard into the freezing waters of the North Sea to see how fast their body temperature dropped and how long it would take for them to die. Today, such data are needed by ocean rescue researchers to help to develop safety devices for ocean users.
Should we use this despicable 'Nazi science'? Develop an argument for or against its use.

NOVEL CHALLENGE

The table below shows the effects of changes to body temperature:

T (°C)	*Effect*
37.0 ± 1	normal oral
35	shivering
34	slurred speech
33	hallucinations
32	shivering stops
30	unconsciousness
26	appears dead

Death is defined as a failure to revive on rewarming above 32°C. When people freeze to death in cold water it has been reported that they do not seem to be in pain as they die. They often seem relaxed. *What could be happening here?*

10.4 MEASURING TEMPERATURE

Measuring temperature requires the use of some property of a substance that changes proportionally with increase in temperature. The property that most temperature measuring instruments use is expansion and contraction. This is the property used in most thermometers. The most common is mercury-in-glass or alcohol-in-glass thermometers that have

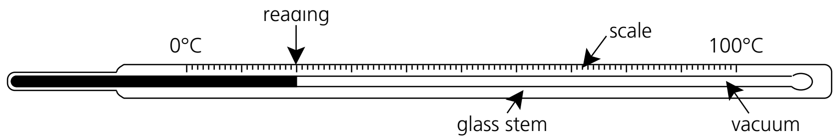


Figure 10.7
The liquid-in-glass thermometer.

been calibrated to indicate the temperature. As temperature increases, the mercury or alcohol expands up a fine tube in the glass thermometer. The markings on the thermometer depend on the scale used. (See Figure 10.7.)

Throughout history scientists made up their own scales to measure temperature. Sir Isaac Newton made up a temperature scale where the freezing point of water was 0 and normal body temperature was 12.

NOVEL CHALLENGE

At what temperature will °C and °F readings be the same? Could the Kelvin temperature reading ever be the same as °C or °F reading?

The Fahrenheit scale

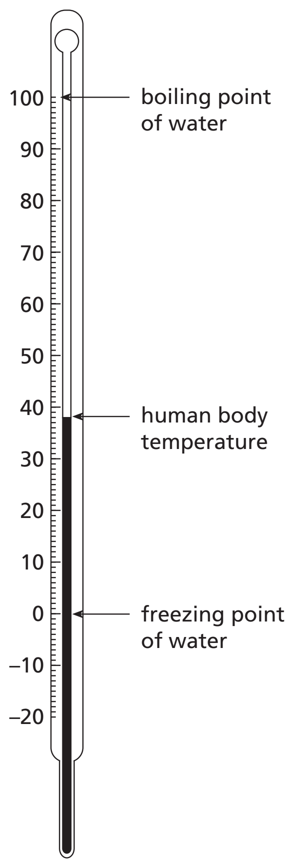
A German physicist, **Gabriel Fahrenheit** (1686–1736), developed a liquid-in-glass thermometer and a temperature scale that took the freezing point of an ice and salt mixture to be 0°F. He took the freezing point of pure water as 32°F and normal body temperature to be 96°F. The boiling point of water is then 212°F. This scale is no longer used in Australia but is still in use in several other countries such as the USA, UK and Canada. The conversion is $°C = (°F - 32) \times \frac{5}{9}$.

The Celsius scale

An easier decimal scale was invented by a Swedish astronomer, **Anders Celsius** (1701–44). On the Celsius scale, also called the centigrade scale, the freezing point of pure water is 0°C and the boiling point is 100°C. Interestingly, he originally took the freezing point to be 100°C and boiling point to be 0°C, but this was changed in the first year. This is the main scale used in measuring body temperature. (See Figure 10.8.)

Did you know that the body temperature of a baby is higher than that of an adult? Did you know also that aspirin is used to lower the body temperature?

Figure 10.8
A Celsius thermometer, a common thermometer used in the laboratory and in the home.



The Kelvin/absolute temperature scale

The two previous scales are relative scales. That is, zero degree on either scale does not mean that this is the lowest temperature obtainable. Since temperature is a measure of the average kinetic energy of the particles, 0°C does not mean that all particle motion has stopped. Then at what temperature does all motion stop? This point would be the true limit of coldness and would produce an absolute zero temperature. **Lord Kelvin** (1824–1907) suggested this temperature was –273.15°C.

When a sample of gas of constant volume is heated its pressure varies with Celsius temperature, as shown in Figure 10.9. Extrapolation of this graph suggests that, at –273.15°C, the pressure becomes zero and therefore all particle motion stops. This is because pressure is caused by particles colliding with the container walls and if there is no motion there are no collisions and therefore no pressure. This point is called absolute zero on the Kelvin scale of temperature. However, one degree on the Kelvin scale is equal in magnitude to one degree on the Celsius scale.

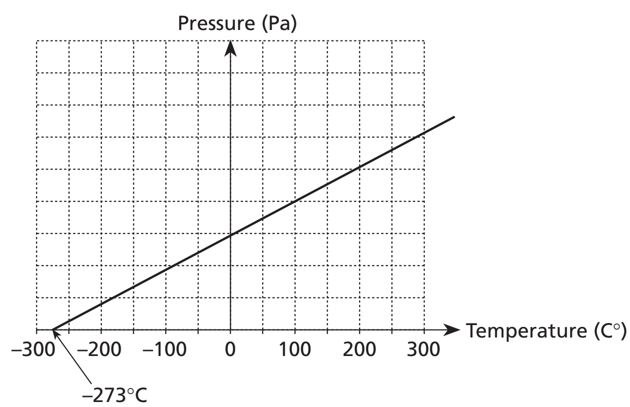
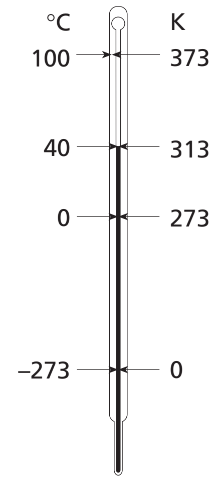


Figure 10.9
The relationship between temperature and pressure, and the establishment of absolute zero.

Therefore changing Celsius temperature to Kelvin temperature simply requires the addition of 273 (to three significant figures) to the Celsius value. Figure 10.10 shows a comparison between the two scales.

Kelvin temperature = Celsius temperature + 273
K = °C + 273

Figure 10.10
A comparison of temperature in degrees Celsius and Kelvin.



Example 1
Convert 50°C to kelvins.

Solution

$$\begin{aligned} K &= °C + 273 \\ &= 50°C + 273 \\ &= 323\ K \end{aligned}$$

Example 2
Convert 486 K to °C.

Solution

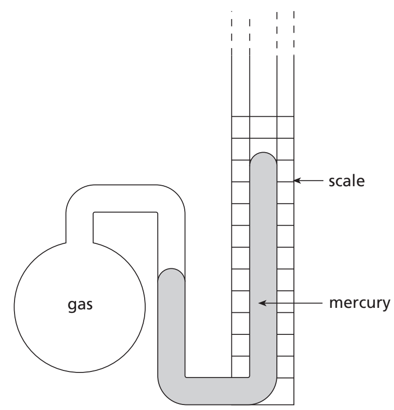
$$\begin{aligned} K &= °C + 273 \\ 486 &= °C + 273 \\ 486 - 273 &= °C \\ &= 213°C \end{aligned}$$

Questions

5 Convert the following temperatures to K: **(a)** 20°C; **(b)** –150°C; **(c)** 520°C; **(d)** –72°C; **(e)** –300°C.

6 Convert the following temperatures to °C: **(a)** 50 K; **(b)** 278 K; **(c)** 1000 K; **(d)** –50 K.

Figure 10.11
A constant volume gas thermometer — a thermometer that relies on the relationship between temperature and pressure.



Other types of thermometers

Even though the liquid-in-glass thermometers are the most widely used in science and in general, they have their limitations. This is mainly due to the liquid freezing or boiling. Alcohol-in-glass thermometers can be used between –100°C and 80°C. Mercury-in-glass thermometers have an operating range of –40°C to 360°C. Glass is also fragile, and mercury is toxic to the body and the environment.

Gas thermometers rely on the expansion of gas. Since change in temperature is proportional to the change in volume of a gas, the expansion of a gas can be calibrated to measure temperature.

Resistance thermometers use the fact that electric current in wires decreases as temperature rises.

Thermocouples consist of two wires made of different metals. The wires are made into a loop including a voltmeter. One end is kept at a reference temperature and the other end is used as a probe. (See Figure 10.12.) When this probe is placed in a substance to be measured the voltage produced is proportional to the difference in temperature between the two ends. Thermocouples can be used to measure temperature over a wide range.

Bimetallic strips rely on the different expansion rates of two different metals. When heated one metal expands more than the other, causing bending and movement of a pointer across a scale. These have a wide working range.

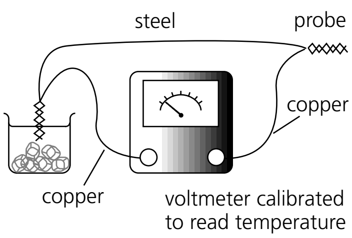
In *liquid crystal thermometers* numbers on a scale are made of different crystalline chemicals. As temperature increases these chemicals change their crystalline structure, which results in colour changes. These are not very accurate.

Pyrometers measure the radiation given off by objects. The characteristic of the radiation changes with temperature. Infrared pyrometers can measure temperature from –20°C to 1500°C. Body temperature is routinely monitored in clinical settings with infrared ear thermometers, which measure the infrared energy emitted from the patient's eardrum in a calibrated length of time. A short tube with a protective sleeve is inserted into the ear, and a shutter is opened to allow radiation from the tympanic membrane to fall on an infrared detector for 0.1 to 0.3 seconds. The device beeps when data collection is completed and a readout of temperature is produced on a liquid crystal display.

Thermistors are semiconductor devices that change their resistance with change in temperature. When these devices are heated their resistance decreases and more current flows. The current is measured on an ammeter, which is calibrated to read temperature.

SPECIFIC HEAT CAPACITY 10.5

Figure 10.12
A thermocouple — a thermometer consisting of two dissimilar metals. The difference in the temperatures of the two ends produces a voltage. This can be calibrated to 'read' temperature.



Do all objects increase their temperature at the same rate as they are heated?

Because of the great variation in molecular structure and bonds that exist between atoms in different substances, energy put into different substances does not result in the same temperature rises. For example, when walking along a beach on a hot day the sand is a lot hotter than the grass or puddles of water. This is because sand only requires 880 J of heat energy from the sun to raise 1 kg of sand by 1°C. Water requires 4200 J.

This property is called the specific heat capacity, *c*, of the substance. It is defined as the amount of energy required to raise the temperature of 1 kg of a substance by 1°C or by 1 K (a change of 1°C is equivalent to a change of 1 K).

Heat capacity is the term used to describe the amount of heat an object contains. Different materials of the same mass and at the same temperature contain different amounts of heat energy. That is, they require different amounts of heat to heat them up and they give out different amounts of energy in cooling down. For example, which contains the more heat — a BBQ plate at 120°C or a coin at the same temperature?

The specific heat capacity of some common substances is given in Table 10.1.

Figure 10.13
A bimetallic strip thermometer. The difference in expansion rates between the two different metals causes a pointer to move across a scale.

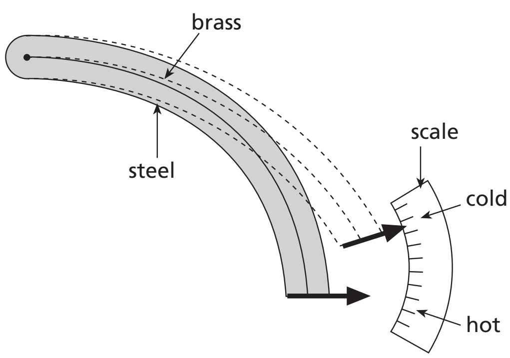


Table 10.1 SPECIFIC HEATS OF SOME COMMON SUBSTANCES

SUBSTANCE	SPECIFIC HEAT CAPACITY c (J kg^{-1} K^{-1})
Lead	130
Mercury	140
Copper	390
Iron and steel	460
Glass	664
Sodium chloride	880
Sand	880
Aluminium	900
Wood	1700
Steam	2020
Ice	2100
Paraffin	2200
Honey	2370
Alcohol	2450
Methylated spirits	2500
Water	4200

Which substance has the highest specific heat capacity?

This is important. Because water has a high specific heat capacity compared with other substances, interesting uses are made of water. Water is used in cooling systems in motor cars. One kilogram of water can take away 4200 J of heat from the engine of a car before its temperature changes by 1°C. What would happen if alcohol was used? The ocean's temperature changes by small amounts compared with the land mass from winter to summer and from day to night. Humans are composed of approximately 70% water, which results in humans responding less to external temperature changes than if they were composed of different materials. Consider how Superman (the man of steel) would be affected by changes in the external temperature.

A useful equation to determine the specific heat of a substance is:

$$Q = mc\Delta T$$

where c is the specific heat capacity in J kg^{-1} °C^{-1}, Q is the quantity of heat in J, m is the mass of the object in kg, ΔT is the change in temperature in °C.
Note: if ΔT is negative then this is the quantity of heat given off by an object.

Example 1

If it takes 4000 J of heat energy to raise the temperature of a 2 kg object by 10°C, what is the specific heat capacity of the object?

Solution

$$\begin{aligned} Q &= mc\Delta T \\ 4000 &= 2 \times c \times 10 \\ c &= 2.00 \times 10^2 \text{ J kg}^{-1} \text{ K}^{-1} \end{aligned}$$

Example 2

How much heat is required to bring a saucepan containing 500 mL of water at 20°C to boiling point? (1 L of water has a mass of 1 kg.)

Solution

$$\begin{aligned} Q &= mc\Delta T \\ &= 0.5 \times 4200 \times (100 - 20) \\ &= 0.5 \times 4200 \times 80 \\ &= 1.68 \times 10^5 \text{ J} \end{aligned}$$

Example 3

If it takes 2850 J of heat to raise the temperature of a 1.5 kg object by 5°C, what substance could the object be made of?

Solution

$$\begin{aligned} Q &= mc\Delta T \\ 2850 &= 1.5 \times c \times 5 \\ c &= 3.80 \times 10^2 \text{ J kg}^{-1} \text{ K}^{-1} \end{aligned}$$

The substance is made of copper (from Table 10.1).

Example 4

A 2 kg block of iron at 25°C is heated by placing it in hot water. If it obtains 4500 J of heat from the water, what is the final temperature of the iron?

Solution

$$\begin{aligned} Q &= mc\Delta T \\ 4500 &= 2 \times 450 \times (T_f - T_i) \\ 4500 &= 900 \times (T_f - 25) \\ T_f &= 50 + 25 \\ T_f &= 75°\text{C} \end{aligned}$$

NOVEL CHALLENGE

When you eat ice cubes, your body uses up energy to melt them and warm them up to body temperature.
Does this mean eating ice would help you lose weight? Could you say they have negative joules?

NOVEL CHALLENGE

In 1700, Dr Charles Blagden took some friends, a dog and a raw beefsteak into a room at 127°C for $\frac{3}{4}$ hour. They all came out unharmed — except for the steak, which was cooked. Why?

NOVEL CHALLENGE

Imagine that you added equal volumes of water and oil to separate beakers and placed them on a hotplate. And after 5 minutes the water began boiling.
Which liquid would be at the higher temperature? Which liquid would have the greater total thermal energy?

Questions

7 Calculate the heat energy absorbed when 1.5 kg of paraffin is raised from 15°C to 50°C.

8 How much heat could be absorbed by a 3 kg block of ice at −10°C before it reaches its melting point?

9 A liquid is heated in a beaker. If it is found that it takes 7500 J of heat to increase the temperature of 500 g of the liquid by 6°C, what could the liquid be?

CALORIMETRY 10.6

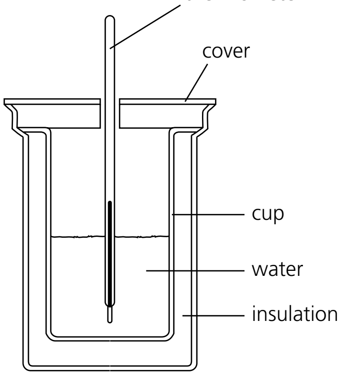
If two substances are placed together in a closed system, that is, one where no energy can escape to the surroundings, then the heat energy lost by one object in the exchange is equal to the heat energy gained by the other. For example, when a cool teaspoon is placed in a hot cup of coffee the heat lost by the coffee and the cup is equal to the heat gained by the spoon.

Heat lost by one substance = heat gained by the other
Q lost = Q gained

This principle is just an extension of the law of conservation of energy — energy is not lost or gained, just transferred or transformed.

In practice there is always some heat lost unless insulation is ideal. However, heat losses to the surroundings can be minimised if experiments are carried out quickly. Scientific experiments also use calorimeters (Figure 10.14), which have good insulation to limit the loss of heat to the surroundings. The process is called calorimetry.

Figure 10.14
A calorimeter — a device used to measure heat exchanges effectively by minimising heat losses to the environment.



Example

If 100 g of alcohol at 50°C is mixed with 250 g of water at 20°C, what is the final temperature of the mixture?

Solution

$$
\begin{aligned}
Q \text{ lost (alcohol)} &= Q \text{ gained (water)} \\
(mc\Delta T)_{\text{alcohol}} &= (mc\Delta T)_{\text{water}} \\
-0.1 \times 2450 \times (T_f - 50) &= 0.250 \times 4200 \times (T_f - 20) \\
-245T_f + 12\,250 &= 1050T_f - 21\,000 \\
12\,250 + 21\,000 &= (1050 + 245)T_f \\
33\,250 &= 1295T_f \\
T_f &= 25.5°\text{C}
\end{aligned}
$$

Note: the negative sign indicates a loss.

TEST YOUR UNDERSTANDING

A common mistake among lower primary students is to say that, when two identical glasses of water both at 40°C are mixed, the final temperature will be 80°C. *Write an explanation suitable for a Grade 3 student about why this is not true. Then explain why you add the volumes together (1 cup + 1 cup = 2 cups).*

Questions

10 A 50 g copper mass is heated by placing it in boiling water. It is then placed in a beaker containing 250 g of an unknown liquid at 20°C. The final temperature of the weight and the liquid is found to be 25°C. What is the specific heat of the liquid? (Assume no heat is lost to the surroundings.)

11 A 1500 W electric jug is used to heat 500 mL of water. Calculate the time for the jug to raise the temperature of the water from room temperature, 20°C, to boiling point. (*Note:* 1500 W means it supplies 1500 J of heat energy every second.) The density of water is 1000 kg m^{-3}.

12 In an experiment, 500 g of copper at 80°C is dropped into 1 kg of kerosene at 20°C. The mixture reached a temperature of 25°C. What is the specific heat capacity of kerosene?

10.7 CHANGE OF STATE

Up to now we have only considered substances changing their temperature as heat is added or taken away. The objects remain in the same state of matter. We will now look at what happens when substances change state, that is, change from a solid to a liquid, a liquid to a gas, or vice versa. This is also called change of 'phase'.

What is happening when a block of ice changes to water involves an understanding of thermal energy and the structure of matter. In Section 10.2 the internal or thermal energy of a substance was defined as the total energy possessed by the particles of the substance. This is made up of both kinetic and potential energies. In solids, like a block of ice, the particles are held firmly in position by the bonds between the particles. They contain kinetic energy in the form of vibrational motion, as well as several forms of potential energy. As the ice is heated, the vibrational motion and therefore the kinetic energy and temperature increase. As the particles vibrate faster they spread apart, also increasing their potential energy. As heating continues the vibrations increase until the molecular forces are no longer strong enough to hold the particles together in fixed positions. The particles break free and are able to slide past one another. The solid melts. It requires a large amount of energy to break the bonds and increase the potential energy of the particles. When this is occurring the addition of thermal energy does not go into changing the kinetic energy of the particles but into increasing the potential energy. Since temperature is a measure of the average kinetic energy of the particles, the temperature does not increase. Thus when a solid melts by the addition of heat, the potential energy of the particles increases without a change in temperature. This is shown graphically in Figure 10.15.

The amount of energy required to melt 1 kg of a substance is called the **specific latent heat of fusion**. (The word 'latent' means hidden.) Why is this heat hidden?

To change 1 kg of ice at 0°C to water at 0°C requires 3.34×10^5 J of energy. Therefore ice has a specific latent heat of fusion of 3.34×10^5 J kg^{-1}.

The reverse is also true; to change 1 kg of water at 0°C to ice at 0°C, that is, to freeze the water, requires the removal of 3.34×10^5 J of energy. The specific latent heat of fusion differs for different substances because of their different bonding. The specific latent heats of fusion of some substances are given in Table 10.2.

Figure 10.15
An effect of heat on the three states of water.

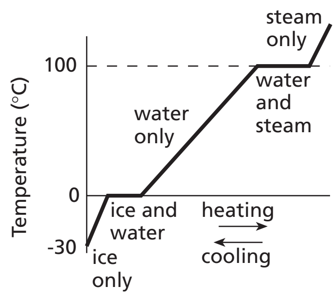


Table 10.2 TYPICAL LATENT HEATS

SUBSTANCE	SPECIFIC LATENT HEAT OF FUSION L_f (J kg^{-1})	SPECIFIC LATENT HEAT OF VAPORISATION L_v (J kg^{-1})
Mercury	1.18×10^4	2.90×10^5
Lead	2.30×10^4	8.64×10^5
Gold	6.30×10^4	1.64×10^6
Silver	1.05×10^5	2.36×10^6
Alcohol	1.09×10^5	8.70×10^5
Aluminium	1.80×10^5	1.14×10^7
Copper	2.05×10^5	4.82×10^5
Iron	2.76×10^5	6.29×10^6
Water	3.34×10^5	2.25×10^6

NOVEL CHALLENGE

At the Le Mans race in France there is 14 km of track and cars reach 360 km h^{-1} (100 m s^{-1}). At the end of the straight drivers approach Mulsanne Corner at 250 km h^{-1} and jam on their brakes to go through the corner at 56 km h^{-1}. The disc rotors glow red hot and sometimes reach 800°C. If a disc rotor (there are four) is 30 cm in diameter and made from steel 0.8 cm thick (density 7.8 g/cm^3), and is 400°C before the corner, show that 95% of the kinetic energy is transferred to heat.

After melting, the addition of heat results in an increase in the kinetic energy (now translational, rotational and vibrational) as well as the potential energies of the liquid. Thus temperature again rises, as shown in Figure 10.15.

As the temperature increases some particles begin to break the cohesion forces holding them together. The forces of attraction between the particles become very weak and the particles move more freely. The substance changes state from a liquid to a gas. At a certain temperature any added thermal energy goes into changing the potential energy of the particles, causing the particles to break the cohesion forces. Again at this point the temperature does not increase as there is no increase in the kinetic energy of the particles. This is shown in Figure 10.15. This temperature is called the **boiling point** of the liquid.

It again requires large amounts of thermal energy to change 1 kg of a liquid to a gas or vapour. For example, it requires 2.26×10^6 J of energy to change 1 kg of water at 100°C into steam at 100°C. The thermal energy required to bring about this change is called the **specific latent heat of vaporisation**.

Specific latent heats of vaporisation of other liquids are given in Table 10.2.

The reverse is again true. It requires the removal of 2.26×10^6 J of thermal energy to change 1 kg of steam into water without change in temperature.

The heat required to melt a mass of a substance is given by the equation:

$$Q = mL_f$$

where Q is the heat required in J, m is the mass of the substance in kg, L_f is the specific latent heat of fusion of the substance in J kg^{-1}.

Similarly, the energy required to vaporise a liquid is given by the equation:

$$Q = mL_v$$

where L_v is the specific latent heat of vaporisation.

Example 1

How much energy is required to change a 2 kg block of lead to liquid at its melting point?

Solution

$$\begin{aligned} Q &= mL_f \\ &= 2 \times 2.30 \times 10^4 \\ &= 4.60 \times 10^4 \text{ J} \end{aligned}$$

Example 2

An ice tray containing 200 g of water at 25°C is placed in the freezer. How much heat energy has to be removed to change the water into ice at −4°C?

Solution

$$\begin{aligned} Q &= (mc\Delta T)_{water} + mL_f + (mc\Delta T)_{ice} \\ &= 0.2 \times 4200 \times (25 - 0) + 0.2 \times 3.35 \times 10^5 + 0.2 \times 2060 \times (4 - 0) \\ &= 21\,000 + 0.668 \times 10^5 + 1648 \\ &= 8.94 \times 10^4 \text{ J} \end{aligned}$$

Questions

13 Find the energy required to melt 2.5 kg of gold at its melting point.

14 Copper has a melting point of 1083°C. Find the energy required to melt 200 g of copper originally at room temperature of 22°C.

15 A 2.0 L bottle of water at 20°C is placed in the freezer of a refrigerator. How much heat must be removed by the refrigerator to freeze this water?

16 A child wanting to make a cordial ice block places 200 g of cordial at 25°C in the freezer. If the freezer can remove energy at the rate of 25 joules per second, what time will it take for the cordial to freeze? (Assume the specific latent heat and specific heat capacity of cordial are the same as water.)

17 Two ice blocks of mass 20 g each are placed in 500 g of water at 40°C. What will be the final temperature of the mixture? (Assume no heat is lost to the container or the surroundings.)

10.8 CHANGING THE MELTING AND BOILING POINTS

Most substances have fixed melting and boiling points as long as they are in pure form. Values given in scientific data are for pure substances. However, the melting point and boiling points can be changed by adding impurities to the substance or by pressure changes. For example, if salt is added to water the melting point is lowered. That is, it freezes at a temperature lower than 0°C.

- Why do councils in some cold countries put salt on the road?
- Why do motorists add another liquid to their radiator in these countries?

Adding impurities raises the boiling point of liquids. If salt is added to water, the water will boil above 100°C.

- How does this affect the cooking rate of potatoes and pasta that are cooked in salty water?

Pressure changes also affect the freezing and boiling points of liquids. Increased pressure lowers the freezing point of liquids that expand when they freeze, such as water, but it raises the freezing point for those that contract when they freeze. This is because increasing pressure pushes the molecules closer together, therefore increasing the temperature at which molecules are attracted to each other and form solids. The following are two examples of this phenomenon.

- If two ice blocks are squeezed together they will melt, but if this pressure is released the water between them will refreeze, gluing them together.
- A thin wire with two weights will cut easily through a block of ice. Why?

Pressure changes also affect the boiling point of a liquid. Increasing the pressure increases the boiling point of a liquid, and decreasing the pressure decreases the boiling point.

- Why is it hard to get a good cup of tea on the top of Mount Everest?

Pressure cookers rely on this property. When the lid is placed on the pressure cooker and the cooker is heated, the pressure can be about twice the normal air pressure. The water boils at about 120°C, and therefore the food cooks faster.

Where else is this principle used?

10.9 EVAPORATION

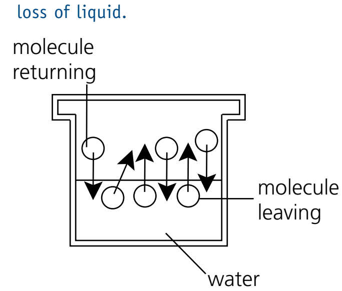
Liquids can change state without boiling. This process is going on all the time in nature. Puddles of water on the road dry up even when the weather is cool. Aftershave and perfume soon disappear from the skin. This is the process of evaporation.

The latent heat of vaporisation plays an important part in evaporation. For molecules of a liquid to change state and become gaseous molecules they require energy. All the molecules in a liquid do not have the same kinetic energy; when molecules evaporate, the faster (hotter) molecules near the surface of the liquid leave first and the slower molecules remain behind. So when water evaporates from the skin your skin feels cooler because the average kinetic energy of those molecules remaining is less. This process, together with sweating, acts as a cooling mechanism for our bodies. When the sweat evaporates from our skin, the skin feels cooler. If there is a breeze blowing, the sweat evaporates faster producing a greater cooling effect.

How can you tell the direction of a breeze by holding up a moist finger?

Evaporation from a container can be stopped or reduced by putting on the lid. In a closed container two processes are occurring: the evaporation of molecules from the liquid and the condensation of molecules back to liquid. If the gaseous molecules are not removed, an equilibrium is reached in which the number of molecules leaving the liquid is equal to the number going back into the liquid. (See Figure 10.16.)

Figure 10.16
In a closed container the rate of evaporation is equal to the rate of condensation, reducing the loss of liquid.



Questions

18 A few years ago it was not uncommon to see cars travelling in country areas with porous water bags made of canvas attached to the front of the car. Why was the temperature of the water in these bags cooler than that of the air?

19 Why does your skin feel cool after placing an alcohol-based aftershave or perfume on it? This is even more evident if you are in a breeze or you blow on it. Is this related to the pain-relief sprays used by footballers?

LAWS OF THERMODYNAMICS 10.10

INVESTIGATING

How could you measure the uneven heating across the carousel inside a microwave if you were given a thermometer, some water and a dozen little plastic containers that rolls of film come in? Be extremely careful if you choose to try it. Start with a low time first, say 20 seconds. Why?

INVESTIGATING

After inventing the first, second and third laws of thermodynamics, physicists decided they had a 'zeroth law'. What on Earth is that?

✔ TEST YOUR UNDERSTANDING

(Answer true or false)

- Heat and temperature are the same thing.
- Heat and cold flow like liquids.
- The hotter of two objects contains the more heat.

What happens when you place a cold body in contact with a warm one, for example, placing an ice cube in a glass of water?

The answer to this has already been discussed. As the molecules collide energy is transferred from those that contain the most to those that contain less, that is, from the hot body to the cool one until equilibrium is reached.

Heat flows from the hotter body to the cooler one.

Under perfect conditions where energy is not lost to the surroundings, the energy lost by the hot body is equal to that gained by the cool one.

What are other ways of 'heating' up a cool body, that is, giving the molecules of the cool body more energy? Think of how you heat up your hands on a cold morning. By rubbing your hands together the mechanical energy of your moving hands, with friction, causes your hands to become hotter.

James Joule (as previously discussed) increased the temperature of water by using the potential energy of the falling weight.

The thermal energy of a system can be increased by adding heat to it or by doing work on it.

The total increase in the thermal energy of an isolated system is equal to the sum of the heat added to it and the work done on it. This is the **first law of thermodynamics**. Notice this is just an extension of the principle of conservation of energy.

The **second law of thermodynamics** formulated by a German physicist Rudolf Clausius (1822–88) relates heat transfer to differences in temperature. For example, what are some devices that take heat out of objects or the air?

Some you may have guessed are: refrigerators, airconditioners, freezers, or even the cooling system of a car.

Each of these devices takes heat from something and transfers it to something of lower temperature. What would happen if the car was driven in a place where the atmospheric temperature was higher than the temperature of the car engine?

This is one of the reasons it is suggested that it is impossible to obtain a temperature of zero kelvins. There is no place to transfer the heat, no point of lower temperature.

To transfer internal energy from a low temperature heat source to a higher one requires work, such as the work done by the motor in a refrigerator or an air conditioner.

Consider now a hot body which is placed in contact with a cold body. Heat is transferred from this hot body to the cold body and after a time equilibrium is reached. If it is an isolated system there is no loss in energy in the process but the system has lost its capacity to do work. There is no longer as great a temperature difference. This is seen in the example of placing a block of ice in hot water. Before placing the ice in the water the molecules of the ice are in a well-ordered crystalline arrangement. After the equilibrium temperature is reached the molecules are in a less ordered, more random motion. The molecules of the ice have become less ordered. The scientific term that defines the orderliness of the molecules is **entropy**.

Entropy is a measure of the disorder of a system. The more disorder the greater the entropy. Natural processes always go in a direction that causes an increase in the total entropy of the universe. This is the second law of thermodynamics. This law would indicate that the availability of energy in the universe is decreasing.

You can, of course, have a decrease in entropy in one part of the universe (e.g. freezing water) but there is an overall greater increase in entropy elsewhere (e.g. heat produced at the rear of the refrigerator).

— Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***20** Compare the temperature and thermal energy of a cup of coffee at 90°C and a swimming pool at 22°C.

***21** Four different masses of the same metal, 100 g, 200 g, 500 g, and 1 kg, were heated by a 2000 J energy supply. If the 100 g mass changed its temperature by 5°C, what would have been the temperature changes of the other three?

***22** State two reasons why a mercury-in-glass thermometer could not be used to measure the temperature of a pottery kiln.

***23** Convert the following Celsius temperatures to kelvins: **(a)** 290°C; **(b)** −25°C; **(c)** 59.2°C.

***24** Change the following Kelvin temperatures to °C: **(a)** 69 K; **(b)** 1376 K; **(c)** 345.6 K.

***25** 1500 J of energy is used to heat a 400 g sample of iron initially at 28°C. What would be the final temperature of the iron?

****26** A beaker containing 200 g of mercury at 15°C was placed in a freezer. Find the energy removed from the mercury if it cooled down to −4°C. Would the freezer have to work harder if the beaker contained 200 g of water instead of mercury? Explain!

****27** An electric kettle was used to heat 500 g of water at 20°C. If the kettle can supply energy at the rate of 1500 J per second, and was turned on for one and a half minutes, what was the final temperature of the water?

****28** A 200 g bar of aluminium was heated in a bunsen burner until its temperature was 150°C. It was then plunged into a beaker containing 500 g of paraffin at 50°C. What was the final temperature of the mixture?

***29** Why is heat needed to change a solid to a liquid but the substance does not change temperature?

***30** Explain why steam at 100°C would be more effective in heating a cup of milk to make coffee (a cup of cappuccino, for example) than water at 100°C?

****31** Dry steam is used to make a cup of coffee by bubbling it through water. If the steam is at 100°C, what mass of steam must be used to heat 200 g of water from 25°C to 95°C?

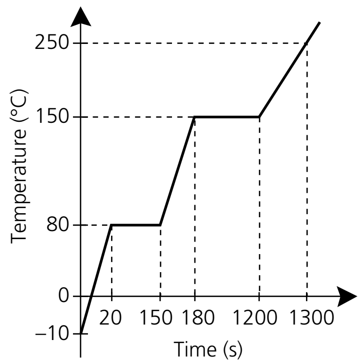
****32** An electric kettle, rated at 2 kW, is filled with water and its total mass determined. The kettle is switched on and the water is allowed to boil for a further 60 s after coming to the boil. The kettle is found to be 80 g lighter. Calculate the specific latent heat of vaporisation. Suggest why this value differs from the stated value of the latent heat of vaporisation.

***33** Suggest the weather conditions that would make clothes on an outside line dry faster.

***34** State two ways in which boiling is different from evaporation.

***35** State two ways in which the rate of evaporation can be increased.

Figure 10.17
A heating curve for question 36.



36 Figure 10.17 shows the relationship between the temperature of 100 g of an unknown substance and the time it is heated by a 1000 W source.
(a) What is the melting point of the substance?
(b) What is the boiling point of the substance?
(c) How much energy is required to melt the substance?
(d) What is the specific latent heat of fusion of this substance?
(e) What is the specific latent heat of vaporisation of this substance?
(f) Calculate the specific heat capacity of the substance in the liquid state.

****37** While carrying out an experiment to measure the boiling point of water at various altitudes, students found that instead of the boiling point decreasing as they went higher in the mountains, the temperature rose slightly. Suggest why this might have occurred.

****38** Students performing an experiment on naphthalene to discover how its temperature changed with time allowed hot naphthalene to cool down. Table 10.3 lists their results.

Table 10.3

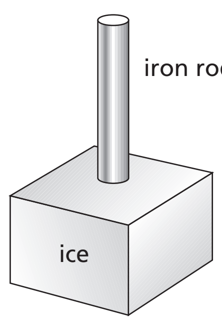
Time (minutes)	0	1	2	3	4	5	6	7	8	9	10
Temperature (°C)	103	93	83	80	80	80	80	75	70	65	60

(a) Plot the graph of temperature against time.
(b) Explain what is happening between the times of 3 minutes and 6 minutes.

Extension — complex, challenging and novel

*****39** An electric shower unit is rated at 5 kW. If water enters it at 15°C and leaves it as hot water at the rate of 5 kg per minute, what is the temperature of the hot water?

Figure 10.18
For question 40.



*****40** An iron rod of mass 100 g and 4 cm in diameter was placed in a furnace and heated until its temperature reached 150°C. It was then placed on its end on top of a large block of ice (temperature of 0°C). (See Figure 10.18.) How far into the block of ice will the rod sink (assume no heat is lost to the surroundings).

*****41** A block of ice of temperature 0°C and mass 20 g was placed in a beaker and weighed. The total mass was 55 g. Steam at 110°C was ducted onto the ice until the ice completely melted. Assuming no loss of heat to the surroundings, find the mass of the beaker and its contents!

CHAPTER 11

Heat and Matter

11.1 INTRODUCTION

- Why is it that you can smell if a gas tap has been left on when you walk into a laboratory but you cannot smell the water spilt on the front bench?
- Have you ever wondered why you can smell perfume or aftershave lotion?
- Why are many of our anti-pollution laws, especially in very industrialised cities and nations such as some cities in America and Japan, concerned with gases?
- Can hot water freeze more rapidly than cold water?

The answer to each of these questions has something to do with the internal structure of gases, liquids and solids. In this chapter we will look at the theory underlying this structure and hence begin to understand the nature of gases.

11.2 THE KINETIC MODEL OF GASES

In the previous chapter we learnt that gases consist of particles that are relatively free in their ability to move compared with those of liquids and solids. These particles also exert almost no force on each other. Because of this the analysis of gas particles is easier. The understanding of the nature of these particles plays an important part in the understanding of heat, heating, and temperature.

To simplify the understanding of how the motion of the particles of a gas affects the properties of a gas, several assumptions are made about these particles.

Assumption 1 The particles of a gas are in constant random motion. They move at high speeds in straight lines unless they collide with the walls of the container or other particles. These collisions are elastic ('elastic' means they do not lose kinetic energy when they collide).

This helps us to explain why gases mix so readily and why gases fill containers.

Assumption 2 The particles of a gas are separated by large distances compared with the diameter of the particles, which are assumed to be negligible in size.

This explains why gases can be compressed and gas densities, in normal situations, are very small compared with those of solids and liquids.

Assumption 3 The force of attraction between particles is negligible, because they are large distances apart.

Assumption 4 Since heating a substance changes the motion of the particles of the substance, the temperature of a gas is a measure of the average speed or kinetic energy of the particles of the gas.

NOVEL CHALLENGE

Imagine a 1 cm square on your skin. Every second there are 10^{22} blows from air molecules. *Can you detect it? If you can't, give some reasons why you can't feel these blows. If the blows stopped, what would you feel?*

11.3 PRESSURE

The pressure of gases plays a major part in everyday life: the pressure of gases in the atmosphere affects the weather; the pressure of air in tyres affects the 'ride' of a car. But what is pressure and how do the particles of a gas exert a pressure?

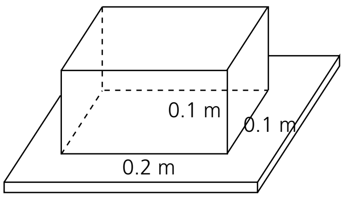
Pressure is the force per unit area ($P = F/A$)

Hence the unit of pressure is a newton per square metre ($N\ m^{-2}$) or the modern SI unit of the pascal (Pa), named in honour of the French scientist Blaise Pascal (1623–1666).

The interchanging of the terms ‘force’ and ‘pressure’ is common with new physics students but there are major differences that can be illustrated with the following example.

A rectangular solid of 2 kg mass is placed on a table as shown in Figure 11.1. The force this object exerts on the table is 20 N. (This is its weight.) But the pressure it exerts is determined by its area of contact:

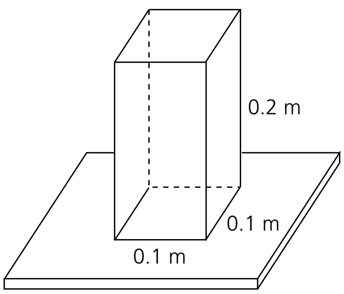
Figure 11.1
Pressure exerted by an object with the 0.2 m × 0.1 m face on the table.



$$P = \frac{20\text{ N}}{0.2\text{ m} \times 0.1\text{ m}}$$
$$= \frac{20\text{ N}}{0.02\text{ m}^2}$$
$$= 1000\text{ Pa}$$

But if the object is now placed on its end as shown in Figure 11.2, the force it exerts on the table will remain at 20 N, but the pressure is now:

Figure 11.2
Pressure exerted by an object with the 0.1 m × 0.1 m face on the table.



$$P = \frac{F}{A}$$
$$= \frac{20\text{ N}}{0.1\text{ m} \times 0.1\text{ m}}$$
$$= \frac{20\text{ N}}{0.01\text{ m}^2}$$
$$= 2000\text{ Pa}$$

It can be seen that if the area of contact is small the pressure is very large. Why do women wearing stiletto heels leave impressions on wooden or cork floors? (And it is not because they are heavy.)

- Where else is this effect seen?
- How does this affect our discussion of gases?

When each gas particle collides with the wall of the container it exerts a force on a small area of the wall. This collision produces pressure. Since gases contain many particles it is the constant collisions with the walls of the container that result in the pressure of the gas in the container. This can be seen when you blow up a balloon. When you start, it contains few particles, which make few collisions with the walls, resulting in low pressure. When the balloon contains more particles there are more collisions, exerting greater pressure on the walls, forcing the balloon to expand.

NOVEL CHALLENGE

A matchstick is placed in a test-tube of water. When you place your thumb over the top and press down, the match sinks. Propose a reason for this and test to see if we’re lying!

THE GAS LAWS 11.4

The properties of a gas are easy to explain because the particles act independently without exerting any significant forces on each other. This is true except in the extremes, when the temperature is very low or when the pressure is high. In these circumstances the particles are maintained in close proximity to each other. The properties of gases that play a part in understanding the behaviour of gases are volume, pressure, temperature and the number of particles in the sample.

The relationships between these variables have been investigated for centuries and affect how we handle gases today.

Boyle's law

One of the earliest scientists to investigate the relationships between the above variables was the British chemist and physicist **Robert Boyle** (1637–91). By experimenting with gases he established that the volume of a gas decreased as the pressure of the gas increased. If the temperature of a confined gas sample was kept constant and the pressure on the gas increased by placing more mass on a piston, as shown in Figure 11.3, the volume of the gas changed, as shown in Figure 11.4. This suggested that pressure was inversely proportional to volume.

When the pressure was plotted against the inverse of volume, Boyle obtained a straight line, as shown in Figure 11.5. This indicated that pressure is directly proportional to the inverse of volume ($P \propto 1/V$),

or $$PV = \text{constant}$$

This relationship is known as **Boyle's law**, which states: For a fixed mass of gas at constant temperature the pressure of the gas varies inversely as the volume. This means for a particular sample of gas at constant temperature an increase in pressure from P_1 to P_2 causes a corresponding decrease in volume from V_1 to V_2.

$$P_1V_1 = \text{a constant} = P_2V_2$$

or $$P_1V_1 = P_2V_2$$

This is normally how Boyle's law is expressed when solving problems.

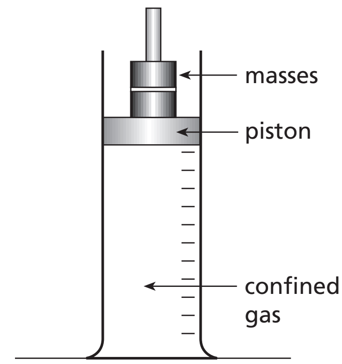


Figure 11.3
A device used to show the relationship between pressure and volume.

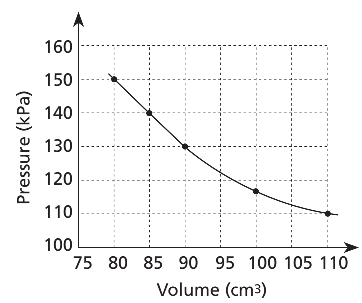


Figure 11.4
Pressure–volume relationship of a confined gas.

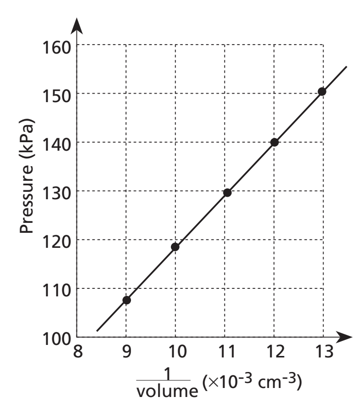


Figure 11.5
Pressure vs 1/volume graph to establish Boyle's law.

Example

A scuba diver releases a 1.0 cm^3 bubble of gas at a depth where the pressure is 4 atmospheres. What will be the volume of that gas at the surface where the pressure is 1 atmosphere (assuming the temperatures are the same)?

Solution

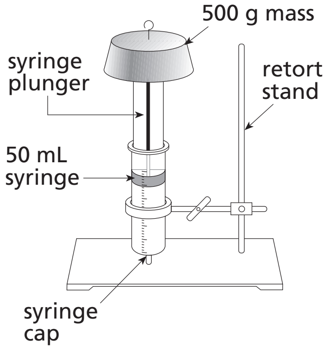
$$P_1V_1 = P_2V_2$$
$$4 \text{ atm} \times 1 \text{ cm}^3 = 1 \text{ atm} \times V_2$$
$$V_2 = 4 \text{ cm}^3$$

Note: the units of pressure and volume are not that important as long as they are consistent. That is, P_1 and P_2 have to have the same units and V_1 and V_2 as well. Some common units of pressure are pascals (Pa), mm of mercury (mmHg) and atmospheres (atm).

Questions

1. A balloon of volume 2.0 L contains air at 230 kPa. What would be the pressure of the gas when its volume is reduced to 0.50 L?
2. A hot-air balloon has a volume of 10 m^3 at sea-level. The balloon then rises to a height in the atmosphere where the pressure is 0.20 atmospheres. What would be the resulting volume of the balloon? (Assume constant temperature.)
3. A diver dives to a depth of 40 m in fresh water where he releases a toy balloon of volume 10 cm^3. What will be the size of the balloon when it reaches the surface? (The pressure increases at a rate of 1 atmosphere for every 10 m descent in fresh water.)
4. A student testing Boyle's law places masses on the top of a syringe as shown in Figure 11.6. With 500 g on the top of the syringe the volume is 50 mL. What mass will need to be placed on the piston for the volume to be 12.5 mL?

Figure 11.6
For question 4.



Charles' law

A French scientist, **Jacques Charles** (1746–1823), investigated the relationship between the volume of a confined gas and the temperature of a gas. For example, imagine heating a gas contained in a syringe in which the plunger is free to move and under atmospheric pressure. As the gas is heated the particles move more rapidly, making more frequent and forceful collisions with the walls and the plunger. The plunger will move outwards to a position that re-establishes the equilibrium between the pressure produced by the gas particles and atmospheric pressure. The pressure and number of particles remain constant but the volume of the gas increases with increasing temperature. This is shown graphically in Figure 11.7.

Notice that there is a direct relationship between volume of the gas and kelvin temperature.
That is:

$$V \propto T \text{ (kelvin)}$$

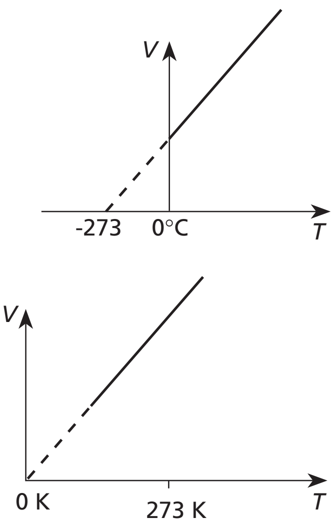
or

$$\frac{V}{T} = \text{constant}$$

or

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

Figure 11.7
Volume–temperature relationship for a confined gas: notice that there is a direct relationship between the volume of the gas and its Kelvin temperature.



Therefore, for a confined gas where the pressure remains constant the volume of the gas is directly proportional to its kelvin temperature. This is **Charles' law**.

This can very easily be seen by placing an inflated balloon in liquid nitrogen. The volume will shrink considerably. A good result can be seen by placing the balloon in a freezer.

When Charles did this experiment he could not cool a gas below −20°C so he had to extrapolate the graph to meet the temperature axis. This occurred at −273°C. This has startling implications. It would suggest that the volume of a gas at −273°C is zero. Charles' law is true for real gases except at low temperatures, and since no substance can exist as a gas below 2 K, it is theoretical and not practical outside this range.

Amontons' law

Guillaume Amontons (1663–1705) was a French physicist who in 1699 discovered that equal changes in temperature of a fixed volume of air produced equivalent variations in pressure ($P \propto T$). This law is sometimes referred to as Gay-Lussac's law. Before chemical fireworks were invented, the Chinese used to throw bamboo onto fires for amusement. The heat would increase the pressure of the gas, causing the pressure of the trapped gas to increase and making the chambers explode. That's why we don't throw pressure-pack spray cans onto a fire.

NOVEL CHALLENGE

Put a lit candle in a jar with no lid and place on a record turntable. Before starting the turntable predict whether the flame will point inwards or outwards. What if there is no jar?

Questions

5 A balloon containing 40 cm^3 of air at 25°C is placed in the freezer where the temperature is −10°C. What will be its volume in the freezer?

6 At what temperature will a 500 cm^3 balloon of gas at 300 K have a volume of 300 cm^3, if the pressure remains constant?

7 Twenty litres of oxygen at 30°C is cooled under constant pressure to −140°C. What will be the new volume of the gas?

8 A 100 cm^3 balloon is filled with hydrogen at 20°C. If the balloon is then released to rise in the atmosphere to a height where the temperature is −50°C, what will be the volume of the balloon? (Assume constant pressure.)

The combined gas equation

Boyle's law, Charles's law and Amontons' law can be combined to obtain an equation that relates the pressure, volume and temperature of a fixed amount of gas.

Since:

- $V \propto 1/P$ (Boyle's law)
- $V \propto T$ (Charles's law)
- $P \propto T$ (Amontons' law)

then $$V \propto \frac{T}{P}$$

then $$\frac{PV}{T} = \text{constant}$$

That is, for a fixed mass of a particular gas:

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

This is called the **combined gas equation**. Notice that for this equation to hold, temperature has to be measured in kelvins. Pressure and volume can be in any units as long as they are consistent across the equation.

Example

A 0.20 L sample of gas at room temperature (20°C) and atmospheric pressure is heated to 200°C and allowed to expand to 0.30 L. What will be the new pressure of the gas?

Solution

- 20°C = (20 + 273) = 293 K.
- 200°C = (200 + 273) = 473 K.

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

$$1 \text{ atm} \times 0.2 \text{ L} / 293 \text{ K} = P_2 \times 0.3 \text{ L} / 473 \text{ K}$$

$$(1 \times 0.2) / 293 \times 473 / 0.3 = P_2$$

$$P_2 = 1.08 \text{ atm}$$

Questions

9 A tank of volume 0.025 m^3 containing a mixture of nitrogen and helium gas is used to inflate party balloons. The pressure in the tank is 2.0×10^7 Pa and at a temperature of 293 K. How many balloons of size 0.0010 m^3, at a temperature of 300 K and at atmospheric pressure, can be filled from the tank?

10 A tank containing 200 L of hydrogen gas at 20°C is kept under a pressure of 200 kPa. The temperature is raised to 90°C and the volume decreased to 150 L. What is the pressure of the gas in the container?

11 Oxygen gas is released from a cylinder at the rate of 10 m^3 per hour at atmospheric pressure. If it is sold at a pressure of 200 atm and has a volume of 2.0 m^3, for how long will it supply oxygen?

NOVEL CHALLENGE

Landmines are deadly explosives buried underground. They are about 10 cm by 10 cm in area and most require 500 kPa to detonate. What mass of person would it take to do that?

The ideal gas equation

The four variables previously discussed that affect the nature of a gas were:

- volume
- temperature
- pressure
- the number of particles in the gas.

So far in each of the previous laws we have considered the number of particles in the gas to be constant. Suppose we measure the temperature, pressure and volume of a gas and then keep the temperature and volume constant but introduce twice the number of particles into the container. What will happen to the pressure? Because the pressure of a gas is due to the collisions of the particles on the walls of the container, the pressure will be twice the original. There are twice as many particles so there are twice as many collisions. Pressure is therefore proportional to the number of particles in the container.

$$P \propto N$$

Since $P \propto 1/V$, and $P \propto T$, it then follows that:

$$P \propto TN/V$$

or

$$PV = kNT$$

where k is a gas constant called **Boltzmann's constant**, named after the Austrian physicist **Ludwig Boltzman** (1844–1906). The units of k are derived from the other variables in the equation.

- From the above, $k = PV/NT$.
- Therefore k has the units of Pa m^3 molecule^{-1} K^{-1} (or J K^{-1} molecule^{-1}).
- The value of k is 1.38×10^{-23} Pa m^3 molecule^{-1} K^{-1}.

Activity 11.1 EQUIVALENT UNITS

Show that the unit Pa m^3 molecule^{-1} K^{-1} is equivalent to J K^{-1} molecule^{-1}.

Because the number of particles in a gas is extremely large we use a **mole** as the unit for the amount or the number of particles. One mole is equal to 6.02×10^{23} particles. This number is called **Avogadro's number** (N_A), after the Italian scientist **Amedo Avogadro** (1776–1856). The enormously large value of Avogadro's number suggests how tiny and how numerous atoms must be. A mole of air can fit into a suitcase. Yet, if these molecules were spread uniformly over the Earth there would be 120 000 of them in every square centimetre. A second example to indicate the size of a mole is that one mole of tennis balls would fill a volume equal to three Moons.

In the previous equation if we change the number of particles to moles we also have to change the proportionality constant, producing the **ideal gas equation**:

$$PV = nRT$$

where R is the universal gas constant, and n is the number of moles of gas.

The value of R is 8.31 Pa m^3 mol^{-1} K^{-1} or 8.31 J K^{-1} molecule^{-1}.

Activity 11.2 CONVERSION OF UNITS

SR

To convert the number of particles (N) to moles (n) we divide by Avogadro's number (N_A – 6.02×10^{23}). If we multiply k (1.38×10^{-23} Pa m^3 particle^{-1} K^{-1}) by the number of particles do we obtain 8.31?

Note 1: since the constant is in terms of standard units, all variables in the equations $PV = NkT$ or $PV = nRT$ have to be in standard units. That is, pressure needs to be measured in Pa, volume in m^3, and temperature in K.

Note 2: gases exist as either atoms or molecules. For example, helium, argon, and neon are single atoms whereas oxygen, hydrogen, and carbon dioxide exist as molecules containing two or more atoms. We use the term 'particles' to refer to both atoms and molecules.

Example 1

Find the number of particles in a 20 L sample of argon gas at a temperature of 273 K and at atmospheric pressure.

Solution

- $P = 1$ atm $= 1.013 \times 10^5$ Pa.
- $T = 273$ K.
- $V = 20$ L $= 20 \times 10^{-3}$ m^3.

$$PV = NkT$$
$$1.013 \times 10^5 \text{ Pa} \times 20 \times 10^{-3} \text{ m}^3 = N \times 1.38 \times 10^{-23} \times 273 \text{ K}$$
$$N = 1.013 \times 10^5 \text{ Pa} \times 20 \times 10^{-3} \text{ m}^3/1.38 \times 10^{-23} \times 273 \text{ K}$$
$$N = 5.4 \times 10^{23} \text{ particles}$$

Example 2

Air has an average molecular weight (molar mass) of 29 g per mole. (That is, 1 mole of gas particles has a mass of 29 g.) What is the volume of 2.0 kg of air at atmospheric pressure and 20°C?

Solution

- $P = 1$ atm $= 1.013 \times 10^5$ Pa.
- $T = 20°C = 293$ K.
- $n = 2$ kg/29 g mole^{-1} = 68.97 mole.
- $R = 8.31$ m^3 Pa mol^{-1} K^{-1}.

$$PV = nRT$$
$$V = nRT/P$$
$$= 68.97 \times 8.31 \times 293/1.013 \times 10^5$$
$$= 1.66 \text{ m}^3$$

Questions

12 What is the pressure of 0.030 moles of hydrogen gas at 37°C if its volume is 80 mL?

13 A 0.50 L sample of oxygen gas has a pressure of 2.0 atm at a temperature of 80°C. How many molecules of oxygen are in the sample?

14 What amount of carbon dioxide gas occupies 2.0×10^{-4} m^3 at a pressure of 150 kPa and a temperature of 20°C?

15 A 500 mL sample of helium gas contains 2.0 moles of gas at 300 K.

(a) Find the pressure exerted by the particles.

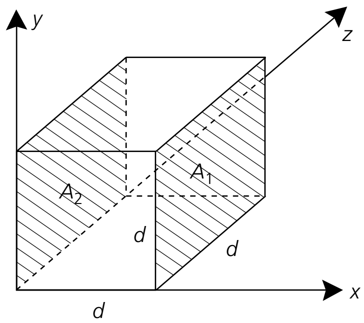
(b) If the temperature is doubled and half of the gas escapes, what will be the new pressure?

Ch11-Walding 4th 25/8/04 11:45 AM Page 261

11.5 TEMPERATURE AND KINETIC ENERGY

Let us now calculate the pressure of an ideal gas from kinetic theory. To simplify matters, we will consider a gas in a cubic vessel whose walls are perfectly elastic. Let each edge be of length d. Call the faces normal to the x axis (Figure 11.8) A_1 and A_2, each of area d^2. Consider a particle that has a velocity $\boldsymbol{v}$. We can resolve $\boldsymbol{v}$ into components $\boldsymbol{v}_x$, $\boldsymbol{v}_y$ and $\boldsymbol{v}_z$, the directions of the edges. If this particle collides with A_1 it will rebound, with its x component of velocity reversed. There will be no effect on $\boldsymbol{v}_y$ or $\boldsymbol{v}_z$, so the change in the particle's momentum $\Delta\boldsymbol{p}$ will be normal to A_1. Hence, the change in momentum of the particle will be:

Figure 11.8
Motion of a particle in a box.



$$\Delta\boldsymbol{p} = \boldsymbol{p}_f - \boldsymbol{p}_i = -m\boldsymbol{v}_x - (m\boldsymbol{v}_x) = -2m\boldsymbol{v}_x$$

Thus the momentum imparted to A_1 will be $2m\boldsymbol{v}_x$ since the total momentum is conserved.

Suppose that this particle reaches A_2 without striking any other particle on the way. The time required to cross the cube will be $d/\boldsymbol{v}_x$. At A_2 it will again have its x component of velocity reversed and will return to A_1. Assuming no collisions in between, the round trip will take a time of $2d/\boldsymbol{v}_x$. Hence, the number of collisions per unit time this particle makes with A_1 is $\boldsymbol{v}_x/2d$, so the rate at which the particle transfers momentum to A_1 is:

$$2m\boldsymbol{v}_x \times \boldsymbol{v}_x/2d = m\boldsymbol{v}_x^2/d$$

To obtain the total force on A_1, that is, the rate at which momentum is imparted to A_1 by all the gas molecules, we must sum up $m\boldsymbol{v}_x^2/d$ for all the particles. Recall from Chapter 4 that $\boldsymbol{F}t = \Delta\boldsymbol{p}$, or $\boldsymbol{F} = \Delta\boldsymbol{p}/t$.

Then, to find the pressure, we divide this force by the area of A_1, namely d^2. If m is the mass of each molecule, we have:

$$\begin{aligned} P &= F/A \\ &= m\boldsymbol{v}_x^2/d/d^2 \text{ for one particle} \end{aligned}$$

The total pressure is:

$$P = m/d^3 \times (\boldsymbol{v}_{x1}^2 + \boldsymbol{v}_{x2}^2 + \boldsymbol{v}_{x3}^2 + \ldots \boldsymbol{v}_{xn}^2)$$

where $\boldsymbol{v}_{x1}$ is the x component of the velocity of particle 1, $\boldsymbol{v}_{x2}$ is that of particle 2, etc. If N is the total number of particles in the container and n is the number per unit volume, then $N/d^3 = n$ or $d^3 = N/n$. Hence:

$$P = mn(\boldsymbol{v}_{x1}^2 + \boldsymbol{v}_{x2}^2 + \boldsymbol{v}_{x3}^2 + \ldots \boldsymbol{v}_{xn}^2)/N$$

But mn is simply the mass per unit volume, that is, the density ρ of the gas we're considering. The quantity $(\boldsymbol{v}_{x1}^2 + \boldsymbol{v}_{x2}^2 + \boldsymbol{v}_{x3}^2 + \ldots \boldsymbol{v}_{xn}^2)/N$ is the average value of $\boldsymbol{v}_x^2$ for all particles in the container. Let us call this $\overline{\boldsymbol{v}_x^2}$.

Then:

$$P = \rho\overline{\boldsymbol{v}_x^2}$$

NOVEL CHALLENGE

French scientist Pierre Gassendi (1592–1655) said that the pressure of a gas doesn't depend on the weight of the gas in a container.
Was he correct? Could you have two equal rigid containers with the same mass of the same gas in each but exhibiting different pressures?

For any particle, $\boldsymbol{v}^2 = \boldsymbol{v}_x^2 + \boldsymbol{v}_y^2 + \boldsymbol{v}_z^2$ (Pythagoras's theorem). Because we have many particles and because they are moving entirely at random, the average values of $\boldsymbol{v}_x^2$, $\boldsymbol{v}_y^2$, and $\boldsymbol{v}_z^2$ are equal and the value of each is exactly one-third the average value of $\boldsymbol{v}^2$. There is no preference among the molecules for motion along any one of the three axes.

Hence, $\overline{\boldsymbol{v}_x^2} = \frac{1}{3}\overline{\boldsymbol{v}^2}$, so that:

$$\begin{aligned} P &= \rho \boldsymbol{v}_x^2 = \tfrac{1}{3}\rho \boldsymbol{v}^2 \\ &= \tfrac{1}{3}mn\,\overline{\boldsymbol{v}^2} \\ &= \tfrac{1}{3}m\,\overline{\boldsymbol{v}^2}\,n \\ &= \tfrac{1}{3}m\,\overline{\boldsymbol{v}^2}\,N/d^3 \\ &= \tfrac{2}{3}\,\tfrac{1}{2}\,m\,\overline{\boldsymbol{v}^2}\,N/V \\ &= \tfrac{2}{3}\,\overline{E_k}\,N/V \end{aligned}$$

From Chapter 9, kinetic energy $(E_k) = \frac{1}{2}\,m\,\boldsymbol{v}^2$, so:

$$PV = \tfrac{2}{3}\,\overline{E_k}\,N$$

This can be equated to the general equation $PV = NkT$.

$$\text{Then} \quad kT = \tfrac{2}{3}\,\overline{E_k}$$

$$\text{or} \quad \tfrac{3}{2}\,kT = \overline{E_k}$$

Example

Find the average kinetic energy of the air particles, and the average speed of the nitrogen molecules in the laboratory at room temperature of 22°C.

Solution

$$\begin{aligned} \overline{E_k} &= \tfrac{3}{2}\,kT \\ &= \tfrac{3}{2}\,1.38 \times 10^{-23} \times (22 + 273)\ \text{K} \\ &= 6.01 \times 10^{-21}\ \text{J} \end{aligned}$$

One mole of nitrogen molecules has a mass of 28 g, therefore the mass of a nitrogen molecule is $28/6.02 \times 10^{23}$ g, equals 4.65×10^{-26} kg.

$$\begin{aligned} \overline{E_k} &= \tfrac{1}{2}m\overline{\boldsymbol{v}}^2 \\ 6.10 \times 10^{-21}\ \text{J} &= \tfrac{1}{2}4.65 \times 10^{-26}\ \overline{\boldsymbol{v}}^2 \\ \boldsymbol{v}^2 &= 6.10 \times 10^{-21} \times 2/4.65 \times 10^{-26} \\ &= 2.6 \times 10^5 \\ \overline{\boldsymbol{v}} &= 5.10 \times 10^2\ \text{m s}^{-1} \end{aligned}$$

(Notice that the speed of gas particles in the room is surprisingly high, 500 m s^{-1}, and if the gas under consideration was hydrogen the speed of the particles would be much greater.)

PHYSICS FACT

It takes an alcohol molecule 6 months to travel from the front of your laboratory to the back (if the windows are closed). During this time it would collide 6000 times per second and travel a total distance of 60 million kilometres. Now that's a lot of 'Fly Buys'.

Questions

16 Find the average speed of the oxygen molecules in the above example. (The mass of oxygen is 32 g per mole.)

17 At a certain temperature a sample of oxygen has a pressure of 2.0×10^6 Pa, and a density of 4.0×10^{-3} g cm^{-3}. Find the average speed of the oxygen molecules of the gas.

18 Samples of two ideal gases, argon and helium, are mixed and heated to 150°C. Find:
(a) the ratio of the average kinetic energy of the argon atoms to the average kinetic energy of the helium atoms;
(b) the ratio of the average velocity of the atoms of argon to the average velocity of the atoms of helium. (The mass of argon is 40 g mol^{-1}, and the mass of helium is 4 g mol^{-1}.)

19 A fluorescent light tube consists of a cylinder 1.2 m long with a diameter of 3 cm. It contains neon gas of density 1.12×10^{-3} g cm^{-3}. If it is heated to 80°C what pressure is the gas in the light? (Mass of neon = 20 g mole^{-1}.)

A FINAL NOTE — REAL GASES 11.6

The above equations are developed for ideal gases, that is, gases that consist of single atoms only. Therefore, all energy put into these gases by means of heat then goes into translational motion (Latin *trans* = 'across', *latio* = 'bringing') — motion in a straight line. That is, it makes the particles move faster in a certain direction. There are several ideal gases: helium, neon, argon, krypton, xenon, and radon. However, most gases have two or more atoms per molecule. For example, nitrogen does not, in gaseous form, exist as an individual atom but as a molecule that consists of two atoms bonded together. Other examples are oxygen (O_2), hydrogen (H_2) and carbon dioxide (CO_2). Energy put into these gases by means of heating does not all go into translational motion. Some goes into rotational motion — the atoms spin around as they travel along; and some goes into potential energy stored in the bonds of the atoms. Because of this, these gases deviate from the ideal gas equation.

Ideal gases are assumed to consist of particles that are insignificantly small and have no attractive force on each other. Real gas particles do have size so it is surprising that at normal temperatures and pressures they, in fact, obey the ideal gas equation very well. However, real gases deviate from the ideal gas equation as the temperature decreases and the pressure increases. At normal temperatures and pressures the particles of oxygen occupy very little volume compared with the total volume of the gas. The particles are moving so fast that they exert little attractive force on each other. But at low temperatures when they are moving slowly, or at high pressures when they are pushed close together, the attractive force is greater and the volume of the particles is significant compared with the volume of the gas. Under these conditions a real gas deviates from the behaviour indicated by the ideal gas equation.

THERMAL EXPANSION 11.7

So far we have looked at the effects of heat on gases. These effects are easier to understand because of the limited effect particles of the gas have on one another (except in collisions). The addition of heat affects the particles of the gas by making them move faster and thus expanding the gas or increasing its pressure on its container. Does the addition of heat have the same effect on the particles of solids and liquids?

Can you think of everyday examples where the addition of heat affects a solid or liquid?

- What causes the mercury in a thermometer to rise (Figure 11.9)?
- Why do trains make the 'clicky–clack' sound when moving over railway lines?
- Have you ever considered why people building ships or high-rise buildings in old movies use white-hot rivets to hold steel plates together? Why not use cold ones? They are easier to handle and do not lead to all those comical situations in cartoons.

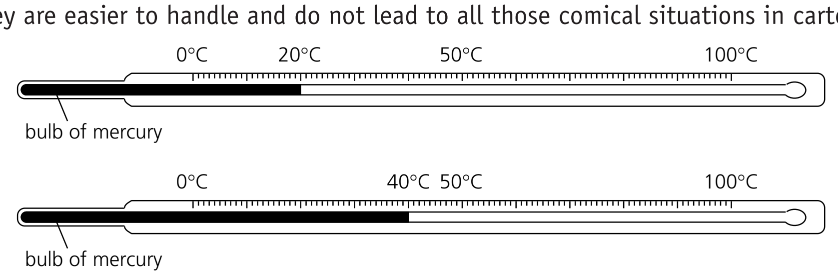


Figure 11.9
Expansion of mercury in a thermometer.

Ch11-Walding 4th 25/8/04 11:45 AM Page 264

11.8 THERMAL EXPANSION OF SOLIDS

With very few exceptions all solids expand when they are heated and contract when they are cooled. But different substances expand at different rates. The rates at which solids expand can be found experimentally.

If we take a 1 m length of aluminium and heat it so as to change its temperature by 1°C, its length will expand by 23.8×10^{-6} m. This may not seem very much, and it is not, for everyday purposes, but when fine measurements are needed or the length is much longer or the temperature change much greater, thermal expansion is very significant.

The change in length of a 1 m length of a substance due to a temperature change of 1°C is called the **coefficient of linear expansion** (α). For example, the coefficient of linear expansion of Pyrex glass is 3.3×10^{-6} m °C^{-1}. This means that a 1 m length of Pyrex glass increases by 3.3×10^{-6} m if heated so as to change its temperature by 1°C. The coefficients of linear expansion of several common substances are given in Table 11.1.

This expansion is more significant if we have greater lengths of materials or the temperature changes by more than 1°C.

For example, if we have a 4 m length of aluminium and change its temperature by 1°C, its length will increase by:

$$4 \times 23.8 \times 10^{-6} \times 1 \text{ m} = 9.52 \times 10^{-5} \text{ m}$$

Table 11.1 COEFFICIENTS OF LINEAR EXPANSION OF COMMON SUBSTANCES

SUBSTANCE	COEFFICIENT OF LINEAR EXPANSION $\alpha \times 10^{-6}$ m °C^{-1}
Diamond	1.2
Glass (Pyrex)	3
Glass (crown)	9
Platinum	9
Steel	10
Iron	12
Brick and concrete	12
Copper	17
Brass	19
Silver	18.8
Aluminium	23.8
Zinc	26.3
Rubber	80

If we have a 1 m length of aluminium and increase its temperature by 100°C, its length will change by 23.8×10^{-6} m every change of 1°C, giving a total increase in length of:

$$23.8 \times 10^{-6} \times 100 \text{ m} = 23.8 \times 10^{-4} \text{ m}$$

Changing the temperature of a 4 m length of aluminium by 100°C causes a change in length of:

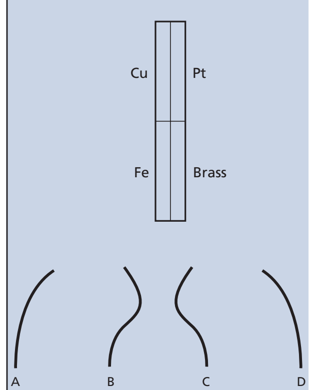
$$23.8 \times 10^{-6} \times 4 \text{ m} \times 100\text{°C} = 9.52 \times 10^{-3} \text{ m} \\ = 9.52 \text{ mm}$$

NOVEL CHALLENGE

When water freezes it expands and can crack pipes; therefore water can do work when it freezes. Where does this energy come from, especially as heat is being removed? Sounds stupid!

NOVEL CHALLENGE

When a rod made up of four metals as shown in the diagram is heated, which of the following diagrams represents its final shape.
What about if it is cooled?



This may have important consequences in everyday activities.

The reverse is also true. Cooling a 4 m length of aluminium from 100°C to 0°C will cause it to contract by 9.52 mm.

This expansion may not seem very much but if the length is big enough and the temperature rise large enough the expansion will be noticeable. For example, in the early 1800s, steam had just been introduced to power the factories. Steam pipes in the cotton mills were often over 400 feet (130 m) long and with temperature rises from a cold 10°C to 400°C the increase in length was such that a carpenter's ruler could be used to measure it. Leading industrialists of the time were enthusiastic in promoting the type of learning needed to deal with the new technology of steam. They saw the academic approach of the universities of the time as being useless in factories. Hence, they supported the establishment of craft guilds and mechanics' institutes to teach 'real-life' knowledge. Universities have changed a lot since then.

Expansion formula A formula to help find the change in length of a substance due to thermal expansion of solids is:

$$\Delta L = L_i \alpha \Delta T$$

where ΔL is the change in length, L_i is the original length of the solid, ΔT is the change in temperature.

It also should be noted that because a solid has three length measurements — length, width and height — all three expand and contract, therefore the volume of a solid changes with temperature change.

This property of solids to expand and contract with temperature change can be an advantage as well as a disadvantage in everyday life.

Can you think of some advantages and disadvantages?

Have you worked out why hot rivets are used yet? This is an advantage.

Other examples where these properties have to be considered include:

- The fitting of gear wheels to axles. If the axle is cooled in liquid nitrogen it contracts and the gear wheel will slip on more easily. When the axle warms up to normal temperature it makes a very tight fit.
- Telephone and electrical cables are hung loosely between poles to allow for contraction in cold weather conditions.
- Bimetallic strips that consist of two dissimilar metals of equal length are used in fire alarms. (See Figure 11.10.)
- Bridges and rail lines have expansion gaps to allow for expansion in hot conditions to stop buckling.
- Next time you pass a large building or concrete paths check for rubber expansion gaps that allow for the expansion of the concrete, to stop cracking.

NOVEL CHALLENGE

The General Electric building in New York has thousands of slabs of Italian Travertine marble bolted on as a skin. Each slab is 6000 mm x 3000 mm. The temperatures in NY can vary from −23°C to +38°C, over which temperature range the length of the slabs increase by 3 mm.

What is the coefficient of linear expansion of marble? How would these slabs be attached to the building so that they didn't crack on expansion or contraction? If a slab was bolted at each corner, how far away from the face of the building would it bow out? Do any buildings in your city have them? Email the authors and it will be added to the next edition.

NOVEL CHALLENGE

In the old days before welding was invented, ships were made by riveting steel sheets together. Rivets are small lengths of steel with a flat head on one end. They were heated to a high temperature and inserted through a hole drilled in the sheets. The rivets were then struck with a hammer to bend the other end over. Why were they heated? Wouldn't they shrink away from the insides of the hole?

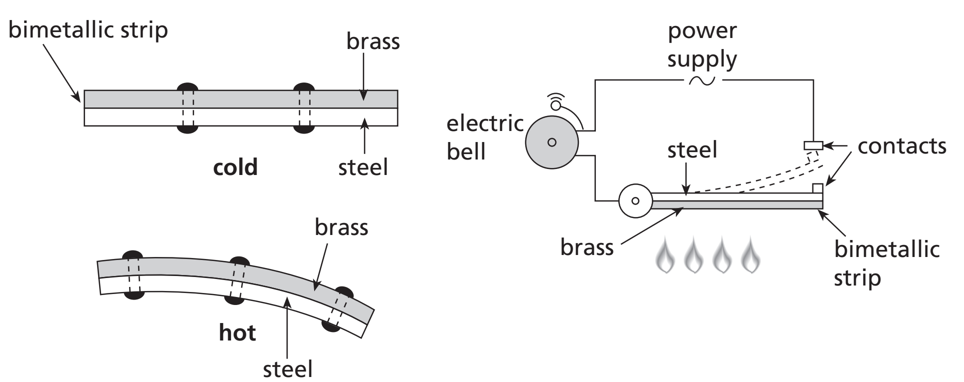


Figure 11.10 Bimetallic strip used in a fire alarm.

- The coefficients of expansion of the iron used for reinforcing concrete and of the concrete itself have to be similar. If they expand at different rates the concrete will crack, and with continual expansion and contraction it may break away from the iron.
- Fillings in teeth and the teeth themselves need to have similar coefficients of expansion. Why?
- Crown glass shatters when you pour boiling water into it but Pyrex does not.
- In aircraft manufacture, rivets are often cooled in dry ice before insertion and then allowed to expand to a tight fit.
- Pipes in refineries often include an expansion loop so that the pipe will not buckle as the temperature rises.

NEI Activity 11.3 THERMAL EXPANSION

Houses with steel roofs on a timber frame will creak when a cloud passes overhead on a hot summer's day. What's going on here? In your response, you should provide quantitative data to support your claim.

Example

An electric company strung an aluminium wire between two piers 200.0 m apart on a day when the temperature was 25°C. They strung it tight so that it would not sag. Find the length of the wire when the temperature fell to –25°C on a cold winter's night? What might happen? What should have been done to prevent this occurring? This does happen in countries that have wide variations in temperature.

Solution

$$\begin{aligned} \Delta L &= L_i\ \alpha\ \Delta T \\ &= 200.0\text{ m} \times 23.8 \times 10^{-6} \times 50°\text{C} \\ &= 2.38 \times 10^{-1}\text{ m} \\ &= 23.8\text{ cm} \\ \therefore L &= 200.0 - 0.238 \\ &= 199.862\text{ m} \end{aligned}$$

PHYSICS UPDATE

Modern train lines are welded together, so they have no expansion gap. How is expansion allowed for so that the lines don't buckle? In the past, fairly light wooden sleepers were used but today heavy steel or concrete sleepers (300 kg) are used and this physically prevents expansion. The force of expansion builds up but it is insufficient to lift the weight of the tracks and sleepers so they just get compressed. If you don't think this can be true, ring *Queensland Rail* and talk to a track engineer.

NOVEL CHALLENGE

A steel ruler has a hole in one end. When the ruler is heated, does the hole get bigger, smaller or stay the same? Be careful — even some science texts get it wrong.

Questions

20 A 3.0 m steel girder is used in the construction of a bridge. If the least and greatest recorded temperatures in the region where the bridge is to be built are –10°C and 45°C, what allowance would need to be made for expansion and contraction, if it was fixed in place when the temperature was 20°C?

21 A very modern car engine is to be made of ceramic material with a coefficient of expansion of 1.8×10^{-6} m °C^{-1}.
- **(a)** What would be the increase in the diameter of the 5.0 cm piston if the temperature of the engine changed from 10°C to 180°C?
- **(b)** How does this compare with an engine made of iron?
- **(c)** How do normal car engines allow for this expansion?

22 An expansion gap between 5.0 m iron railway lines was set at 2.0 mm on a day when the temperature was 10°C. To what temperature would the lines have to increase before they begin to buckle?

23 A 100 m long iron water pipe was laid in the ground on a hot summer's day when the temperature was 25°C. By what length would it have contracted if the ground temperature fell to –15°C?

THERMAL EXPANSION OF LIQUIDS 11.9

A very common device making use of the expansion of liquids is a thermometer. As the temperature increases, the mercury or alcohol in the thermometer increases in volume and moves up the fine tube. Other examples include the explosion of bottles filled with liquid and left in the hot sun.

As liquids take the shape of the container we are mainly interested in the volume changes of liquids with temperature. Again these changes can be found experimentally. The **coefficients of volume expansion**, β, of some common liquids are given in Table 11.2.

Table 11.2 COEFFICIENTS OF VOLUME EXPANSION OF LIQUIDS

LIQUID	COEFFICIENT OF VOLUME EXPANSION $\beta \times 10^{-4}$ m °C^{-1}
Mercury	1.82
Water	2.07
Petrol	9.55
Turpentine	9.73
Alcohol	11.2
Acetone	14.87
Ether	16.56

As with solids, the change in volume of liquids can be found by the formula:

$$\Delta V = \beta\, V_i\, \Delta T$$

where ΔV is the change in the volume of the liquid, β is the coefficient of volume expansion, V_i is the initial volume of the liquid, ΔT is the change in temperature.

Example

What would be the increase in the volume of 0.20 L of acetone if it was heated from 10°C to 40°C?

Solution

$$\begin{aligned}\Delta V &= \beta\, V_i\, \Delta T \\ &= 14.87 \times 10^{-4} \times 0.20 \times (40 - 10) \\ &= 89.2 \times 10^{-4}\ \text{L} \\ &= 8.9 \times 10^{-3}\ \text{L}\end{aligned}$$

Questions

24 By what volume would 25 L of alcohol increase if its temperature was increased from 10°C to 50°C?

25 How much extra petrol would you get if you bought 50 L at 5.0°C instead of at 40°C?

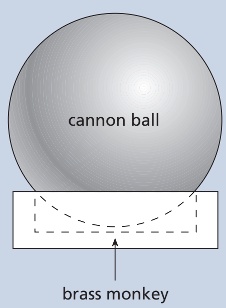
26 A student measured 500 mL of water into a measuring cylinder at a temperature of 25°C and placed it in the refrigerator where the temperature was 4.0°C. What will be the measurement on the measuring cylinder? (Assume the cylinder does not contract.)

NOVEL CHALLENGE

The coefficient of volume expansion (β) for iron is 0.36×10^{-4}.
How many times greater is this than the coefficient of linear expansion (α)? Propose a proof for this.

NOVEL CHALLENGE

In days gone by, warships had cannons mounted on the decks with the iron cannonballs resting in shallow brass ashtray-shaped containers called 'brass monkeys'. These would invariably fill with water as the sea washed over the decks. In the Atlantic ocean sometimes it would get so cold that sailors would say: 'It's cold enough to freeze the balls off a brass monkey'.
What do you suspect they meant?

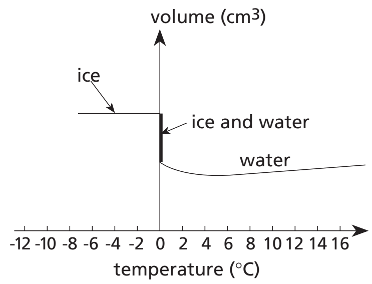


11.10 THE ABNORMAL EXPANSION OF WATER

Most liquids contract with a decrease in temperature. Water is different. Water in fact expands as it cools below a certain temperature. Why do soft drink bottling companies leave air at the top of bottles of soft drink? But expansion also is seen if a bottle of soft drink is left in a freezer. Consider the graph (Figure 11.11) showing the change in volume of water with temperature changes.

As water is cooled it contracts as expected with all liquids; however, at approximately 4°C it stops contracting. Between 4°C and 0°C it actually expands as temperature decreases. This is due to the rearrangement of the particles that make up water. This rearrangement takes up more volume. A certain amount of water therefore has a minimum volume and a maximum density at 4°C. When water freezes at 0°C it undergoes considerable expansion. The volume of 100 mL of water changes to 109 mL of ice. This is very important in cold countries. Unless water pipes are very well insulated the water in them will freeze and this may cause the pipes to burst.

Figure 11.11
The change in the volume of water with temperature.



This abnormal (or anomalous) expansion of water can cause many problems:

- In cold climates the water in the engine block and radiator of a car can freeze and the expansion can shatter the engine. 'Antifreeze' is usually added to the water to lower its freezing point and to prevent freezing from occuring. It also raises the boiling point. Can you see the use of this?
- When water freezes in pipes in your house it can cause them to crack. Some people leave taps dripping to prevent this happening. Do you think that this would work?

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***27** A cylinder contains 0.50 m^3 of helium gas at 2.0 atm. What volume of gas is able to escape if it is released into the atmosphere?

****28** An oxygen cylinder releases gas at the rate of 4 m^3 h^{-1} at atmospheric pressure. If it is sold at a pressure of 200 atm and has a volume of 2 m^3, for what time will it supply oxygen?

***29** A container of gas has a pressure of 80 cm of Hg when its volume is 800 cm^3 and its temperature is 60°C. What will be its pressure when its temperature is increased to 90°C, and its volume is reduced to 400 cm^3?

****30** Two cylinders of volumes 2.0 m^3 and 3.0 m^3 are at pressures 4.0 atm and 6.0 atm respectively. If they are then joined by a thin, short tube, what will be the new pressure in each?

***31** What is the effect on the volume of a gas **(a)** whose pressure is tripled at the same time as its temperature is halved; **(b)** whose pressure is kept constant while twice the number of molecules are added at the same temperature?

***32** What volume will two moles of gas occupy at a pressure of 8.2×10^4 N m^{-2} and a temperature of 290 K?

***33** A container has a volume of 20.0 L at a temperature of 360 K. Gas is forced into the container until the pressure is 2.0×10^5 N m^{-2}. How many molecules of gas are there in the container now?

*34 A flask is open to the atmosphere at room temperature of 26°C. To what temperature must the flask be heated before only two-thirds of the original number of molecules remain in the flask?

*35 A cylinder has a volume of 1.5 m^3 and contains neon gas at a pressure of 1.0×10^{-1} cm height of mercury at the temperature of 20°C. What is the number of particles in the cylinder?

**36 A flask contains gas at a temperature of 30°C and a pressure of 2.0×10^3 Pa. Find the number of molecules per cubic metre in the flask.

**37 If we have a cylinder of volume 4.4 m^3 containing 5.0×10^{22} molecules at a pressure of 3 atm, what will be the average kinetic energy of the centre of mass motion of the molecules?

*38 Find the average kinetic energy of translation of a molecule at 24°C.

*39 How much energy is needed to raise the temperature of 1 mol of a monatomic gas by 10°C?

*40 A steel water pipe line of 2000 m is fixed in place on a day when the temperature was 30°C. What will be its new length when the temperature drops to −20°C?

*41 The space between 10 m steel railway lines is 8.0 mm at 5.0°C. What would be the space at 30°C?

**42 Mercury has a density of 13.6 g cm^{-3} at 20°C. What would be the density of mercury at 100°C?

**43 The data given in Table 11.3 were collected by heating a 50 cm copper rod in a water bath. From the data calculate the coefficient of linear expansion for copper.

Table 11.3

TEMPERATURE	LENGTH (CM)
20	50.000
40	50.017
60	50.033
80	50.052
100	50.070

*44 Explain why iron and not steel is used to reinforce concrete.

Extension — complex, challenging and novel

***45 Twenty-five percent of the energy put into a certain non-monatomic gas causes increased rotation and vibration of the atoms within the molecule. How much energy is required to raise the temperature of 2 mol of this gas from 10°C to 35°C?

***46 If 2 mol of helium at 50°C is mixed with 4 mol of argon at 20°C, find the final temperature of the mixture.

***47 Gas X is monatomic, and gas Y has one-quarter of its total energy involved as energy within the molecules.

(a) If 2 mol of X at 70°C is mixed with 1 mol of Y at 25°C, find the final temperature of the mixture.

(b) If 2 mol of Y at 80°C is mixed with 2 mol of X at 35°C, find the final temperature of the mixture.

***48 A bubble expands to three times its original volume while rising from the bottom to the surface of a lake.

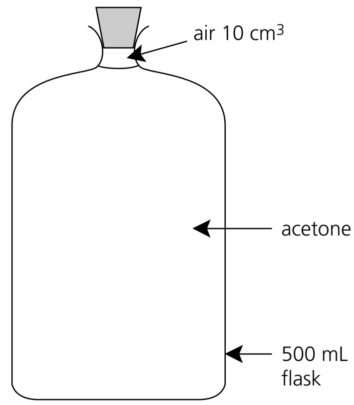
(a) Assuming that the lake throughout is at the same temperature as the surrounding atmosphere, how deep is the lake?

(b) If there had been a temperature increase from 7.0°C at the bottom to 27°C at the surface, by what factor would the bubble have expanded?

***49 Two glass flasks, one of which has twice the volume of the other, are connected by a thin tube of negligible volume. They contain dry air at a temperature of 20°C and a pressure of 76 cm of mercury. The larger flask is then immersed in steam at 100°C and the smaller in melting ice at 0°C. Neglecting any change in volume of the flasks, find the resulting pressure in them.

***50 A 500 mL flask as shown in Figure 11.12 was filled almost to the top with acetone at 20°C. It was inadvertently left on the bench in the sunlight on a hot day when the temperature reached 32°C. Assuming no acetone changed to vapour, calculate the pressure of the air in the neck of the flask. (Neglect the expansion of the flask itself as this is small compared with the expansion of acetone.)

Figure 11.12
For question 50.



***51 A new type of temperature scale has been created by Martians. Measurements of a gas held at constant pressure give the following data. The temperature is in degrees Martian (°M) which has the same size *degree* as in the Celsius and Kelvin scales:

Temperature (°M)	90	120	150
Volume (L)	30	45	60

Create an equation that converts a temperature on the Martian scale into a temperature on the Celsius scale. It should look something like the Celsius/Kelvin conversion equation (K = °C + 273).

°M = ?

CHAPTER 12

Heat Transfer

INTRODUCTION 12.1

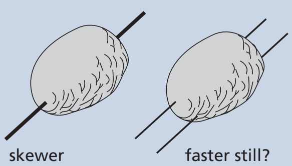
The transfer of heat energy from one place to another may seem unimportant to many. How does this affect me? What use is this to me? However, like many applications of physics, heat energy transfer unknowingly affects our everyday life more than a casual glimpse would suggest.

- On cold winter nights what keeps you warm? Why does that quilt, particularly a down quilt, keep you warm?
- Why do saucepans have plastic, wooden etc. handles? Would it not be better if they were all steel or aluminium? They would be easier to manufacture and clean.
- How do you feel the warmth of the electric heater from across the room? Could electric heaters be used in outer space to keep astronauts warm?
- If you were interrupted while making a cup of coffee, would it be better to leave it before putting in the cold milk, or put the milk in before you do that little job? In which case would the coffee be hotter when you return?
- How does the Sun's heat energy reach the Earth?
- Why are there heat shields on the Space Shuttle?
- Double glazing of windows is very beneficial in the conservation of energy for households or for large buildings. Why? (If you don't know what double glazing is, check the encyclopaedia or look in any building magazine.)

All the above examples have something in common. They all can be explained by the understanding of heat energy transfer. There are several ways in which heat energy can be transferred. Let's look at them in turn.

NOVEL CHALLENGE

Cooks sometimes put a metal skewer through potatoes to make them cook more quickly. *Would you speed things up by using a skewer of twice the diameter, or two of the smaller skewers? If you used two, where is the optimal place to put them? Why?*



NOVEL CHALLENGE

When you get out of bed in the morning, carpet feels warmer under your feet than tiles. Why is this if they are both at the same temperature?

Activity 12.1 HEAT LOSSES

SR

NEI

You can probably think of many more situations where the loss, gain, or transfer of heat energy from one place to another plays a role in our everyday life. For example:

1. Why doesn't the Earth get hotter and hotter as sunlight falls on it? How does the term 'albedo' apply to this situation?
2. Computer CPUs have big metal 'heat sinks' with large-surface-area fins attached. What is the purpose of this?
3. The bony plates on the back of a stegosaurus have been claimed to be part of its cooling system. How might they work? Research this and discuss arguments for and against this proposal using 'discussion' genre.

12.2 CONDUCTION

Conduction (from the Latin *conducere* meaning 'to lead together') is the process by which heat energy is transferred through a medium by the vibrating particles of the medium, but without the particles actually moving. For example, when a metal teaspoon is placed in hot water the handle becomes hot. Heat energy travels from the hot water through the spoon to your hand. The reason this occurs has already been suggested in Chapter 10. (This may be the time to revise this section.) The molecules of the hot water are moving faster than those of the spoon — they have more energy because they are hotter. When they collide with the particles of the spoon they transfer some of their energy to those particles of the spoon. These molecules then collide with others adjacent to them. This continues until all the molecules of the spoon and water are in equilibrium. Heat energy is thus transferred from the hot water to the spoon and eventually to your hand. Of course, you might say that the spoon's handle does not get as hot as the water. This is true. But where else is the spoon's handle transferring some of its energy? The air around it has molecules! Notice that the energy is transferred from the hot water to the spoon and your hand but the particles themselves do not move. They may vibrate but they do not move with the transfer of heat energy.

So to transfer heat energy by conduction the medium must contain particles and the closer together the particles the better. Therefore solids, liquids and gases can **conduct** heat energy, but a vacuum cannot.

(From now on we will refer to heat energy transfer as heat flow, which is a simpler way of expressing the idea of a transfer of heat energy from one medium to another.)

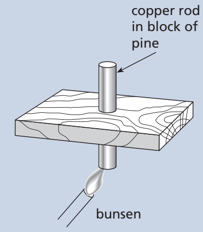
This would also suggest that solids are better conductors than liquids, which are in turn better than gases. This, in general, is true, as the particles in most solids make closer contact with each other than those of liquids or gases. Table 12.1 indicates the rate of heat flow through particular materials. It will be noticed that this table reinforces the above statement. This will be discussed more fully later. The table might also suggest why copper-based saucepans are better than iron-based saucepans.

Table 12.1 THERMAL CONDUCTIVITY OF SOME MATERIALS

MATERIAL	THERMAL CONDUCTIVITY, k (W m^{-1} K^{-1})
Silver	430
Copper	400
Aluminium	240
Brass	105
Iron	67
Steel	46
Concrete	0.8
Glass	0.8
Brick	0.6
Water	0.6
Asbestos, paper	0.2
Rubber	0.2
Plasterboard	0.13
Wood	0.08
Cork	0.05
Carpet	0.05
Bone	0.042
Fibreglass wool	0.04
Plastic foam	0.03
Air	0.024
Fat	0.021

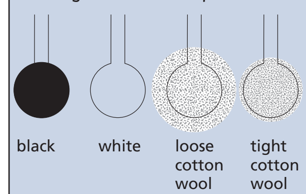
NOVEL CHALLENGE

A copper rod is placed through a hole in a piece of pine and heated. Charring occurs more along the grain that across it. *Now why is this? Propose a physics explanation.*



NOVEL CHALLENGE

Four thermometers as shown are placed in the Sun for 10 minutes. List them in order from highest reading to lowest. Explain.

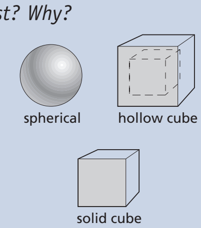


NOVEL CHALLENGE

How can you cook a hamburger thoroughly in the shortest time? Would you cook it on an open grill (large heat, but some charring) or in a pan (small heat). Explain using physics principles. Suggest to your physics teacher that you have an end-of-term BBQ and that the school pay for the hamburger patties. Good luck!

NOVEL CHALLENGE

You have three ice-cubes of the same mass. *Which one will melt first? Why?*



Ch12-Walding 4th 25/8/04 11:46 AM Page 273

Activity 12.2 CONDUCTIVITY OF LIQUIDS

Put some ice in a test-tube and hold it in place with some steel wool (Figure 12.1). Half-fill the test-tube with water. Hold the upper part of the tube over a candle or a Bunsen burner until the water boils. What do you notice about the ice? What does this suggest about the conductivity of liquids?

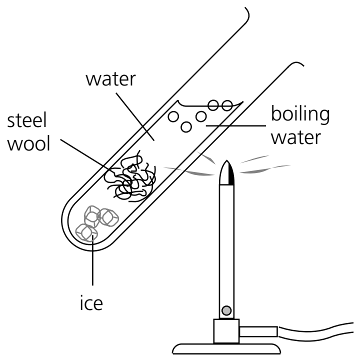


Figure 12.1
The water boils but the ice remains because water does not conduct heat very well.

Bonding

The bonding of the atoms in materials controls how easily the atoms vibrate and therefore conduct. The bonds between the atoms in metals allow the atoms to vibrate freely in all directions, whereas the bonds in non-metals hold the particles more firmly, and are more rigid, thus not allowing the particles to vibrate as freely. So metals are good **conductors** whereas non-metals are poor conductors, or **insulators**. This again is shown in Table 12.1.

Both good conductors and poor conductors (insulators) have their uses. Good conductors are used for such things as the bases of saucepans, car radiators, cooling fins on air-cooled engines such as those used in VWs, and as heat sinks on semiconductor electronic devices. Poor conductors are used to insulate roofs, insulate water pipes in cold countries, and for jumpers, wet suits, and the handles on pots and pans.

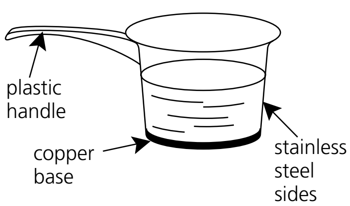


Figure 12.2
Copper-based saucepans conduct heat well whereas poor conductors are good for handles.

Staying cool or hot

A special mention has to be made of those materials (many synthetic) that are poor conductors because they trap air within their fibres. Since air is a poor conductor (Table 12.1), materials that trap air do not transfer heat energy very well. Materials such as wool, fur, polystyrene, carpet, fibreglass fibres, etc. all have these qualities. Fibreglass or wool insulation is used in the ceilings of houses as it does not allow the heat energy to be transferred readily from the atmosphere to the interior on hot days or the reverse on cold days, thus improving the living conditions within the house and reducing the cost of heating or cooling. Carpeted floors always feel warmer then wooden or concrete floors on cold mornings. Carpet reduces the rate at which heat is lost from your feet to the floor, therefore your feet will retain their heat longer and feel warmer, except for the loss of heat to the atmosphere — to stop this you had better wear slippers (woollen ones).

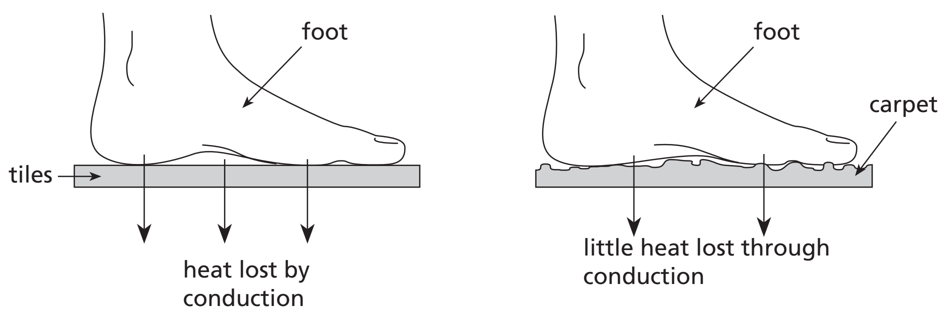
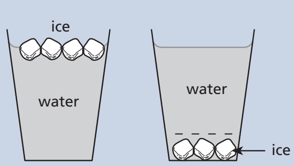


Figure 12.3
Carpet feels warmer than concrete because concrete conducts the heat from your feet more rapidly.

NOVEL CHALLENGE

Which will cool the water more quickly — leaving the ice to float or keeping it submerged? Provide the physics principles behind this.



Rate of heat flow

Table 12.1 indicates that heat energy is transferred through materials at different rates. Heat reaches your hand quickly when the ends of some metals are placed in a Bunsen flame, while other materials such as wood do not transfer the heat energy nearly as fast or as readily. The rate of heat flow depends on several properties of the material. The rate of heat flow (R) is defined as the heat energy transferred per second, and is measured in joules per second or watts.

What do you think controls the heat flow from the stove through the bottom of a copper-based saucepan to the water in the saucepan?

Commonsense would suggest several factors control the heat flow. These have been verified through experimentation:

- Experiments have shown that if the material were thicker, heat would take longer to pass through. That is, if our saucepan's base was thick it would take longer for heat to penetrate. The rate of heat flow (R) is inversely proportional to the thickness (d).

$$R \propto \frac{1}{d}$$

- The rate of heat flow has also been shown to be dependent on the temperature difference across the material. That is, if the temperature difference on either side of the saucepan's base is greater, heat will flow more quickly.

$$R \propto (T_2 - T_1)$$

- The area of the material influences the flow rate. The greater the area the quicker the energy transfer.

$$R \propto A$$

- The rate of heat flow also depends on the type of material. As already suggested, heat flows readily through some materials and not so readily through others. The **thermal conductivity** of a material (k) is a measure of the rate of flow of heat energy through 1 m^2 of a material of thickness of 1 m and having a temperature difference of 1 K between the sides. For example, copper has a thermal conductivity of 400 W m^{-1} K^{-1}, which means 400 joules per second flow through a 1 m square, 1 m thick piece of copper when there is a temperature difference of 1 K between the sides.

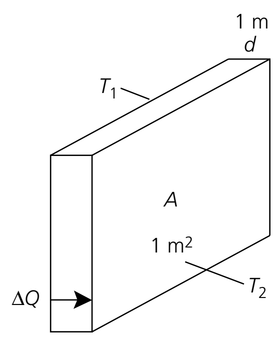


Figure 12.4
Heat flow through a piece of copper 1 m^2 × 1 m.

Putting all the above variables together the rate of heat flow becomes:

$$R \propto \frac{(T_2 - T_1) \times A}{d}$$

then

$$R = \frac{k(T_2 - T_1)A}{d}$$

The units for thermal conductivity, k, result from the above equation:

$$k = \frac{Rd}{(T_2 - T_1)A}$$

The units thus become W m/K m^2, or W m^{-1} K^{-1}.

Table 12.1 gives the thermal conductivity of several common substances. This table is worth more than a passing glance. It indicates why plastic, wood etc. are used for handles of saucepans; why fibreglass, wool and plastic foam are used for insulating in houses; why cork is used for place mats for hot materials; and why animals that live in very cold climates have a great deal of fat on their bodies.

NOVEL CHALLENGE

Does dark water (e.g. tea) cool more quickly or more slowly than white water (e.g. milk in water)? Explain this before you try it. While you're thinking about it, here's another: does brown bread toast more quickly than white bread?

INVESTIGATING

Have you ever been over the Gateway Bridge in Brisbane and noticed the huge storage tanks for BP oil?
The ones filled with unrefined oils or lubricants are painted black, whereas the tanks filled with petrol are painted silver. Propose a reason for this and if you are able to, find out why they really do it. We had to ring BP in Melbourne.

Example

Calculate the initial rate at which heat flows through a copper-based saucepan that has a 15 cm diameter, 1 cm thick, base. The temperature of the water in the saucepan is initially 18°C while the stove hotplate is 120°C.

Compare this with a saucepan made of steel.

Solution

$$R = \frac{k(T_2 - T_1)A}{d}$$

$$= \frac{400(120 - 18)\pi 0.075^2}{0.01}$$

$$= 7.2 \times 10^4 \text{ W}$$

For the steel saucepan:

$$R = \frac{k(T_2 - T_1)A}{d}$$

$$= \frac{46(120 - 18)\pi 0.075^2}{0.01}$$

$$= 8.3 \times 10^3 \text{ W}$$

The flow of heat energy through copper is approximately 10 times that of steel. This should have been seen from Table 12.1, since all other variables were the same.

Photo 12.1
The Texas TI-83 graphing calculator and the CBL (Computer Based Laboratory) are another useful way of accumulating temperature/time data.

NOVEL CHALLENGE

80 mL of cold water is placed in a polystyrene cup and the cup is placed in a beaker of hot water. Thermometers are placed in both containers. Predict the shape of the temperatures versus time graph. The experiment is repeated but a small cube of ice is placed in the cold water. Now show how the graph shapes will change (in red ink). If you had a TI-83 graphing calculator with temperature probes (see photo 12.1), you could follow the progress.

Questions

1. When one end of a piece of glass rod is placed in the flame of a Bunsen burner the other end becomes hot.
   (a) Explain how the heat energy travels from one end to the other.
   (b) What other laboratory materials would transfer heat faster?
2. Many birds on cold winter mornings are seen to 'fluff' their feathers. What is the purpose of this?
3. **(a)** Calculate the rate at which heat energy is lost through a 1.0 m^2 laboratory window on a day when it is 15°C on the outside and 25°C inside. The glass is approximately 5.0 mm thick. **(b)** How much heat energy is lost in 1 hour?
4. Calculate the heat lost from a seal in 30 minutes. Assume the seal's total surface area is 1.1 m^2 and the thickness of the fat is 2.0 cm . The atmosphere temperature is –25°C and the seal's body temperature is 37°C.

Fire-walking

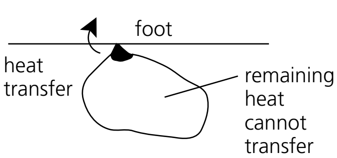
Though fire-walking has long been associated with Far East mysticism it has recently been adopted by the New Age movement in California. It has also recently come under the scrutiny of physicists in search of an explanation for how people can walk on 600°C red-hot coals in bare feet and not get burnt.

To obtain the coals a pile of wood is set on fire and allowed to burn to red-hot over a period of an hour or two. The bed, about 4 m long and 1 m wide, is raked evenly and when paper is thrown on it will burst into flames. People can then walk over it, taking about 7 seconds, without getting burnt. How do physicists explain this?

Two effects are in operation:

- poor thermal conductivity of coals.
- the Leidenfrost effect.

Figure 12.5
Coals provide a poor conductor and there is little surface area in contact with the feet.

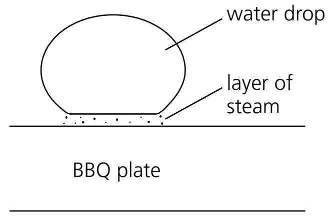


Coal is a poor conductor of heat, unlike, say, a steel barbecue plate. The heat leaves the edge of the coal but does not transfer to the feet very quickly.

The other effect is called the Leidenfrost effect. If you have ever poured water on a very hot barbeque plate, you would have noticed that the water forms little drops and they dance around on a layer of steam on the plate (Figure 12.6).

In a similar way this layer of steam keeps the skin away from the coals and protects the foot even more. The Australian Skeptics often arrange demonstrations of fire-walking to debunk the mysticism of the process. There is no magic in it — just pure physics. But do not try it! People have been burnt.

Figure 12.6
Water bubbles floating on steam on a barbecue plate.

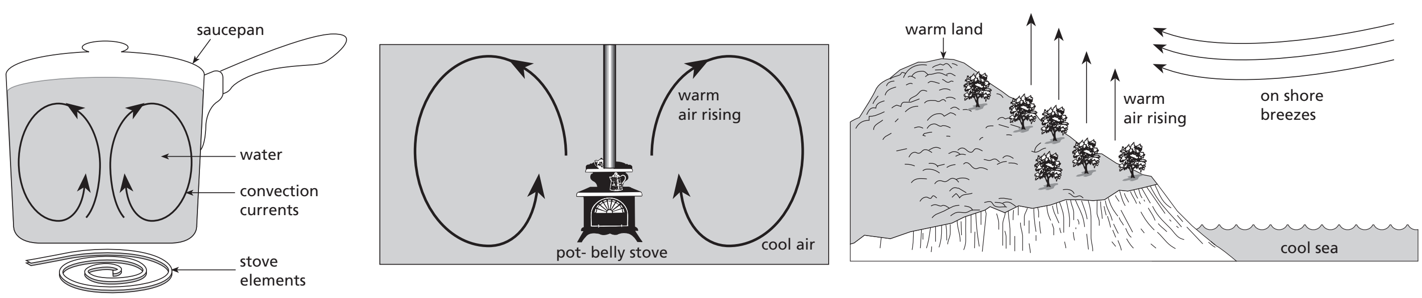


12.3 CONVECTION

The second method of heat transfer is by the **convection** process. (The word convection comes from the Latin *convehere* meaning 'to carry together'.) Convection is similar to conduction but in this case the particles of the materials themselves actually move. While conduction is the transfer of heat by the vibration of particles of the material, convection is the transfer of heat by the movement of particles. As solid particles do not move, convection is confined to liquids and gases.

Convection is used to explain how pot-belly stoves heat rooms, why fireplaces 'draw' properly, how water is heated in saucepans, and how onshore and offshore sea breezes develop.

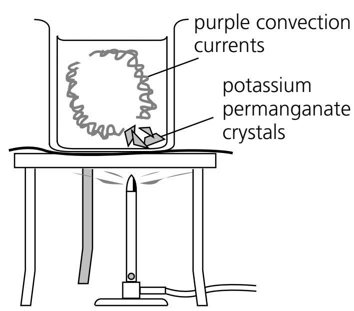
Figure 12.7
Convection. **(a)** Convection currents in a saucepan of water; **(b)** a pot-belly stove warms a room by setting up convection currents; **(c)** convection currents arise because the difference in temperature creates onshore breezes.



For example, when a pot-belly stove is placed in the centre of the living room it heats the air in its immediate vicinity. This air expands, becoming less dense and thus rising. Cooler surrounding air moves in to replace the hotter air that rises. The hot rising air cools as it goes higher and therefore recirculates, as shown in Figure 12.7**(b)**. Convection currents are thus set up. A similar process happens when water is heated in a saucepan on the stove.

Convection can be demonstrated very effectively in the laboratory. If one or two crystals of potassium permanganate are placed in a beaker of water as shown in Figure 12.8 and heated, the crystals dissolve as the surrounding water becomes hot. Purple convection currents of potassium permanganate solution are formed.

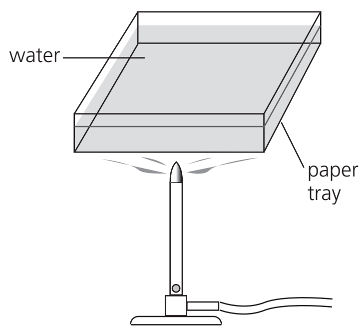
Figure 12.8
Heating potassium permanganate crystals produces observable convection currents.



EI Activity 12.3 BOILING WATER IN A PAPER BAG

Make a small paper container out of paper (Figure 12.9). Put some water in it and hold it over a candle. You can boil the water without burning the paper. Describe these effects in terms of conduction and convection.

Figure 12.9
Water will boil in a paper tray when heated.



Activity 12.4 RESEARCH

Research one of the following and be prepared to explain your research to the class:

- Count Rumford spent his summer holidays in 1794 by the Italian beach. Instead of swimming he investigated convection currents. What did he find out?
- How are convection currents associated with afternoon onshore breezes and morning offshore breezes?
- How do gliders and hang-gliders use convection currents ('thermals')?
- How is convection used in solar hot water systems?
- What is the difference between convection ovens and fan-forced ovens?

In cold countries many homes are heated by convection (Figure 12.10). Proper design of convection systems allows the hot air to circulate and the cool air to return to the furnace to be reheated. The rush of air this creates near the vents can be very noticeable.

Figure 12.10
Convection currents are used to circulate heat throughout houses.

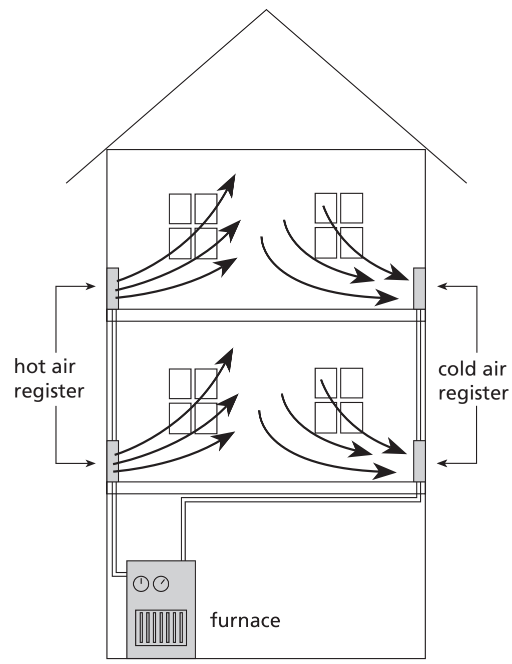


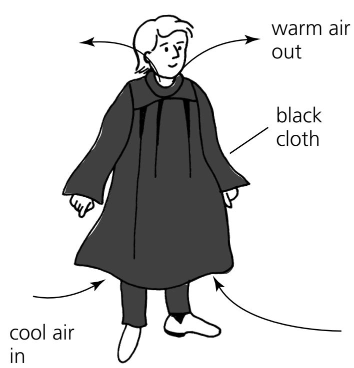
Photo 12.2
Hang-glider and thermals.

Some currents formed in the oceans, for example the gulf stream and the Japan current, are large-scale examples of convection.

Staying cool in the desert

In even the hottest regions of north Africa, nomadic travellers wear black loose-fitting clothing to keep cool. We wear white to keep cool; so what is going on? The answer is that the black cloth heats up the air between the cloth and the skin creating an updraft, which draws cool air in as the warm air exits through the neck opening. This keeps the people from overheating.

Figure 12.11
Loose black clothes set up convection currents to control the temperature.



Questions

5 What is the advantage of placing the heating element at the bottom of an electric kettle?

6 Why can heat energy from the Sun not be transferred to the Earth by conduction or convection?

12.4 RADIATION

In conduction and convection the vibration or the movement of particles results in heat energy transfer, but how does heat energy move between places where no particles exist? How does heat energy travel between the Sun and the Earth through the vacuum of space?

The heat is transferred by a process called **radiation**. The word radiation is from the Latin *radiate* meaning 'to emit beams'. This process involves the movement of heat energy by waves — **electromagnetic waves**. The properties of waves will be discussed in Chapters 13, 14 and 15.

SR Activity 12.5 HOT BULBS

Put your hand on the glass of a light bulb that is turned off (and is cool). Turn the light on and immediately you will feel the heat from the electromagnetic radiation. Turn if off. Is the glass hot? Explain.

All bodies radiate energy in the form of electromagnetic waves whose wavelength is in the infrared region. The wavelengths of these waves are longer than those of visible light and therefore cannot be seen. However, they can be detected. Most people have seen films in which various pieces of apparatus are used to detect the differences in infrared radiation given off by humans compared with those of the surroundings.

Hot bodies give off more of this radiation than cooler ones. The wavelength of the infrared radiation emitted depends on the temperature of the radiating body. For example, a table in a darkened room can be photographed with infrared sensitive film while it cannot be seen by the naked eye. This property of bodies to emit infrared radiation has many applications in industry and medicine, as well as for the military:

- Tumours below the skin's surface have a higher metabolic rate than the surrounding tissue. They therefore produce more heat. Infrared **thermography** is used to map the infrared radiation given off by the tumour and surrounding tissue. Electronic processing can produce coloured pictures of a person's body.

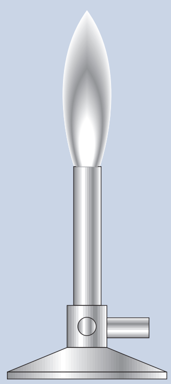
INVESTIGATING

For thirty years the science literature has reported that warm water freezes more quickly than cold water. This sounds like nonsense, but under certain conditions it will happen. It all has to do with the different convection currents. Your design should include: different conditions (lid on/off); different container (polystyrene/glass; tall/short); and a search for trends (try 40°C, 60°C, 80°C). It is a perfect experiment for using a TI-83 and CBL data-logger (or similar).

Photo 12.3
A thermogram. An infrared photograph shows the different amounts of heat emitted from different parts of the human body.

NOVEL CHALLENGE

The temperature around a Bunsen burner is lower the further you are from it. But does the temperature fall away evenly in all directions? Draw a Bunsen flame as in the diagram below and predict where points of similar temperature will be, keeping in mind that both convection and radiation are operating. Join these points of equal temperature — they are called 'isotherms' (Greek *iso* = 'equal'; *thermos* = 'heat'). Try it!



- Some animals such as rattlesnakes and some birds use special heat receptor cells to detect and track their victims. The sidewinder missile used by air force fighters to 'home in' on the heat emissions from the exhaust of enemy aircraft obtains its name from the rattlesnake (or sidewinder).
- The military has developed many applications of the detection of infrared radiation. In the Second World War infrared sensitive film was used to detect the damage done to enemy factories even when there was cloud cover. With the Vietnam conflict, sophisticated scopes for rifles were developed. These could pick up and amplify the body heat of the enemy, let alone the heat from a match or the glow of a cigarette in the mouth — not a good idea!
- Satellites using infrared photography are now in regular use. They have been used to detect troop movements and missiles being launched, as well as to detect hot underground rock formations. One coal-mining community in Pennsylvania in the USA disappeared when oxidation of coal seams caused huge underground caverns to cave in. Another community was saved when satellite controllers using infrared photography were able to give sufficient warning.

The Sun, a very hot body, radiates infrared heat energy, which reaches the Earth. Approximately 1400 joules of heat energy is arriving per second at every square metre of the upper atmosphere of the Earth. When this radiation strikes an object it causes the molecules of the object to gain energy and vibrate faster, resulting in the object becoming hotter. Some objects absorb this radiation better than others. Dark objects absorb radiation better than white objects, while white shiny surfaces also reflect heat.

- Why is it that sports-people who play sport in the Sun (cricket, tennis etc.) wear white clothing?
- What colour is the traditional clothing of desert peoples of north Africa?

NEI Activity 12.6 HOUSEHOLD APPLIANCES

It would be a worthwhile exercise to check the appliances around the house to identify those that rely on the reflection of radiation to improve their performance. Make a list of these and select one to explain its operation.

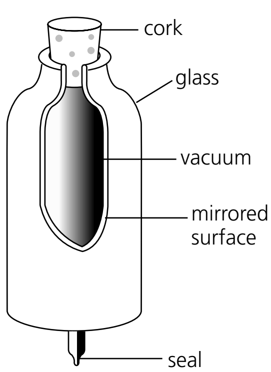
FURTHER APPLICATIONS 12.5

An understanding of conduction, convection and radiation is useful in the designing of better appliances and improving living conditions.

- A thermos flask or vacuum flask is a good example of how an understanding of heat energy transfer has produced a more efficient product.

  Double glass walls containing a vacuum reduce the loss of heat by convection and conduction. The walls are also silvered to reflect heat back into the flask and reduce the loss of heat by radiation. However, the fluid in a vacuum flask does change its temperature with time. Can you suggest where most heat loss occurs?
- Greenhouses have always been efficient at maintaining a warm environment for better plant growth. Shorter high energy infrared waves from the Sun pass through the glass walls and heat the plants and the soil. These in turn emit longer, lower energy infrared waves that cannot penetrate glass. Thus the inside of the glass house remains hot. Carbon dioxide and other gases in the environment act in the same way as the glass in a glasshouse. These gases in the atmosphere let in ultraviolet waves, which are converted to infrared and cannot get out, thus heating the Earth. This produces the 'greenhouse effect'. However, a certain level of this is needed. It has been suggested that the Earth's atmosphere would be, on average, 133°C cooler but for the greenhouse effect.

Figure 12.12
A simple thermos flask.



Activity 12.7 INSULATION

'Silver batts' are advertised as being an excellent method of insulating houses. Investigate the characteristics of these 'batts' that might help you to determine the truth in these advertisements.

Activity 12.8 BUNSEN ISOTHERM

The temperature around a lit Bunsen burner gets lower the further you are from the flame. But it is not a regular decrease because of convection currents taking hot air upwards.

1. Predict the shape of lines joining points of equal temperature about a Bunsen flame (as viewed from the side). These lines are called **isotherms** (Greek *iso* = 'equal').
2. Use a thermometer to measure temperatures around the flame at say 10 cm intervals away from the flame. Do this for points directly above the flame (12 o'clock), 1 o'clock, 2, 3, 4 and 5 o'clock. Do you need to do both sides of the flame? Now draw an isotherm diagram. Explain why the shape is not symmetrical.

NOVEL CHALLENGE

The roofs of two houses are covered in snow. *In which house is the ceiling insulation better: the one in which the snow melts quickly or slowly?*

NOVEL CHALLENGE

Cut a grape almost half-way through and pull each half apart slightly so that there is a thin bridge of skin between the two halves. Put it in a microwave and give it 10 seconds on high. *Now explain that!*

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

*7 Which allows heat to flow better: cork or iron? Why do you think this is so?

*8 Explain why convection occurs in fluids but not in solids.

*9 If a couple of blocks of ice are placed in a test-tube with water as shown in Figure 12.13, the ice will float at the top — ice is less dense than water. Explain the process by which the ice obtains heat energy needed to melt.

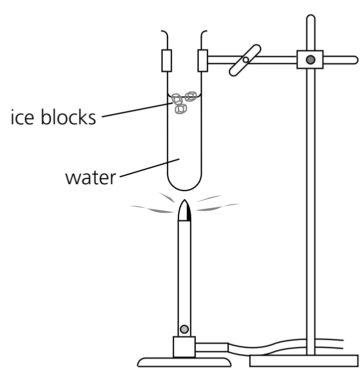


Figure 12.13
For question 9.

*10 Explain how a hot cup of coffee sitting on the kitchen table loses heat energy.

*11 A concrete floor is 15 cm thick. If the temperatures on opposite sides of the floor are 4°C and 20°C, calculate the rate of heat flow per unit area of the floor.

*12 Calculate the heat lost through a 3 m by 3 m brick wall if the wall is 30 cm thick and the temperature of the exterior is 5°C and the internal temperature is 28°C.

INVESTIGATING

The Space Shuttle has 27 416 tiles each of area 15 cm^2 as surface insulation. They are made of low-density silica fibre coated with waterproof borosilicate coating. They can stand 1648°C whereas the aluminium skin underneath melts at 660°C. Find out the thermal conductivity of these tiles.

INVESTIGATING

Sunflower seeds germinate faster if microwaved for 30 s first. But if you do it for 60 s they don't germinate at all. On the other hand, carrot seeds take 14 days whether they are given 30 s or 60 s. Propose a testable hypothesis.

***13** Calculate the heat lost in 30 minutes through a 2 m × 1 m × 6 mm thick glass window if the temperature difference between the two sides is 20°C.

***14** Does wood burning in a pot-belly stove warm the room by conduction, convection, or by radiation? Explain.

***15** What colour are the copper pipes used in solar hot water systems? Why are they this colour?

***16** What is the difference between infrared radiation and visible light?

***17** Explain why an iron rod at 1000 K is red-hot while a similar rod at 2000 K is white-hot.

***18** Why on a cold morning do the silver handlebars of bicycles feel colder than the black rubber hand-grips?

***19** On a cold winter's morning why does a metal spoon feel colder than the table it rests on?

***20** Why do knitted jumpers keep you warm in winter?

***21** Brass bases are good for saucepans, but why are the sides not made of brass? There may be economical and weight reasons for this, but what is a good reason in terms of conduction?

***22** A 375 mL can of 'Coke' and a 375 mL bottle of 'Coke' were placed in the freezer at the same time.
(a) Which would cool the faster?
(b) What other physical characteristics of the containers would be worth considering in arriving at your answer to part **(a)** of this question?

***23** Metal roofs of houses were traditionally left white or silver, but manufacturers over recent years have been producing them in all colours. Is this a good move? Explain.

***24** In cold countries where icy conditions can make roads slippery, gravel or soot is thrown on the icy roads. What effect would this have?

***25** Why do many people die in intense bush fires without being touched by the flames?

***26** A Davy Safety Lamp consists of a wire mesh box placed over a lit candle (see Figure 12.14). If a stream of Bunsen burner gas is directed at the mesh cage the gas burns but the flame does not come through and ignite the Bunsen burner. Explain why this is so and its advantage in coal mines.

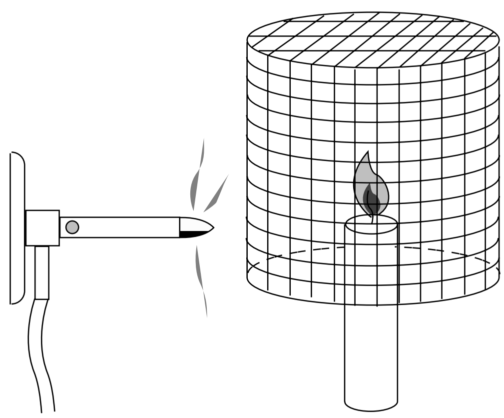


Figure 12.14
A Davy Safety Lamp.

***27** A glass box is equipped with glass chimneys. A small candle is placed under one of the chimneys and smoke is introduced into the other chimney.
(a) In what direction does the smoke move? Explain why.
(b) A lid is put on the top of the chimney above the candle. What happens now?

***28** Will a candle burn in zero gravity, such as on board a space shuttle, even if there is a normal supply of oxygen? Explain your answer.

*****29** A candle is lit and placed in a can open at the top. The can is dropped from shoulder height to the ground but remains upright while it descends. Will the candle go out? Explain. If you cannot decide, try it.

*****30** A match held near a powerful light bulb (say 100 W) will not ignite, but if you blacken it with graphite from a 'lead' pencil it will. Explain.

*****31** People can dip their moist fingers in molten lead (400°C) without getting burnt. How can this be? Do not try it!

****32** Do you think you could make a lens out of ice and use it to focus the Sun's rays on some paper to ignite it? Would the ice melt before the paper was hot enough to catch on fire? Explain in terms of radiation and conduction.

Extension — complex, challenging and novel

*****33** Modern house designers pay a great deal of attention to the conservation of energy. In doing so they insulate houses to reduce the loss of heat in winter and the absorption of heat in summer. Heat is lost through the roof, the walls, the windows, the doors and cracks. Suggest where most heat is lost. Also suggest how these losses can be reduced. This will involve describing the types of materials that may be used.

*****34** Heat sinks on CPUs and transistors have certain characteristics that improve their performance. They are normally black, thin, made of aluminium, and have many vanes to increase their surface area. Critically analyse each of these characteristics and the ability of each to help to remove heat from the device.

*****35** Windows in office buildings are double glazed to conserve energy. Analyse the construction of double glazed windows to determine how they achieve this purpose.

*****36** Many modern low-set school buildings are constructed of 20 cm concrete blocks. If a school room has dimensions of 4.0 m $\times$ 3.5 m $\times$ 2.5 m, calculate the heat lost from this room through the walls to the exterior in 30 minutes if the external temperature is 15°C and the internal temperature is 25°C.

*****37** Imagine a baby duck's body to be a perfect sphere of 7 cm diameter with an internal temperature of 32°C (Figure 12.15).

Some data:

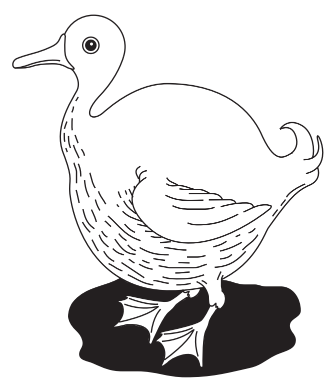
- Volume of a sphere = $\frac{4}{3}\pi r^3$.
- Surface area of a sphere = $4\pi r^2$.
- Thermal conductivity of feathers = 0.03 W/m/°C.

(a) Calculate the rate of heat flow from the duck on a day when the outside temperature is 15°C. Make whatever assumptions about the duck and its feathers you find necessary.

(b) How much heat does the duck have to generate per hour to maintain its body temperature at 32°C?

(c) Due to some genetic engineering, if the baby duck was a perfect cube instead of a sphere but still had the same volume, would the rate of heat loss be any different? Explain!

Figure 12.15
Spherical duck.



UNIT 05

Wave Motion

CHAPTER 13

Wave Motion in One Dimension

13.1 INTRODUCTION

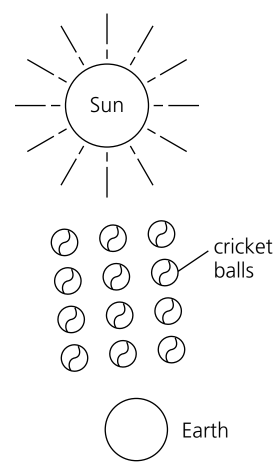
When a prize fighter trains by hitting a punching bag, energy travels from the closed fist to the bag, and if someone is holding the bag, to that person.

When a baseball is hit, energy travels with the ball to the person who catches it. The amount of energy it carries can be felt by the stinging of the hands.

But how does light energy travel from the Sun to be used by solar cells on Earth? Is the energy carried by moving cricket balls, perhaps?

There seems to be no particle like a fist or a ball to carry the energy — at least no visible particle. Where the answer lies may be seen by observing a surfboard rider in the ocean. As the unbroken wave passes the rider, he or she goes up and down but does not move forward. However, this water wave carries energy — very much energy. Consider the energy carried by a 'tidal wave'. (A poor or misleading term. A giant wave produced in the ocean often by volcanic activity or earthquakes is called a 'tsunami', and has nothing to do with the normal tide movements.) One of the largest tsunamis recorded was caused by a volcanic explosion on the island of Krakatoa on 27 August 1883. The resulting 40 m high tsunami lashed the coast of Indonesia killing some 36 000 people. This tsunami was even registered by the tide gauges in the English Channel. Tsunamis created this way often cause more deaths than the disturbance that created them. As tsunamis often cause havoc in Japan, Hawaii, and other Pacific Islands, great efforts are made to detect the epicentre of the earthquakes by measuring wave velocities. This allows determination of the expected time of arrival of the tsunami so people in low-lying areas can be warned and evacuated.

Figure 13.1
Does the energy from the Sun come like this?



The above is one example of wave motion in nature. There are many more examples. Can you think of some?

The understanding of waves is also important in modern-day conveniences. Water beds have baffles in them to stop waves when a person rolls over. Wave generators are being used to create waves in theme parks for the entertainment of patrons. The motion of the waves in oceans is one of the latest methods of generating electricity.

So waves can carry energy, but how do they do this?

13.2 WAVES

Waves are classified according to the method of transfer of energy. If a medium is required for the transfer of energy, then the waves are called **mechanical** waves. If no medium is required and the waves are able to travel through a vacuum, then the waves are called **electromagnetic**. In this chapter we will discuss only mechanical waves, while in Chapter 15 forms of electromagnetic waves such as radio waves, light and X-rays will be discussed.

A good working example of a mechanical wave can be created by dropping a stone in a pool of water. A circular wave is seen to radiate outward from the point the stone enters the water. In this case water is the 'medium'.

Photo 13.1
Circular waves produced when a stone is dropped into a pool.

Notice that, to create the wave, you have to create a **disturbance** in an undisturbed medium. The wave continues to go outward until it runs out of energy. How is this loss of energy seen? The height of the wave is called the **amplitude** of the wave. It is the maximum displacement of the wave from its equilibrium position shown as '*A*' in Figure 13.2.

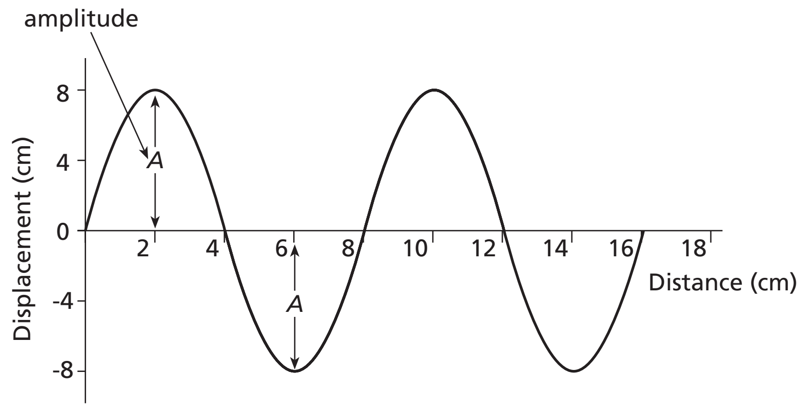


Figure 13.2
The amplitude of a wave is the maximum distance from the equilibrium position.

What other quantity is determined by the amplitude of a wave?

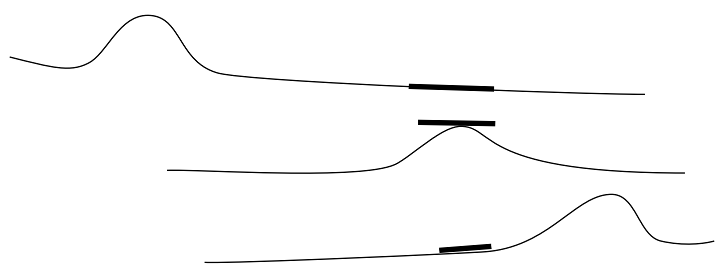
The amplitude of a large water wave might be 10 m. This suggests that the wave would have large amounts of energy. The larger the amplitude the more energy the wave possesses. The energy of the wave comes from the disturbance. Some of the energy of the stone is transferred to the water wave. As you go further from the source of the wave the amplitude of the wave becomes less as the energy dissipates. Waves similar to these water waves can be created in many objects. Children can often be seen holding the ends of a piece of skipping rope or a hose and flicking it. A wave or **pulse** moves from the flicked end to the other end. The energy put into the wave can easily be felt by the child at the other end. The energy can be so great that it may cause the rope to flick out of the hands of the receiver. Notice that the energy and the pulse moves along the rope without the particles that make up the rope moving toward the receiver.

NOVEL CHALLENGE

Put a lit candle in a room and open a door quickly. How long will the breeze take to get to the candle? Measure the distance and the time. Do you think the breeze would travel at the speed of sound in air?

If there is a small branch floating in a pond what happens to it as a wave passes? The branch goes up and down as the wave passes but returns to its original rest position once the wave has passed. The same thing happens to you and your small fishing dinghy as the wash from a large boat passes under you. The rest position is also called the 'equilibrium' position.

Figure 13.3
A branch in a pond moves out and back as a wave passes.



Wave types

Water waves or rope waves are particular types of waves. As seen in the water wave, the water (and branch) move upward as the wave passes. The particles that make up the water move at right angles to the direction the wave is travelling. This is the same for the rope wave. The rope moves upward as the pulse passes and then back to its original position.

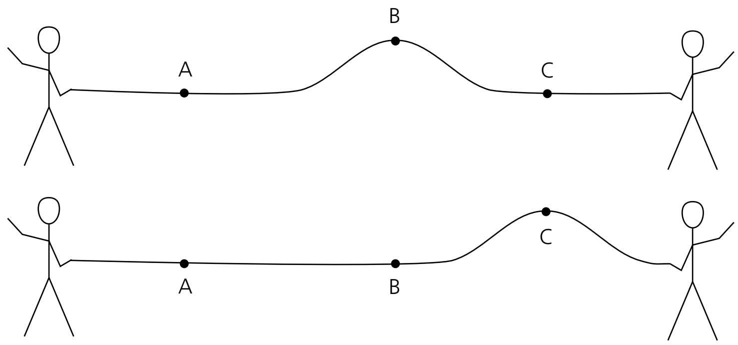


Figure 13.4
The parts of a rope move outward as waves pass.

Waves that do this are called **transverse waves**. (The word transverse comes from the Latin *transvertere* meaning 'to turn across'). Each point of the wave vibrates perpendicularly to the direction the wave is travelling — perpendicular to the direction of **propagation** of the wave (Latin *propago* = 'layer of a plant'; adds layers as it grows outwards). Examples of waves that are transverse in nature are waves in water; waves in ropes, hoses, and springs; and electromagnetic radiation, examples of which are light, radio waves, and television waves.

Notice the direction of the motion of the particles of a spring as a transverse wave passes as shown in Figure 13.5.

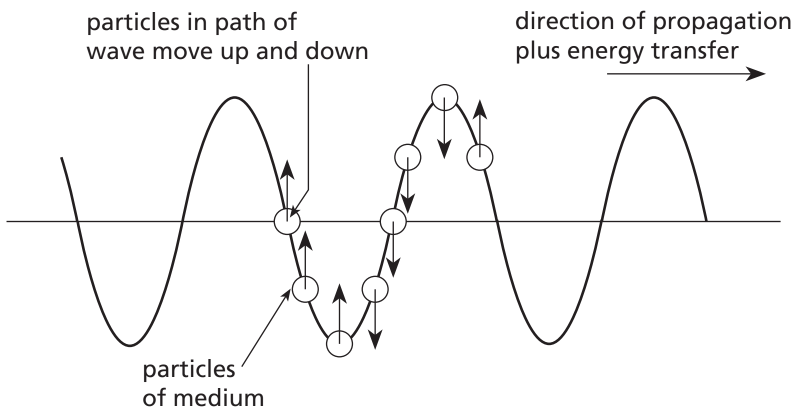


Figure 13.5
In a transverse wave the particles move at right angles to the motion of the wave. The direction of individual particles is given by the arrows.

Another type of wave is a **compressional** or **longitudinal** wave. Examples of these can be created in springs by compressing a part of a spring and then letting it go so the compression travels down the spring.

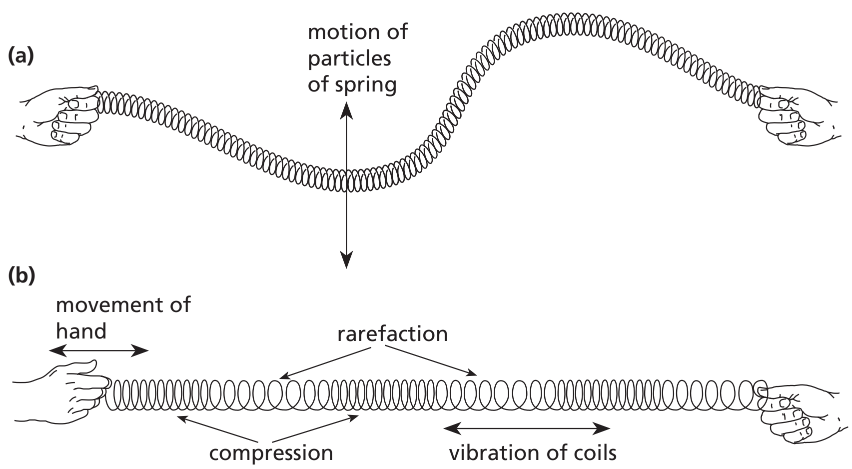
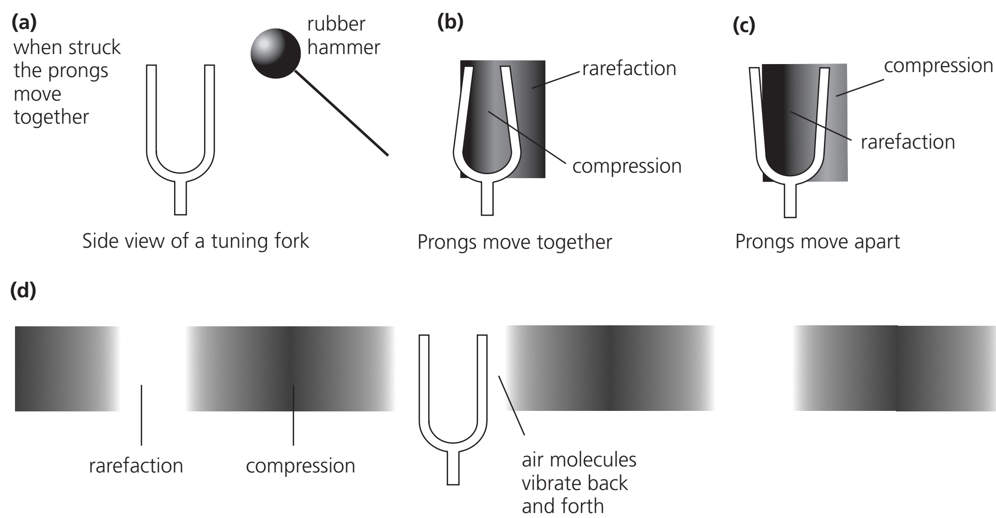


Figure 13.6
(a) Transverse waves.
(b) Longitudinal waves.

The particles of a spring propagating longitudinal waves vibrate in the same direction as the pulse is moving. This creates **compressions** and **rarefactions**. Can you think of other types of longitudinal waves?

Musical instruments create longitudinal waves by their action on the air particles in close proximity to the vibrating instrument. For example, when a drum membrane moves out it forces air particles together, generating a compression. When it moves in it produces a rarefaction. Tuning forks work in the same way (Figure 13.7). Would these instruments work in outer space?

Figure 13.7
Compression and rarefaction occur in the air molecules surrounding the prongs of a tuning fork.



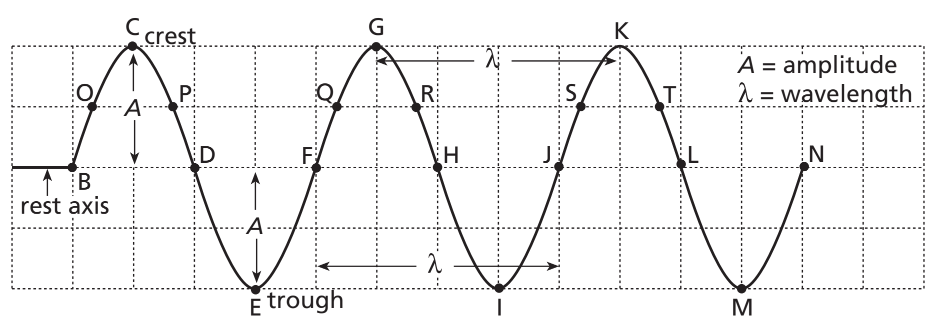
WAVE CHARACTERISTICS 13.3

PHYSICS FACT

The symbol λ is the Greek 'L' (lambda) for length.

A single disturbance produced by a source such as a flicking rope is called a **pulse**, but if a continuous set of pulses is produced by a source with a constant time interval between the generation of each pulse, the result is a wave, which has several characteristics. The **wavelength** (λ) is the minimum distance between two points on the wave that are in **phase**, for example, the distance between two consecutive **crests** or **troughs**. If the two points are in phase they are at the same distance from the rest position and are moving in the same direction at the same time. For example, D and H in Figure 13.8 are in phase, as they are on the equilibrium position and about to move up. C and E are out of phase — C is about to move down whereas E is about to move up. The wavelength of a wave is shown in Figure 13.8 between G and K, O and Q, or B and F. It is a little harder to see and measure the wavelength of a longitudinal wave. It is the distance between the middle of adjacent compressions, or adjacent rarefactions, as shown in Figure 13.9.

Figure 13.8
Wave characteristics of typical transverse waves.



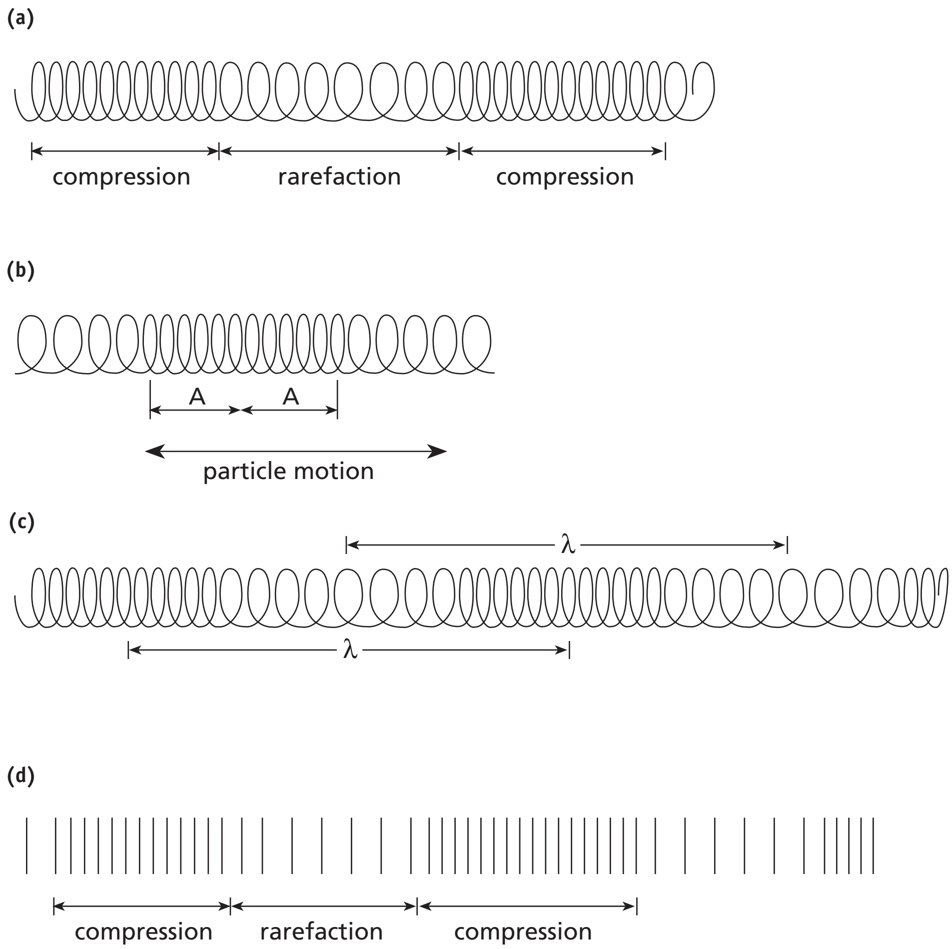


Figure 13.9
Wave characteristics of typical longitudinal waves.

NOVEL CHALLENGE

The June 2000 issue of the very prestigious *New England Journal of Medicine* reported that the average rate of jaw movement of gum chewers in the USA is 100 Hz. What do you suppose they really meant?

The **frequency** (f) of a wave is the number of waves passing a given point per second or the number of waves created by the source per second. The unit for frequency would thus become waves per second, or cycles per second (c s^{-1}), or the modern unit of a **hertz** (Hz) named after the German physicist **Heinrich Hertz** (1853–94) who in the 1880s discovered a technique for transmitting and receiving radio waves. A hertz is a cycle per second. For example, if four crests pass a point in 1 second then the frequency is 4 Hz. The frequency of visible light waves is between 4×10^{14} Hz and 8×10^{14} Hz. The ranges of frequencies heard and produced by some animals and produced by some musical instruments are given in Figure 13.10.

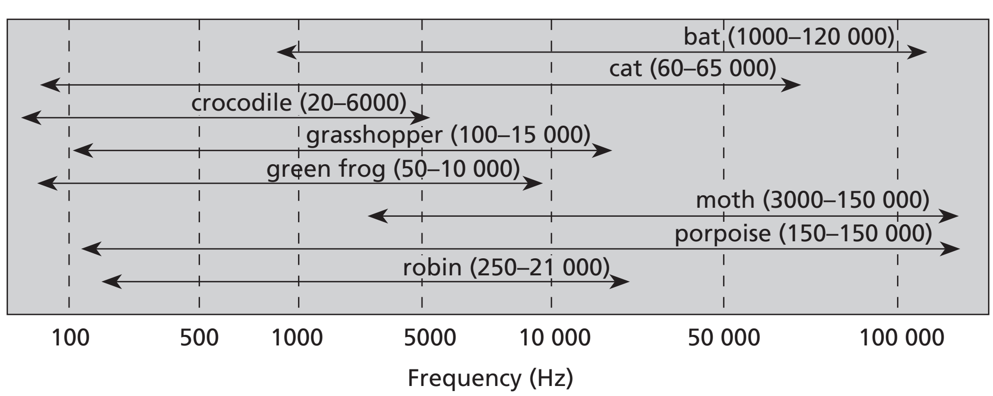


Figure 13.10
Range of frequencies of sounds produced by various animals.

NEI Activity 13.1 HEARING RANGE

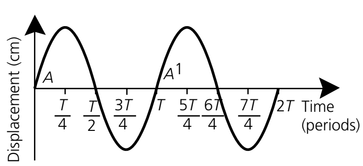
1. See if you can find out the frequency range for sounds that humans can hear.
2. How does this range change as you get older?
3. Dogs are said to be able to hear sounds of higher frequency. Find out what range of frequencies is audible to them.

The **period** of a wave is the time it takes for one full wave to pass; that is, one complete cycle to pass. If the frequency of a wave is 10 Hz or 10 cycles per second, then 10 waves pass per second. It will then take 1/10 second for one wave to pass. The period of the wave is 1/10 second. Therefore period (T) = 1/frequency.

$$T = \frac{1}{f}$$

The wave equation

Figure 13.11
The propagation of a transverse wave with time.



How can we measure the speed of a wave? How fast is the wave travelling? You could measure how far the wave travels in a certain time or you could use the wave characteristics (Figure 13.11). At one instant, point 'A' will be at the origin as shown. One period later point A will be at A′. This means the wave has travelled one wavelength in one period.

The speed of the wave = distance travelled/time taken.

$$\boldsymbol{v} = \frac{1\lambda}{T}$$

$$\boldsymbol{v} = \frac{1}{T}\lambda$$

$$\boldsymbol{v} = f\lambda$$

This is known as the **wave equation**, where $\boldsymbol{v}$ is the speed of the wave in m s^{-1}, λ is the wavelength of the wave in m, f is the frequency of the wave in Hz.
Note: this equation will apply to all wave forms, both mechanical and electromagnetic.

Example

An observer sitting on a shore counts the waves and finds that there are 6 waves per minute hitting the shore. She measures the distance between consecutive crests to be 10 m. What is the velocity of the waves?

Solution

$$\boldsymbol{v} = f\lambda$$
$$\boldsymbol{v} = 6/60 \text{ Hz} \times 10 \text{ m}$$
$$\boldsymbol{v} = 1 \text{ m s}^{-1}$$

PHYSICS FACT

Breweries monitor the percentage of alcohol in beer as it flows through pipes by passing it over a special beer tuning fork. The rate of vibration changes according to the density (and hence alcohol concentration) of the beer. Clever huh?

Questions

1. Two students 5.0 m apart 'flicked' a spring to create waves. They found that when they flicked it twice per second, 10 wave crests were created between the two students. Calculate the velocity of the waves.
2. The speed of light is 3.0×10^8 m s^{-1}. If the wavelength of red light is 5.0×10^{-7} m, what is its frequency?

3 Find the frequency of the light emitted from the common laboratory laser. You will find the wavelength of the laser marked on it.
4 The frequency of a 5.0 m sound wave in water is 300 Hz. What is the speed and the period of the sound waves in the water?
5 Figure 13.8 shows a transverse wave. If this wave is travelling to the right, what is the direction of movement of points B, G, P, D, and M at this time?

NOVEL CHALLENGE

A boat catches fire and the skipper jumps overboard and swims away. The skipper hears an explosion while underwater, lifts his head out of the water and hears another explosion. Bystanders say there was only one explosion but the skipper says there were two. Who was correct? Explain.

13.4 WAVES AND BOUNDARIES

The speed of a wave is a characteristic of the medium in which it is moving, and changes when a wave moves from one medium to another. This is shown in Table 13.1.

Table 13.1 SPEED OF WAVES IN VARIOUS MEDIA

WAVE TYPE	MEDIUM	SPEED ($m\ s^{-1}$)
Sound	carbon dioxide	260
	air	331
	hydrogen	1290
	pure water	1410
	salt water	1450
	glass	5500
Light	vacuum	2.997×10^8
	air	2.988×10^8
	glass (crown)	2.0×10^8
Earthquake	crust	3500 (transverse)
		8000 (longitudinal)
	mantle	6500 (transverse)
		11 000 (longitudinal)

NOVEL CHALLENGE

Scientists monitor global warming by measuring ocean temperatures. They pass a 57 Hz signal at Heard Island in the southern Indian Ocean and time its arrival at points in major ocean basins of the world. How does speed vary with water temperature? Draw a graph.

How can we best explain what happens when a wave hits a barrier or moves from one medium to another? For example, what happens when light hits a mirror? What happens when a pulse sent down a spring by one student hits the firm hand of another at the other end? We will use waves in springs to investigate this principle. The spring pulse is easier to visualise and to do experiments with than light waves. Why?

Figure 13.12 shows a spring attached to a wall. If a pulse is sent down this spring what happens when it hits the wall? This might be a good time to observe this.

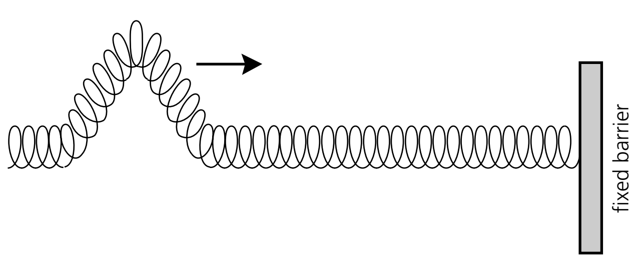


Figure 13.12
A slinky spring used to investigate how waves are reflected from fixed barriers.

EI Activity 13.2 REFLECTION OF PULSES

Attach a spring to a wall or have a friend hold it firmly. This is called a 'fixed end'.

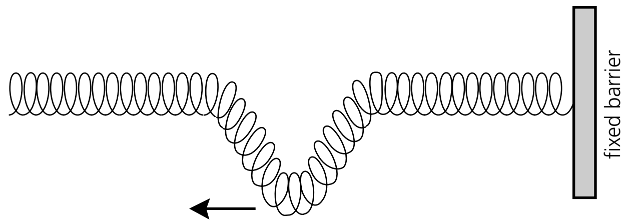
1 Send a pulse (the **incident pulse**) down the spring and observe the reflection of the pulse from this fixed end.

2 Measure the time taken for the pulse to travel from the source to the wall and from the wall back to the source.

The pulse is **reflected**, which means it comes back along the spring. But on what side of the spring does it return?

It will be observed to come back on the opposite side of the spring with approximately the same amplitude. The **reflected pulse** is said to be inverted, **out of phase**, or **180° out of phase** with the incident pulse. It is said to have undergone a **phase reversal.** The speed of the reflected pulse and the incident pulse will be the same, as the speed in a particular medium remains constant.

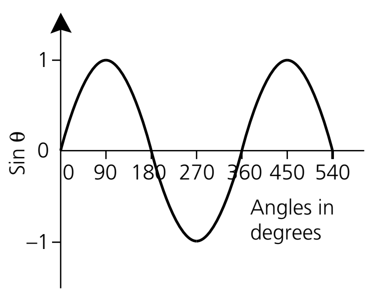
Figure 13.13
A pulse reflects from a fixed end with a phase change of 180°.



SR Activity 13.3 PHASE

This activity will help you to understand phase change in relation to angular degrees. Rule up a piece of graph paper, making the vertical axis the value of the 'sin' of the angle. Make about 5 cm equal to one unit. Place the angle in degrees on the horizontal axis. Using your calculator find sin θ, where θ = 0°, 90°, 180°, 270°, 360°, 450°, and 540°. On the graph paper plot the value obtained for the sine of the angle against the angle. Draw a smooth curve through the points. You should obtain something similar to Figure 13.14.

Figure 13.14
The wave produced by plotting the value of sin θ against the angle θ.



This curve resembles a perfect transverse wave. Notice the position of the points on the curve at 90°, and 270°. They are the same distance on opposite sides of the *x*-axis with 180° between them. When waves are reflected from fixed ends the reflected waves are on the opposite side of the spring. This might help to explain why the reflected wave is said to be 180° out of phase with the incident wave.

Another way of understanding this is to view an object going around in a vertical circular path, such as a seat on a Ferris wheel. When viewing from the side you will notice a particular seat on the wheel appears to go up and down like a cork in water when waves are passing. The cork on a crest would correspond to the particular seat on the Ferris wheel being at the top, and the cork would be in a trough when the seat is at the bottom. Since you go through 360° to complete a single circle, then half a circle corresponds to 180°, which is the difference between the cork being on a crest or in a trough.

Reflection can also occur if the spring has a free end, that is, it is not fixed to a wall. However, this time the pulse comes back on the same side as (in phase with) the incident pulse.

EI Activity 13.4 REFLECTION FROM FREE ENDS

Try it!

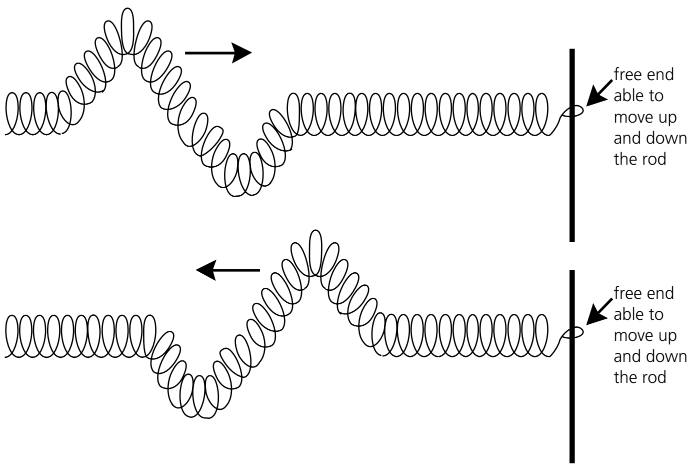


Figure 13.15
A pulse reflects from an open end with no phase change.

The above cases are the two extremes. We will now consider a wave that meets a boundary between two different springs.

EI Activity 13.5 LIGHTER TO HEAVIER SPRINGS

Try generating a pulse that travels from a lighter (less dense) spring to a heavier (more dense) one as shown in Figure 13.16, and observe the resulting pulse(s) after the pulse meets the boundary.

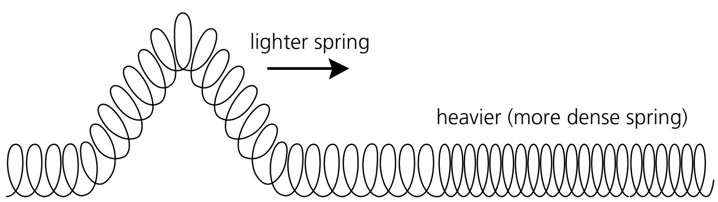


Figure 13.16
Two media — a lighter spring and a heavier spring.

When it meets the boundary (or join) some of the pulse is **transmitted** (continues on into the more dense spring) and some is reflected. At this point it will be noticed that the transmitted pulse is upright, in phase, or on the same side of the spring as the incident pulse, but the reflected pulse is upside-down or out of phase. As far as the reflected pulse is concerned, the boundary is behaving like a 'fixed end'.

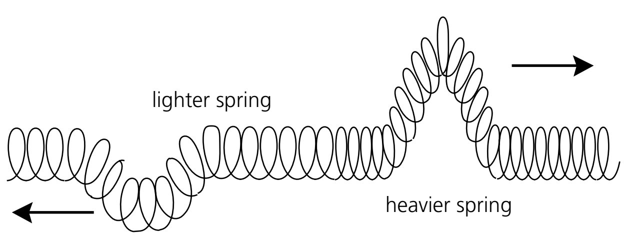


Figure 13.17
Waves going from a lighter spring to a heavier one are reflected, inverted and transmitted in phase.

Since the pulse has broken up, the amplitudes of both reflected and transmitted pulses are smaller than the incident pulse. The amplitudes of the transmitted pulse and reflected pulse are determined by the relative densities of the two media. If the second medium is much heavier, the transmitted pulse will be small and the reflected pulse will be large. If the second medium is only a little more dense, the transmitted pulse will be larger and the reflected pulse will be smaller. The velocity of the transmitted pulse depends on the medium but will be less than the incident or reflected pulses. However, the velocity of the reflected pulse will be the same as the incident pulse as the pulses are travelling in the same medium.

EI Activity 13.6 HEAVIER TO LIGHTER SPRINGS

Now create a pulse that goes from a heavier spring to a lighter spring as shown in Figure 13.18, and observe the resulting pulses.

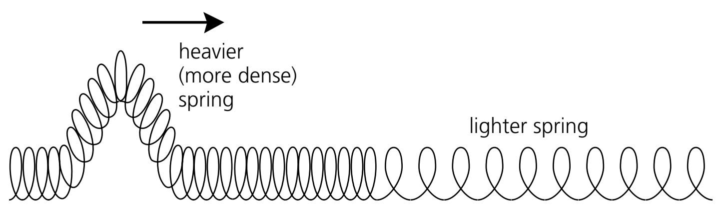


Figure 13.18
Waves going from a heavier spring to a lighter spring.

It will be observed that both reflection and transmission occur. In this case both the transmitted and reflected pulses will be seen to be on the same side of the spring as (in phase with) the incident pulse. The boundary is behaving like a 'free end', as far as the reflected pulse is concerned.

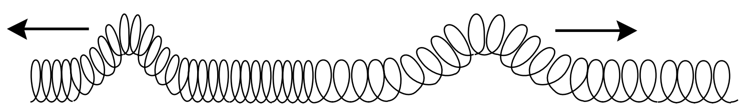


Figure 13.19
Waves going from a heavier spring to a lighter spring are reflected in phase and transmitted in phase.

In general when waves go from a less dense to a more dense medium the reflected pulse will be out of phase and the transmitted pulse will be in phase; and when waves go from a more dense medium to a less dense medium the reflected and transmitted pulses will both be in phase. This is summarised in Table 13.2.

Table 13.2 THE PHASE RELATIONSHIPS BETWEEN REFLECTED AND TRANSMITTED PULSES, WITH RESPECT TO THE INCIDENT PULSE WHEN IT MEETS VARIOUS BOUNDARIES

BOUNDARY	REFLECTED PULSE	TRANSMITTED PULSE
Fixed end	out of phase	—
Free end	in phase	—
Light medium to heavy medium	out of phase	in phase
Heavy medium to light medium	in phase	in phase

This principle of reflection of waves is important when applied to musical instruments, such as open and closed wind instruments. This will be discussed in Chapter 15.

13.5 SUPERPOSITION OF WAVES

You may have noticed that there are many instances when one wave meets another, either those produced by nature or those produced by man. For example, what happens when waves produced by two passing boats in the ocean cross over one another? What happens when two sound waves or two light waves meet? Do they cancel each other out, resulting in the elimination of both waves?

The intersection of two waves can be seen by producing pulses in a spring, one from each end.

EI Activity 13.7 INTERSECTION OF PULSES

1 Produce pulses simultaneously from either end of a spring with:
   (a) pulses on the same side of the spring;
   (b) pulses on the opposite side of the spring.
2 Observe the resulting wave form when they meet, and after they pass.
3 Draw diagrams to represent these interactions.

In the above activity pulses are seen to add together when they pass over each other, producing a much larger, or smaller, or differently shaped wave. The type of resulting wave depends on whether the pulses were produced on the same side or on opposite sides of the spring. However, once they have passed they continue as though they had not met. Figure 13.20 shows the resulting pattern produced when two pulses intersect. If they are produced on opposite sides of the spring **destructive interference** occurs, producing a smaller wave, or no wave at all at that instant. If they are produced on the same side of the spring **constructive interference** occurs, producing a super crest.

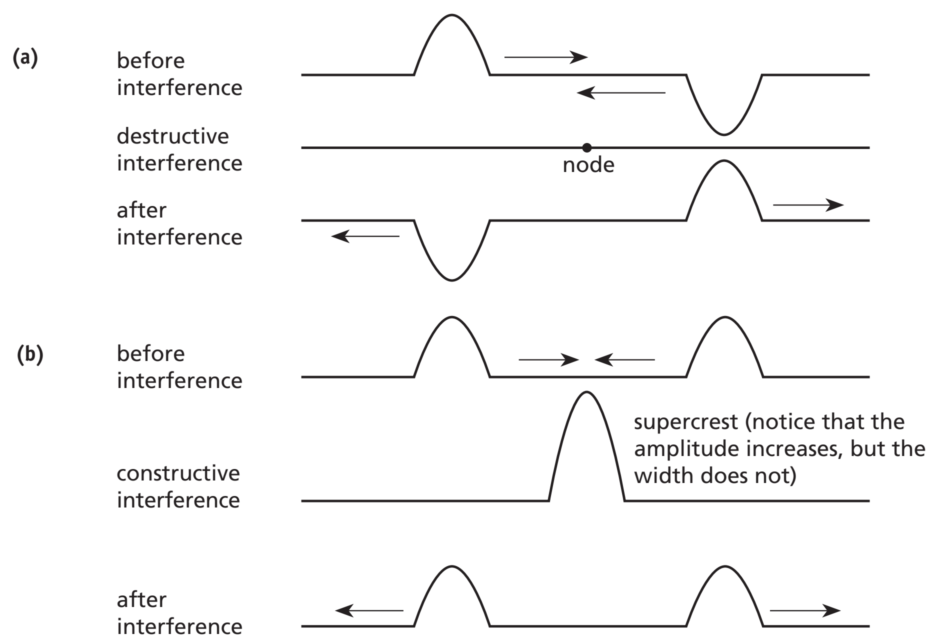


Figure 13.20
Superposition of waves may cause **(a)** destructive or **(b)** constructive interference.

This process is called the principle of **superposition**. The resulting wave can be obtained by adding the pulses' displacements, from the equilibrium positions, at several points, as shown in Figure 13.21.

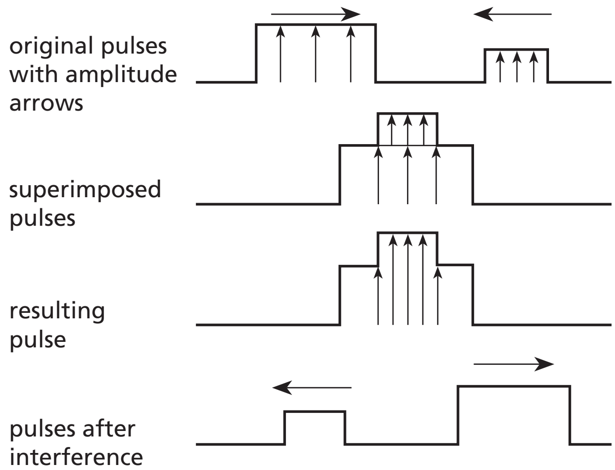
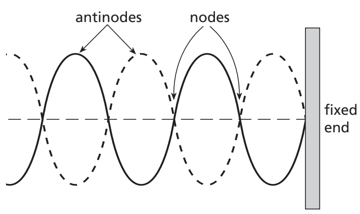


Figure 13.21
When two waves are superimposed their displacements add and they continue unaltered.

Standing waves

If a series of waves are created from each end of a spring of the **same amplitude and frequency**, a **stationary wave** or **standing wave** is created. This occurs because of the continued cancellations and additions of the waves as they travel along the spring and pass through each other. When the first crests meet they produce a pulse of twice the amplitude. A short time later (a quarter of a period) the pulses have moved so the crest of one is interacting with the trough of another, producing a point of zero displacement — a **node**. (The word 'node' comes from the Latin word *nodus*, meaning 'knot' — it looks as though the spring is knotted together.) Another quarter of a period later each pulse has moved another quarter of a wavelength and the two crests and two troughs again meet, producing super crests and troughs — **antinodes**. The characteristic standing wave pattern, as shown in Figure 13.22, is difficult to produce by two students flicking a spring from either end as they have to continually flick the spring in phase. But it is easy to produce by attaching the spring to a wall at one end and flicking the other as shown in Figure 13.22. The resulting standing wave pattern oscillates between the fixed and dotted lines as shown. Notice that the distance between two successive nodes is a half of a wavelength.

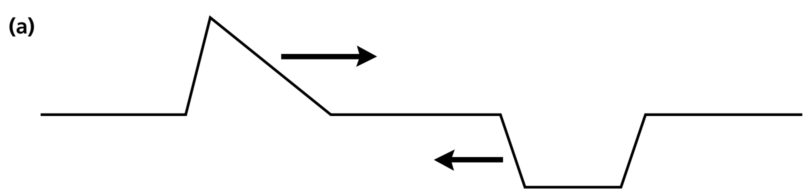
Figure 13.22
A standing wave produced by the continual generation and reflection of waves off a fixed wall. Oscillation only occurs between the dashed and solid lines.



Example

Construct the wave pattern produced when the two pulses shown in Figure 13.23 meet.

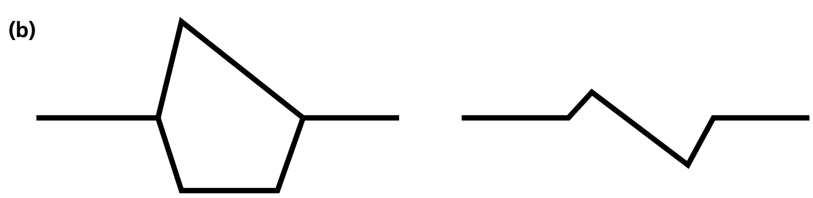
Figure 13.23
For example question.



Solution

See Figure 13.24.

Figure 13.24
For example question.



Questions

6 Use the principle of superposition to determine the resulting pulse when the pulses shown in Figure 13.25 are superimposed on each other.

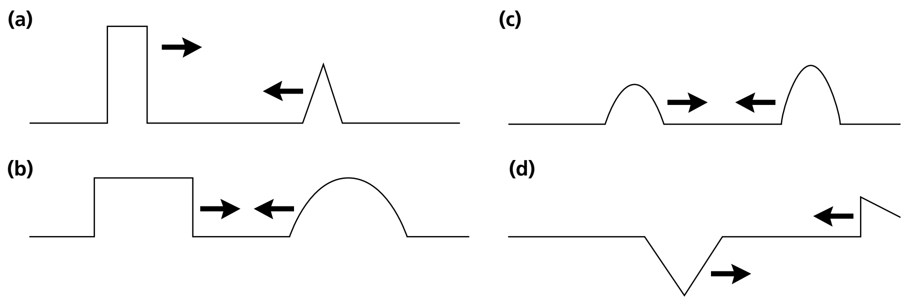


Figure 13.25
For question 6.

7 Two identical waves are produced on either side and either ends of a rope (Figure 13.26). Draw the resulting wave when they are in the positions shown. What do you notice about point X?

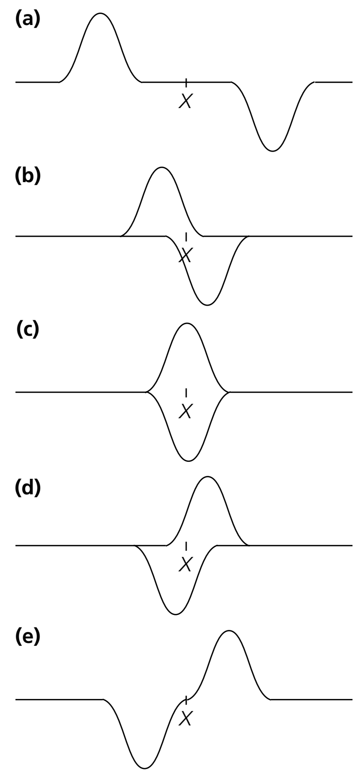


Figure 13.26
For question 7.

Wave motion in sports equipment

As you have seen in the chapters on force, momentum and energy there are several 'sweet spots' in bats and racquets. One is the centre of percussion — the point that produces no jarring in the hand when a ball is struck. This point is a nodal point for standing waves in the equipment.

Tennis racquet When a ball is hit, the racquet rings as waves run up and down its length (Figure 13.27). The string node is just above the centre of the strings.

Baseball bat If you hold a bat loosely by the handle and tap it with a hammer you will hear ringing at most points. But at about 16 cm from the far end (of a 78 cm aluminium bat) there will be very little sound. This is the nodal point and a ball struck here produces no stinging in the hand if held at the node at the other end. Try it!

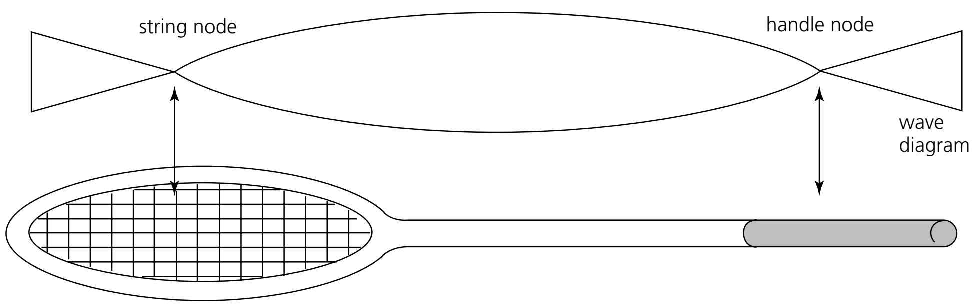
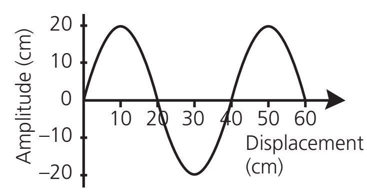


Figure 13.27
The string node of the standing wave of a tennis racquet is just above the centre of the strings.

13.6 GRAPHICAL ANALYSIS OF MOTION

Wave motion can be represented graphically in two ways: amplitude–displacement and amplitude–time graphs.

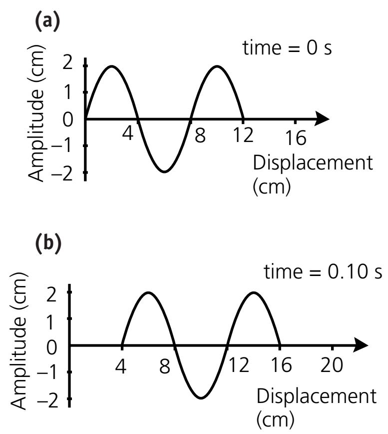
Figure 13.28
An amplitude–displacement graph for wave motion.



Amplitude–displacement graphs

The graph in Figure 13.28 represents the position of a wave at a certain time. From this graph the amplitude and the wavelength of the wave can be determined. If the position of the wave at another time is given, the characteristics of the wave that involve time, such as the speed and the frequency of the wave, can be calculated.

Figure 13.29
For example question.



Example

Figure 13.29**(a)** shows the position of a wave at time equals zero, and Figure 13.29**(b)** shows the position of the wave 0.10 second later.

(a) Calculate the speed and frequency of the wave.
(b) Draw the wave after another 0.2 second.

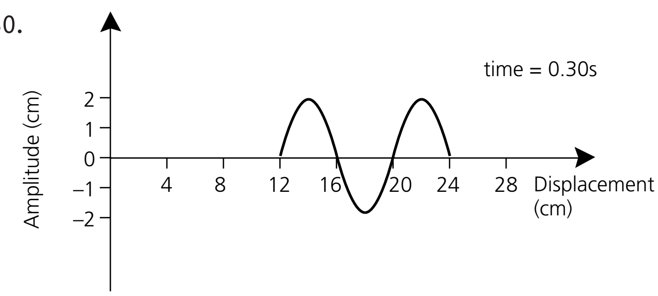
Solution

(a) The wave has travelled a distance of 4 cm in the 0.1 s. Therefore:

$$\begin{aligned}
\boldsymbol{v} &= \frac{d}{t} \\
&= \frac{4\ \text{cm}}{0.1\ \text{s}} \\
&= 40\ \text{cm s}^{-1} \\
\boldsymbol{v} &= f\,\lambda \\
f &= \frac{\boldsymbol{v}}{\lambda} \\
&= \frac{40\ \text{cm s}^{-1}}{8\ \text{cm}} \\
&= 5\ \text{Hz}
\end{aligned}$$

Figure 13.30
For solution to example question.

(b) See Figure 13.30.



INVESTIGATING

Hold an aluminium or steel rod (unscrew a retort stand) vertically and tap the top with a hammer. Hold it at different places and note the change in sound. You can get frequencies greater than 20 000 Hz (painful). How can you get beats?

Amplitude–time graphs

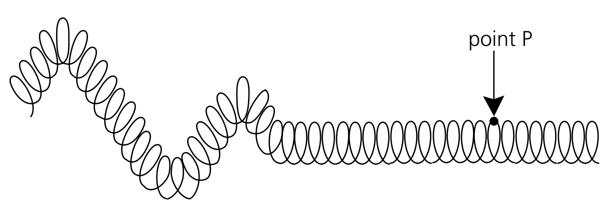
As well as indicating the amplitude of the wave these graphs indicate the position of the wave and the position of points on the wave at certain times. Thus the velocity of the wave can be calculated.

Example

A wave is created on a spring as shown (Figure 13.31). The displacement of point P is given by the graph (Figure 13.32).

(a) Calculate the period of the wave.
(b) What is the amplitude of the wave?
(c) If the wave is moving at 5 cm s^{-1} to the right what is the wavelength of the wave?

Figure 13.31
For example question.



Solution

(a) The wave repeats itself after (1.0 – 0.4) seconds, therefore the period is 0.6 s.
(b) Amplitude = 20 cm.

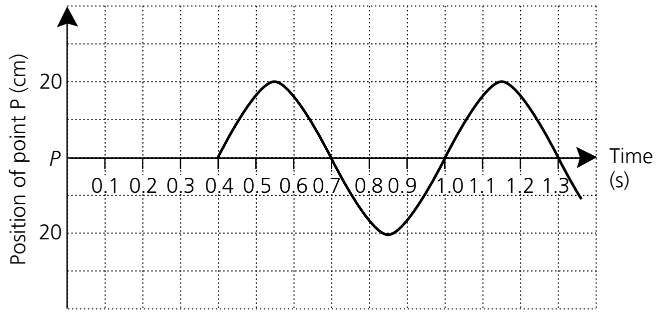


Figure 13.32
For solution to example question.

(c)

$$v = f\lambda$$

$$\lambda = \frac{v}{f}$$

$$= v \times T$$

$$= 5\text{ cm s}^{-1} \times 0.6\text{ s}$$

$$= 3.0\text{ cm}$$

Questions

8 Figure 13.33 represents the displacement of particles in a rope with time as a wave passes. Calculate **(a)** the amplitude of the wave; **(b)** the period of the wave; **(c)** the frequency of the wave.

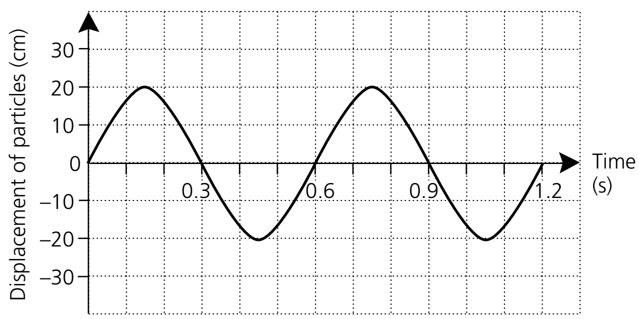


Figure 13.33
For question 8.

9 For the waveform shown in Figure 13.34 find the following:

(a) the wavelength of the disturbance;

(b) the amplitude of the wave;

(c) if the wave is travelling to the right at a speed of 80 cm s^{-1}, find the frequency of the disturbance.

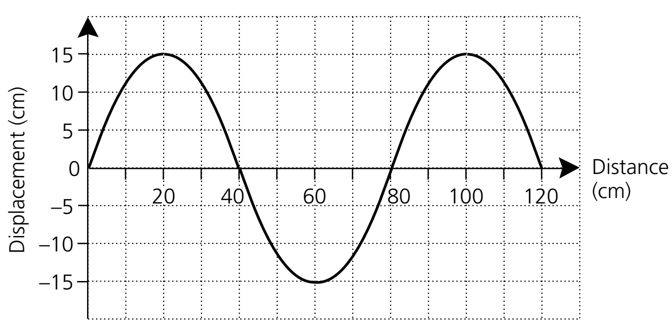


Figure 13.34
For question 9.

10 Figure 13.35 shows the position of a wave at two instances in time. From these graphs determine **(a)** the amplitude of the wave; **(b)** the wavelength of the wave; **(c)** the velocity of the wave.

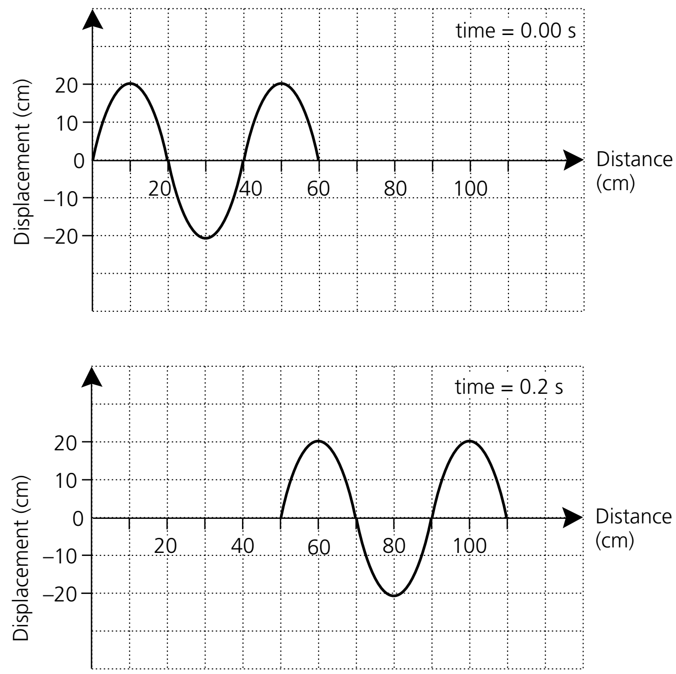


Figure 13.35
For question 10.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***11** What happens to the speed of a pulse in a slinky spring if it is stretched?

***12** Explain the difference between a longitudinal and a transverse wave.

***13** A dinghy anchored in the ocean is seen to bob up and down as waves pass.
- **(a)** What types of waves are they?
- **(b)** How do you know this?
- **(c)** What physical feature of the wave indicates the energy the wave possesses?

***14** Name a particular type of wave where the particles move at **(a)** right angles to the direction of propagation; **(b)** in the same direction as the wave is moving.

***15** Describe what is meant by the wavelength of a wave in relation to a longitudinal wave.

***16** State the wave equation indicating the meaning of the symbols used.

***17** Two students shaking a slinky spring create 10 waves in 5 seconds. The wavelength of the waves is 50 cm.
- **(a)** What is the frequency, period, and speed of the wave?
- **(b)** If the wave was shaken at a greater frequency **(i)** what characteristics of the wave would change; **(ii)** what characteristic would remain the same?
- **(c)** How can the speed of the wave in the spring be changed?

***18** Students creating waves in two slinky springs joined together found that the waves travelled down the first one, and were reflected upside-down from the junction of the springs.
- **(a)** What is the relationship between the heaviness of the two springs?
- **(b)** Will waves be transmitted into the second spring and if so will they be in phase or out of phase?

***19** What conditions are necessary for the creation of standing waves?

***20** Two identical waves are created from either ends of a long spring. Each wave has a wavelength of 20 cm and an amplitude of 10 cm. This produces a standing wave pattern.

(a) What is the maximum displacement of the resultant waveform as they pass through one another?

(b) How far apart are the nodes?

***21** Two weekend anglers find that their 4 m boat bobs up and down 3 times in 20 seconds, and exactly 3 wave crests can fit under the boat at any one time. What is the velocity of the waves?

***22** When a tuning fork is hit it vibrates at the rate of 300 vibrations per second. If the speed of sound in air is 340 m s^{-1}, calculate the wavelength of the sound waves produced.

***23** AM radio stations transmit radio waves that are electromagnetic waves similar to light waves, but have a frequency from about 500 kHz to 30 MHz. If they travel at 3×10^8 m s^{-1}, what is the wavelength of radio waves?

***24** The speed of sound in salt water is 1450 m s^{-1}, and 340 m s^{-1} in air. If the frequency of the sound being generated by a motor boat engine is 550 Hz, what is **(a)** the wavelength of the sound in air; **(b)** the frequency of the sound in salt water; **(c)** the wavelength of the sound in salt water?

****25** Figure 13.36 shows transverse waves being generated at the rate of 10 per second. From this diagram determine:

(a) the amplitude of the waves;

(b) the wavelength of the waves;

(c) the period of the waves;

(d) the velocity of the waves.

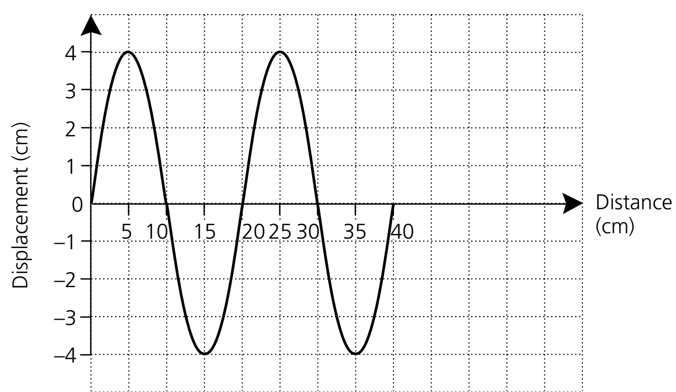


Figure 13.36
For question 25.

****26** When earthquakes occur they create waves that spread outward from the source. Three types of waves occur. The primary or P waves, which are caused by the back and forth movement of rocks; the secondary or S waves, which travel as a result of the up and down movement of rocks; and the L waves, which are ripples that travel on the surface and are set up when the P and S waves reach the surface. The L waves cause rocks to vibrate in an up and down motion.

(a) What type of waves are P, S, and L waves?

(b) Earthquake waves have a wavelength of 10 m and a speed of 3.0 km s^{-1}. What is their frequency?

****27** Figures 13.37 **(a)** and **(b)** show a transverse wave and a longitudinal wave moving to the right. Indicate the direction of motion of the points A and B.

Ch13-Walding 4th 25/8/04 11:50 AM Page 301

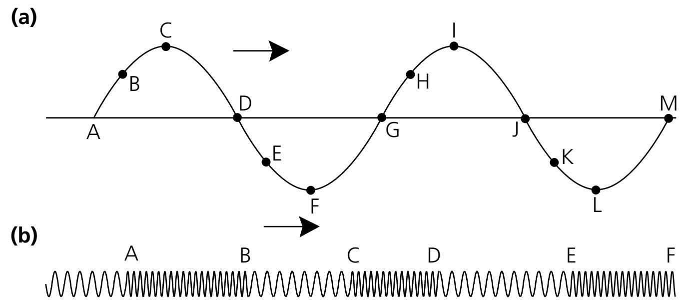


Figure 13.37
For question 27.

****28** Waves are sent down a set of connected ropes and later return to the sender as shown in Figure 13.38. What is the relationship between the density of the rope sections A, B, and C?

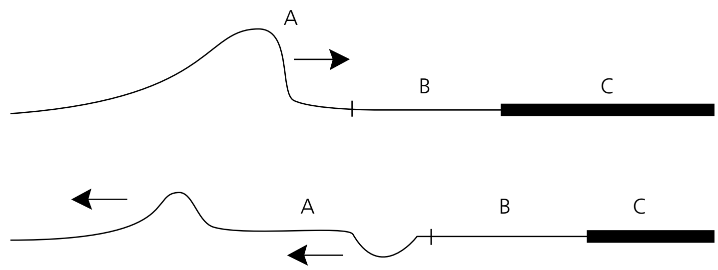


Figure 13.38
For question 28.

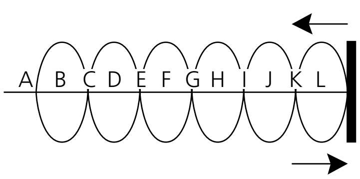


Figure 13.39
For question 29.

****29** A standing wave pattern is set up by reflecting waves off a wall as shown in Figure 13.39.

(a) Indicate the wavelength of the waves producing the standing wave.

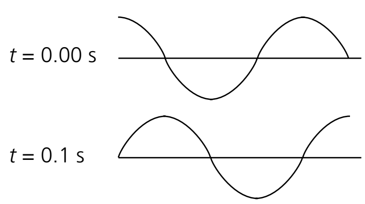
(b) Which points are nodes?

(c) Which points are antinodes?

(d) If the distance between 'A' and 'F' is 5.0 m, what is the wavelength of the waves?

****30** Figure 13.40 shows a wave in the same section of a string at two different times. What is the greatest possible period of the wave?

Figure 13.40
For question 30.



****31** A transverse wave is travelling from right to left through a series of particles. At a certain instant the waveform is as shown in Figure 13.41. Each of the vibrating particles is observed to perform two complete oscillations in 20 seconds.

(a) Find the following quantities: wavelength, frequency, amplitude, and speed of the wave.

(b) At the instant shown, which of the particles (A–H) are **(i)** moving upward; **(ii)** moving downward; **(iii)** momentarily still?

(c) What will be the position of particle C one-quarter of a period later?

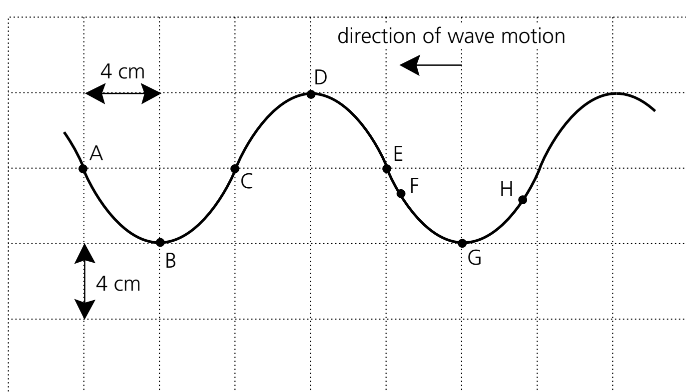


Figure 13.41
For question 31.

***32** Figure 13.42 shows a wave of frequency 10 Hz at an instant in time. The wave is travelling to the right.

(a) What is the wavelength of the wave?
(b) What is the speed of the wave?
(c) Using the letters shown, name any two points on the wave that are in phase.
(d) If a reflecting barrier is placed at F, sketch on an appropriate diagram the shape of the reflected wave.

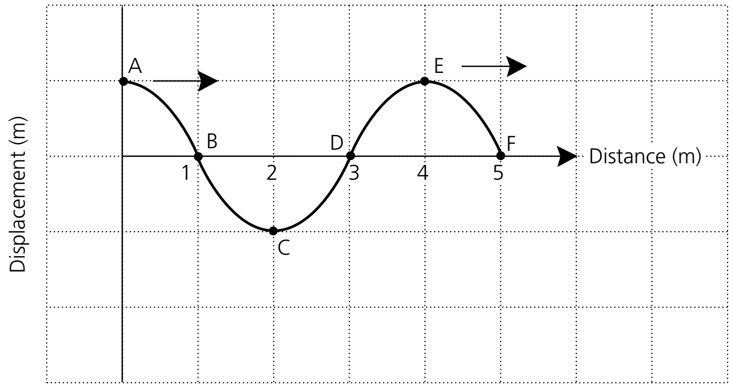


Figure 13.42
For question 32.

****33** Draw neat diagrams to illustrate the reflected pulses in the four situations shown in Figure 13.43. Each pulse is created in a rope.

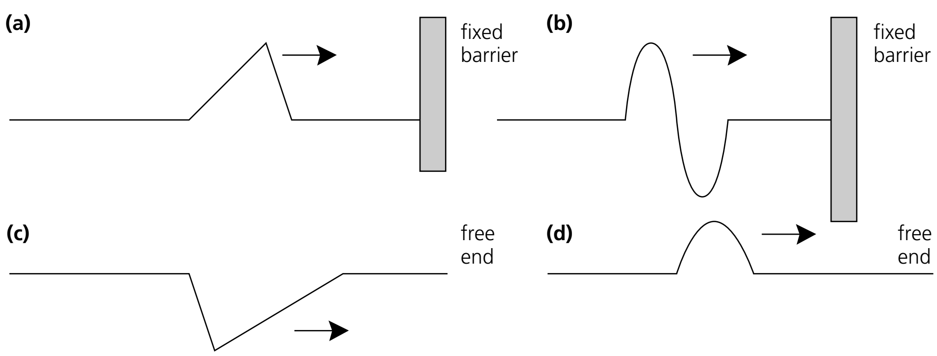


Figure 13.43
For question 33.

****34** A heavy spring is attached to a lighter spring and then fixed to a wall as shown in Figure 13.44. A pulse is sent along the heavier spring towards the wall.

(a) Draw the pulse that is transmitted into the lighter spring.
(b) Draw the shape of the pulse reflected from the lighter spring.
(c) What will happen when the pulse hits the wall?

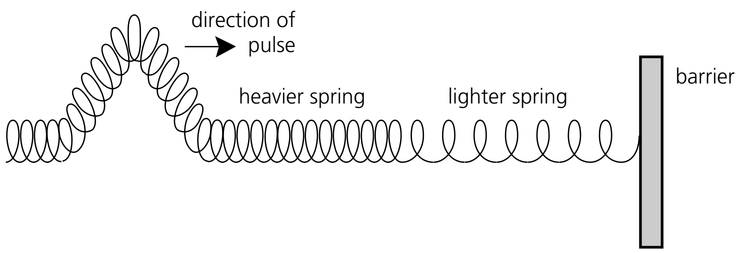


Figure 13.44
For question 34.

Ch13-Walding 4th 25/8/04 11:50 AM Page 303

**35 Figure 13.45 illustrates a pulse that is moving to the right along a stretched rope at a speed of 4.0 cm s^{-1}.
(a) Draw the position of the pulse 1 s later.
(b) If A is a point on the rope, where will it be in 0.50 s?
(c) What is the amplitude of the wave?

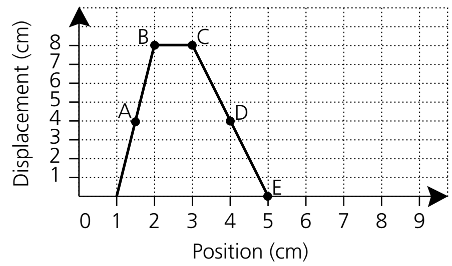


Figure 13.45
For question 35.

36 Figure 13.46 shows the position of a wave in a rope at two instances in time. Determine **(a) the wavelength of the wave; **(b)** the frequency of the wave; **(c)** the amplitude of the wave; **(d)** the speed of the wave.

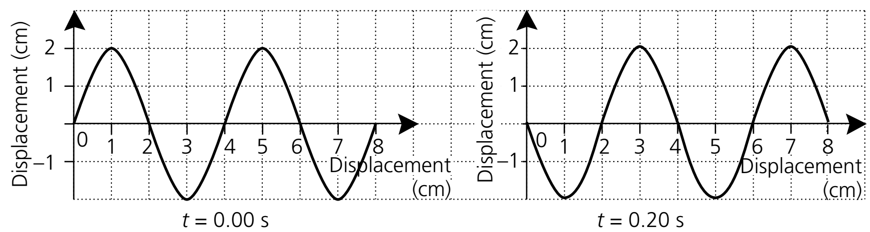


Figure 13.46
For question 36.

Extension — complex, challenging and novel

***37 Students sitting 50 m from the start of an athletic competition hear the starting pistol start the race. (See Figure 13.47.) They hear a second noise 0.9 s later due to the sound being reflected from the grandstand 150 m from the start. What is the speed of sound on this day?

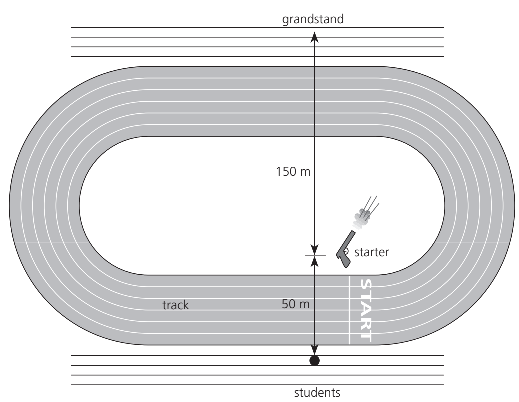
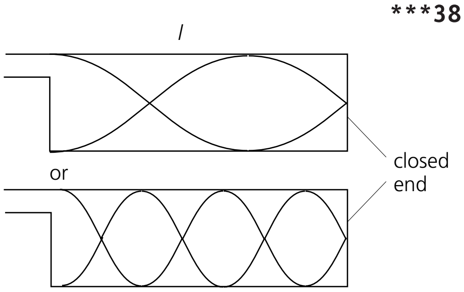


Figure 13.47
For question 37.

Figure 13.48
For question 38.



***38 Sounds are produced in stringed musical instruments by setting up standing waves in strings by plucking them. Longitudinal stationary waves can be produced in the air columns of wind instruments, like a flute, by blowing in them. The air column can vibrate in a number of different ways. An example of two modes of vibration in a closed-end pipe is given in Figure 13.48. Draw the first five different modes of vibration for an open-ended pipe and determine a general formula relating the length of the pipe to the wavelength of the sound produced.

*****39** A fisherman using an echo-sounder to locate fish finds that the reflected pulses return after 0.20 s and 0.25 s. Interpret these two times and determine how far from the ocean floor are the fish (if any).

*****40** Can the 'Mexican wave', historically started at a World Cup soccer match in Mexico in 1986 to distract the competitors, be regarded as a wave? Critically analyse the features of this phenomenon in the light of wave characteristics.

*****41** Figure 13.49 illustrates a pulse moving to the right along a stretched spring at a speed of 5.0 m s^{-1}.

(a) Which of the points A, B, C has the greatest speed at this instant?

(b) Calculate the instantaneous velocity of point C.

(c) Draw the displacement–time graph and the velocity–time graph for the point X on the spring. Take the zero for time at the instant shown in the graph.

(d) Suggest why this particular situation is very unlikely.

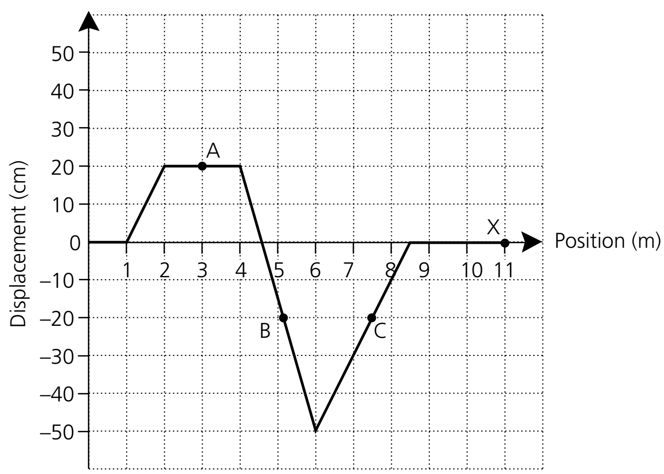


Figure 13.49
For question 41.

CHAPTER 14

Wave Motion in Two Dimensions

INTRODUCTION 14.1

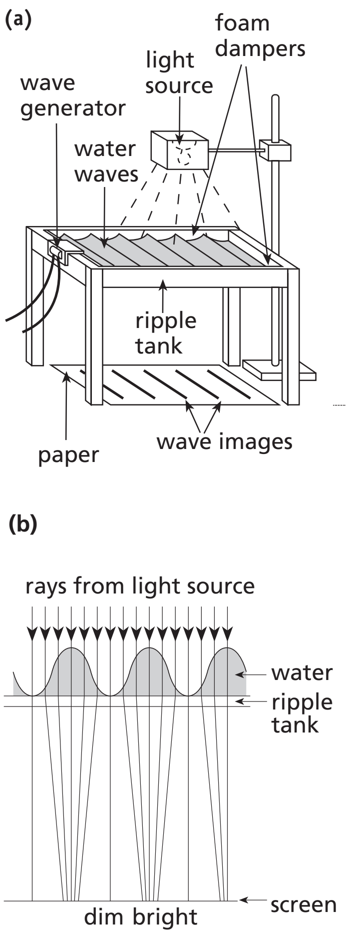
Waves in springs, strings, or even hoses move in only one dimension — along the material and back again. However, sound waves, water waves and light waves can move in any number of directions. This is true for the majority of waves that occur in nature. Can you think of other types of waves that are not confined to movement in one direction?

In this chapter we will look at some of the characteristics of waves that propagate in all directions.

Did you know the following facts?

- Locating earthquakes, using ultrasound on unborn babies, and bats' echolocation have something in common. Can you name it?
- The highest wave ever recorded was 34 m from crest to trough, produced during a hurricane in 1933.
- A landslide in 1958 produced a wave 24 m high in a canyon-like fiord in Alaska.
- The highest wave ever ridden was a tsunami that struck Hawaii in 1868. It was surfed in by a man named Holua, to save his own life.

Figure 14.1
(a) A ripple tank used to study water waves. **(b)** Crests focus light to produce bright regions while troughs spread the light out to produce dark regions.



WATER WAVES — WAVEFRONTS 14.2

The problem with analysing and observing sound and light waves is that they cannot be seen. It is impossible to see and observe light waves. Why is this?

However, water waves can be generated and observed quite readily. All of us at some stage have observed or enjoyed waves at the beach. Surfers use the energy of waves to carry them forward. Many water amusement parks have wave generators to produce waves for the enjoyment of the patrons. In the laboratory, wave generators create water waves in a **ripple tank** (Figure 14.1**(a)**).

The ripple tank consists of a square tray with a glass bottom to allow a light to shine through. The sides are normally metal and are lined with foam rubber to absorb waves and thus stop reflections that may interfere with what is being observed. About 2 cm of water is placed in the tray. A light source is placed above the tray, and shines through the water and transparent bottom onto a screen placed under the tank. Waves are generated by means of an electric motor to which is attached beads to create circular waves, or a straight rod to produce straight waves. These dip into the water to create the waves as the motor turns.

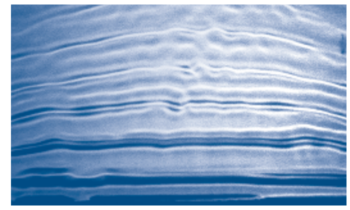
The waves produced in the ripple tank have the same characteristics as any transverse waves. They consist of crests and troughs. The crests act like converging lenses to the light from the light source and focus the light, creating bright areas on the screen. The troughs act like diverging lenses and spread the light out, producing dark areas on the screen, as shown in Figure 14.1**(b)**.

The shape of the **wavefront** depends on the shape of the '**dipper**' producing the wave. If a bead is used, circular wavefronts will be produced. If a straight bar is used, straight wavefronts will be produced as shown in the photo. The distance between crests or wavefronts is a wavelength.

The **crests** of these waves move away from the source. The direction of propagation of the wave is perpendicular to the wavefront, as shown in Figure 14.2. Points on the wavefront are moving in phase. That is, all points are moving the same way, up or down, at the same time.

For a single point source the wavefronts radiate outward, forming circular waves. Other examples of wavefronts radiating outward include those produced when bombs or firecrackers explode.

Photo 14.1
Straight waves and circular waves being generated in a ripple tank.



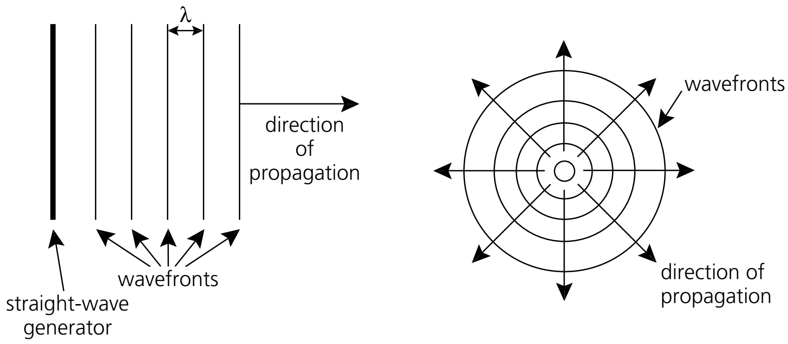


Figure 14.2
A straight-wave generator produces wavefronts that are parallel to the generator. A single dipper produces circular wavefronts that propagate radially.

It is often difficult to measure the wavelength of these waves as they are continually moving. However, by using a stroboscope the wave pattern can be observed more clearly.

Activity 14.1 STROBOSCOPES

A small cross is drawn on a three-bladed electric fan. When the fan is illuminated with a stroboscope, the cross appears stationary at strobe frequencies of 100, 150, 200, 250 and 300 Hz. What is the most likely frequency of the fan blade with the cross on it?

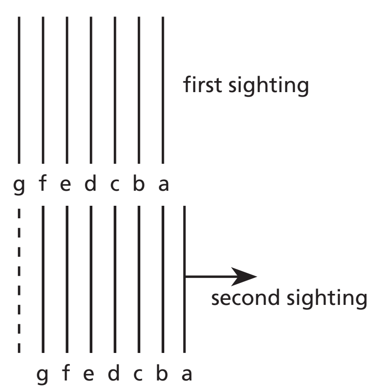
If the hand-held stroboscope is rotated so that the time it takes to rotate the stroboscope from one slit to the next is the same time taken for one wave to move to the position of the wavefront in front (one wavelength), the waves will appear to be stationary. (Refer to Figure 14.3.) This will make measurement of wavelength easier. The frequency of the waves can also be easily measured. If the stroboscope has 10 slits and is rotating at 4 times per second, 40 slits pass the eye in 1 second. Therefore, the time between each sighting through a slit, and thus the period of the waves, is 1/40 second. The frequency is therefore 40 cycles per second or 40 Hz. If electronic stroboscopes are used the job becomes much easier.

The speed of the waves can therefore be calculated using the wave equation, $\boldsymbol{v} = f\lambda$.

If the depth of the water is constant the speed of the wave will be constant, but if the depth of the water varies the speed changes. **Surface waves** on water, which are a mixture of transverse and longitudinal waves, travel more slowly in shallow water and faster in deeper water. If a water wave moves from one depth of water to another it is similar to moving from one medium to another. This will result in the wave being transmitted, and reflected in various ways as discussed in Chapter 13.

For other types of waves the speed depends on other factors. The speed of sound waves depends on the density, pressure and type of gas they pass through. The speed of earthquakes depends on the type of rock through which they move.

Figure 14.3
Waves appear stationary if the strobe is rotated at the correct speed. One wavefront moves to the position of the previous one.



Huygens's principle

Wavefronts are seen to radiate outward from a vibrating source. But how do these wavefronts move?

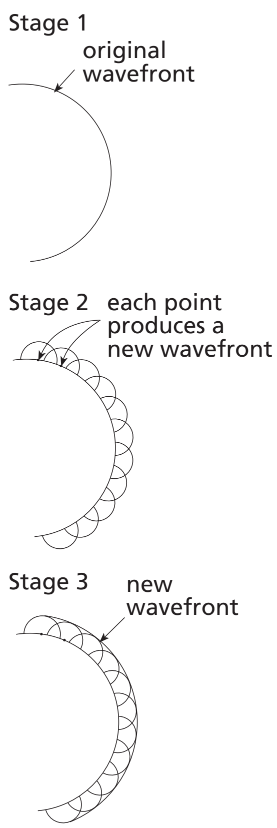
The movement of the source of the disturbance causes those water particles in the near vicinity to vibrate in harmony with the source. These vibrating particles cause those next to them to vibrate. Thus each particle on a wavefront is thought to be the source of a small

Ch14-Walding 4th 25/8/04 11:52 AM Page 307

circular secondary wavelet (Figure 14.4). This principle was put forward by the Dutch physicist **Christian Huygens** (1629–95) in the seventeenth century. Since all points on the wavefront are in phase, each produces a wavelet moving outward with the same velocity. A short time later, after these wavelets have travelled a short distance, the wavelets are connected within a common envelope, producing a new wavefront, which is the tangent to the wavelets. We thus produce a new wavefront travelling outward. All points on this new wavefront then become the source for new wavelets, and the process continues.

Straight waves can be created by dipping a straight bar such as a ruler into the ripple tank. All points on the ruler act as point sources for wavelets. Since all points are in phase the envelope enclosing all wavelets produces a straight wave whose wavefront is parallel to the source and whose direction of propagation is perpendicular to the wavefront.

Figure 14.4
Huygens's principle says that all points on the wavefront produce secondary wavelets, which are the source for the resulting wavefront.



Questions

1 Explain clearly, with the use of diagrams, the terms 'wavefront' and 'direction of propagation'.

2 What is the relationship between the direction of propagation of a wave and the wavefront?

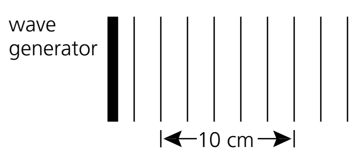
3 A strobe with 10 slits is used to 'freeze' the motion of water waves in a ripple tank. It is found that the highest speed of rotation of the stroboscope needed to produce a stationary wave pattern is 60 revolutions per minute. The wavelength of the waves is measured to be 10 mm. Calculate the speed of the waves in the tank.

4 The wavefronts of straight waves produced in a ripple tank are shown in Figure 14.5. Calculate the speed of the waves if the wave generator produces 5 waves per second.

5 Two dippers used to create waves in a ripple tank are 5.0 cm apart. They are oscillating at the rate of 20 Hz, and the circular waves produced travel at 25 cm s^{-1}.
- **(a)** What is the period of the waves?
- **(b)** What is the wavelength of the waves?
- **(c)** How many wavelengths fit between the two dippers?
- **(d)** If the frequency of the dippers is doubled what is the new wavelength of the waves?

REFLECTION 14.3

Figure 14.5
For question 4.



When a water wavefront meets a fixed barrier perpendicularly it is reflected back with the same velocity. If it meets the barrier at an angle, the reflected wavefront leaves at an angle equal to the angle of incidence. To see how this occurs examine the diagram of a wavefront ABC hitting a barrier (Figure 14.6).

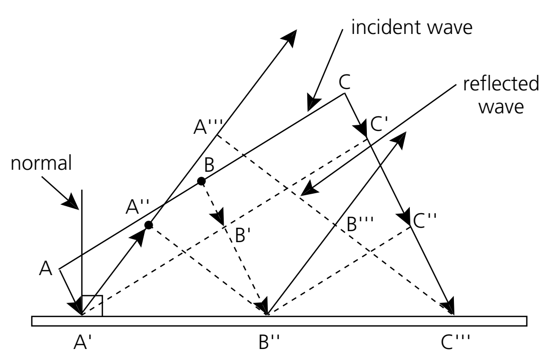


Figure 14.6
The reflection of a straight wave ABC from a straight barrier. The process is complicated by the interference between incident and reflected waves.

Point A is the first point to hit the barrier at A′ and be reflected. It hits the barrier and bounces back at the same angle to the **normal** (the line perpendicular to the barrier) as it collided, as shown in Figure 14.6. When point B on the wavefront hits the barrier at B″, B has travelled a distance of $B'B''$. This means point A has moved away from the barrier the same distance (as all points travel the same speed in the one medium). Point A moves from A' to A'' a distance equal to $B'B''$. The new wave is $A''B''C''$ as shown, with a part of the wave $A''B''$ moving away from the barrier and a part $B''C''$ still moving toward the barrier. A short time later point C hits the barrier at C'''. Point A will then have moved a further distance away from the barrier, a distance equal to $C''C'''$, to point A'''. The entire reflected wave $A'''B'''C'''$ is now moving away from the barrier as shown.

In general, water waves follow the same rules as light when reflected from a mirror, or marbles when thrown at a brick wall. (This will be discussed in Chapter 17.) The angle between the direction of propagation of the incident wave and the normal to the barrier is equal to the angle between the direction of propagation of the reflected wave and the normal as indicated in Figure 14.7.

This is stated as **the angle of incidence equals the angle of reflection**.

You can see how difficult it is to visualise, and explain, what is happening even using simple diagrams, let alone to observe this in a ripple tank. The interactions between the incoming and the reflected waves makes observation and analysis confusing. The reflection of waves and the interaction of incoming waves and reflected waves can be observed in nature when the incoming waves in the ocean interact with those reflected off the headland, off large boats or sea-walls.

Figure 14.7
When a straight wave is reflected, angle *i* equals angle *r*.

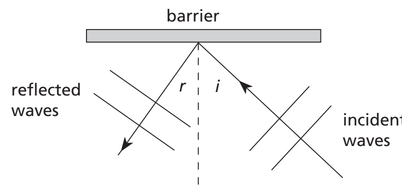
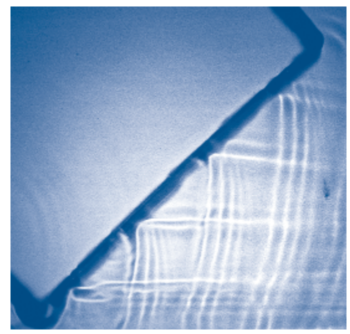


Photo 14.2
Straight waves being reflected from a barrier in a ripple tank.



EI Activity 14.2 REFLECTION

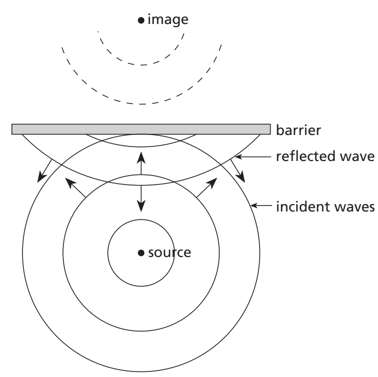
Generate straight waves in a ripple tank and observe the resulting pattern when these waves are reflected from a barrier.

The same principles of reflection apply when circular waves interact with straight barriers or when straight waves reflect from curved barriers, as shown in Figure 14.8.

Notice for the circular wave in Figure 14.8 that the front part of the wave hits the barrier first, therefore it is reflected first, creating a curved reflected wave that gives the impression that it was made by a source behind the barrier.

Straight waves reflecting from a curved barrier result in the focusing of the waves to a point. After passing through the point where this occurs (the **focal point**) the sides get further behind, resulting in a curved wave. (See Figure 14.9). Curved waves generated at the focal point and reflecting from the correctly curved barrier can produce straight waves, as shown in Figure 14.10.

Figure 14.8
When a curved wave reflects from a straight barrier the reflected wave appears to originate from a point behind the barrier.



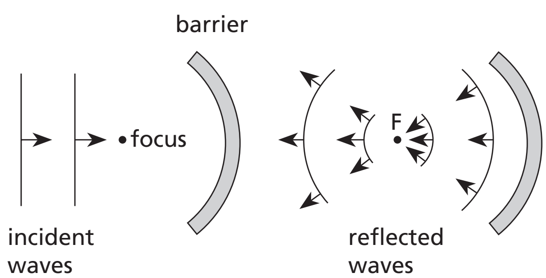


Figure 14.9
A straight wave becomes curved and passes through the focal point when reflected from a curved barrier.

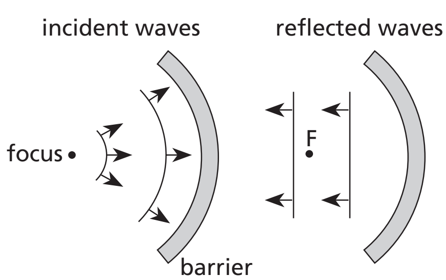
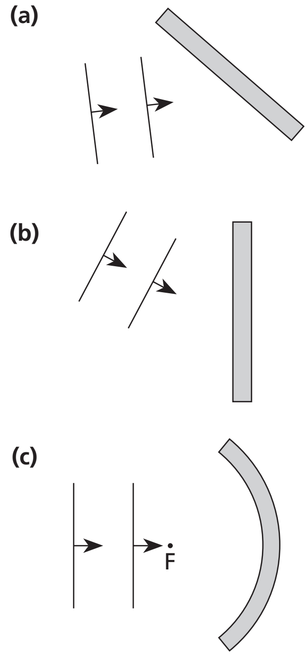


Figure 14.10
A wave originating from the focal point of a curved surface is reflected as a straight wave.

Figure 14.11
For question 8.



Questions

6 Give a definition of the terms 'normal', 'angle of incidence', and 'angle of reflection'. Use a diagram to help to explain the meaning of these terms.

7 A straight wave strikes a straight barrier at an angle of incidence of 25°. Draw a diagram to show the incident and reflected waves.

8 For each situation in Figure 14.11 where waves are incident on reflecting barriers, draw the wavefronts of the waves after reflection.

REFRACTION 14.4

When waves pass from one medium to another further properties of waves can be observed. To change the medium for water waves only requires changing the depth of the water. This can be accomplished by placing a sheet of glass or Perspex in the ripple tank so the depth of water over it is less than the surrounding water in the tank. This divides the tank into two areas, as shown in Figure 14.12.

If straight waves are created so that the wavefronts hit the shallow area parallel to the boundary, they will pass from the deep region into the shallow region without change in direction. However, waves travel more slowly in shallow water compared with deep water, and since the frequency of waves in both regions is the same because they are produced by the one source, the wavelength of the waves changes. Since $\boldsymbol{v} = f\lambda$, and because '$\boldsymbol{v}$' becomes less in the shallow water, and 'f' remains the same, 'λ' must decrease. The resulting pattern is shown in Figure 14.13. Table 14.1 shows the relationship between the depth of water, speed, and wavelength of waves.

Figure 14.12
Placing a piece of Perspex or glass in a ripple tank divides the water into two regions of different depths so refraction of waves can be investigated.

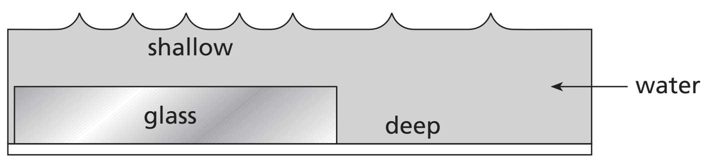
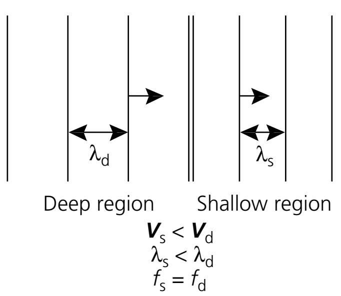


Table 14.1

Depth	shallow	deep
Wavelength	short	long
Speed	slow	fast

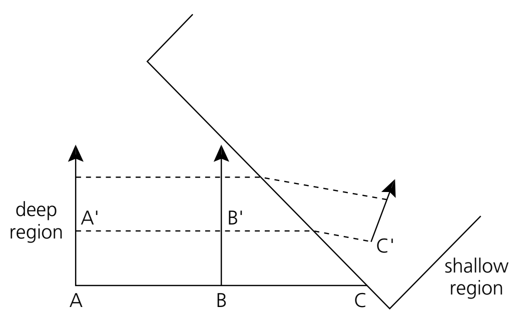
Figure 14.13
When a wave passes from one depth to another its speed and wavelength change.



Remember it this way: shallow, short, slow (the 3 s rule).

The reverse is also true. When waves go from a shallow to a deeper region they speed up and their wavelength increases. However, if the waves hit the junction between media at an angle other than 90° they change direction. This change in direction of the waves as they go from one medium to another is called **refraction**, which comes from the Latin *refractus* meaning 'broken off'. This property is explained in Figure 14.14.

Figure 14.14
When a wave hits the boundary between two depths of water at an angle, it changes direction.



When wave ABC, travelling in a deep region, hits a junction between a deep and shallow region at an angle, as shown, C hits the junction first. A short time later A and B have moved a small distance to A′ and B′, but because C is travelling in a medium where the wave travels more slowly it has fallen behind and only moved to C′. The wave now becomes A′B′C′. As the direction of propagation of the wave is perpendicular to the wavefront, the wave has changed direction in the shallower region.

If a set of periodic waves is moving from one medium (deep) to another (shallow) as shown in the photo, all waves change direction and the wavelength decreases.

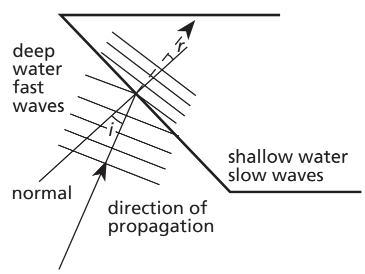
The size of the change in direction and the wavelength will depend on the relative change in depth of the water. If the difference between the depths of water is large the waves change direction and wavelength a great deal.

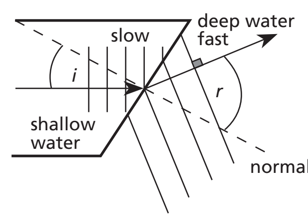
Again the reverse is also true. If waves move from a shallow to a deeper region at an angle to the junction, the wavelength becomes larger as the waves get further ahead. The angle between the direction of the propagation of the incident wave and the normal is called the **angle of incidence**. The angle between the normal and the direction of propagation of the refracted wave is called the **angle of refraction** (see Figure 14.15). These will be discussed more fully in Chapter 17 (Optics) where they can be viewed and measured more easily. When using light, refraction will occur when the waves propagate across the junction between air and glass, for example.

Note, however, that in going from deep to shallow water the angle of refraction is smaller than the angle of incidence. You can add this other ‘s’ word to your list: shallow, short, slow, and small. Note that while speeds, wavelengths and angles change, frequency does not.

Photo 14.3
Straight waves being refracted as they go from a deep area to a shallow area in a ripple tank.

Figure 14.15
The wavelength decreases when waves go from deep to shallow water, and increases when going from shallow to deep water.





Seismic waves

The study of waves produced by earthquakes (**seismology**) has been, and is, very important in determining the structure and properties of the Earth. Earthquakes produce three types of waves that radiate out at high speeds from the epicentre of the earthquake. They are P (primary) waves, which are longitudinal waves, S (shear) waves, which are transverse waves, and surface waves, which travel along the surface of the Earth and are transverse and more complicated elliptical waves. The speeds of these waves are given in Table 14.2. However, their speeds and direction are affected by the material in which they move.

Table 14.2 THE APPROXIMATE SPEEDS OF SEISMIC WAVES

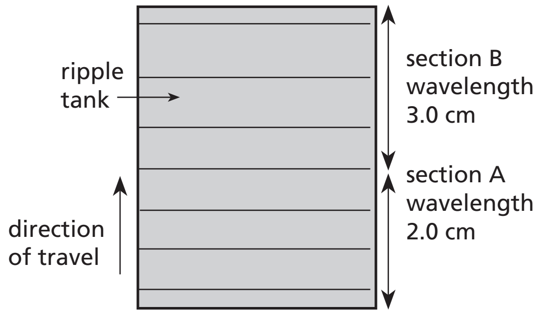
WAVE TYPE	WAVE SPEED (km s^{-1})
P wave	8.0
S wave	4.5
Surface wave	2.0

Seismic waves can be detected by seismographs, which record the movement of the Earth. Because seismic waves travel through the whole planet and can be detected at various stations on the Earth's surface, a great deal of information about the Earth's crust, mantle, and core has been found. Continual research has identified the depth of the crust, the size of the mantle, the existence of a liquid outer core, and even the thickness and probable constituents of finer layers within the mantle.

Questions

9 Figure 14.16 indicates the position of straight waves in a ripple tank at a particular instant. The velocity of the waves in section A is 10 cm s^{-1}.
- **(a)** Explain the reason for the different wavelengths in the two sections.
- **(b)** Calculate the frequency of the waves in section A.
- **(c)** What is the frequency of the waves in section B?
- **(d)** What is the speed of the waves in section B?

Figure 14.16
For question 9.

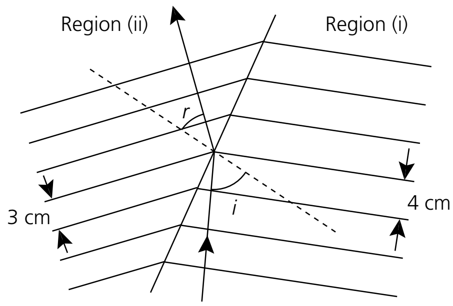


10 Straight waves of wavelength 2.5 cm have a speed of 8.0 cm s^{-1} in a region in a ripple tank where the depth of water is 2.0 cm. They pass into a region of depth 1.0 cm where the speed is determined to be 6.0 cm s^{-1}. Find the wavelength of the waves in the new region.

11 Straight waves in a ripple tank move from region **(i)** where their wavelength is 4.0 cm to region **(ii)** where their wavelength is 3.0 cm (Figure 14.17).

(a) Which region is the deeper?

Figure 14.17
For question 10.



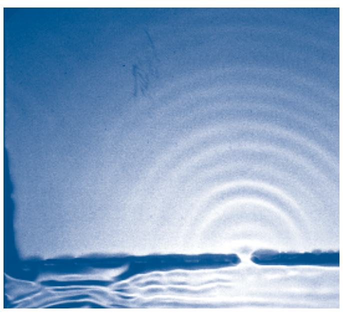
(b) Calculate the ratio of the wavelength in region **(i)** to that in region **(ii)**.

(c) Calculate the ratio of the speed of the waves in region **(i)** to the speed of the waves in region **(ii)**.

(d) Calculate the ratio of the frequency of the waves in region **(i)** to the frequency in region **(ii)**.

(e) What is the name given to this phenomenon?

Photo 14.4
Straight waves passing through a small opening in a barrier in a ripple tank are diffracted.



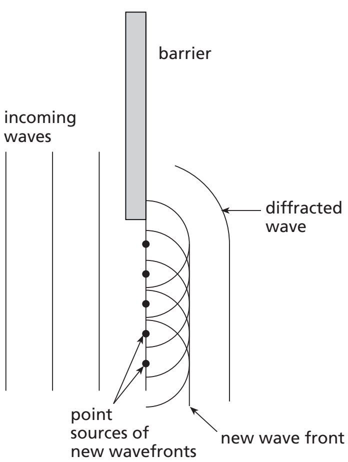
DIFFRACTION OF WATER WAVES 14.5

Another property of water waves is often observed when waves from the ocean enter an inlet. They form a circular pattern. It can be easily observed in the ripple tank by placing two large blocks of glass or metal obstacles, to produce a slit, in front of the straight waves. The waves, as they pass through the slit, produce a circular wave, as shown in the photo.

This bending of waves as they pass through a slit is called **diffraction** (from the Latin *diffractus* meaning 'to break apart'). Similar bending occurs if the waves pass around the end of an obstacle.

Diffraction can be explained in terms of Huygens's principle. As a straight wavefront enters the aperture (the slit), a secondary wavelet is produced by each point on the wavefront in the aperture. This continues to occur as the wavefront travels outward. The envelope enclosing the wavelets adds, to produce a straight wave in the centre but the edges remain curved. The wave curves around the aperture's edges, as shown in Figure 14.18.

Figure 14.18
Diffraction of a wave at an edge can be explained by Huygens's principle.



Changing the slit width

It can be easily reasoned using Huygens's principle that if the aperture was smaller the shape of the waves passing through the aperture would be more curved, more circular, and if the aperture were larger the resulting waves would be straighter except for the edges. This is in fact what occurs. This can be observed easily in the ripple tank by changing the position of the obstacles making up the slit. Figure 14.19 shows the resulting patterns.

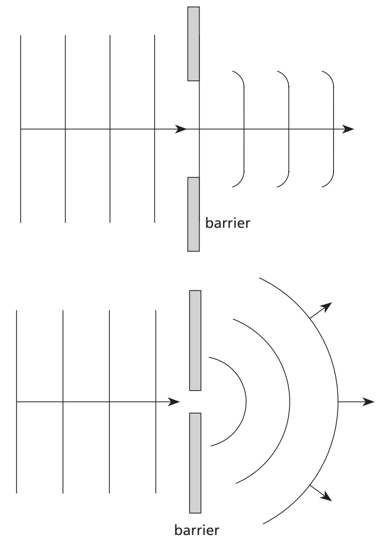


Figure 14.19
Diffraction is more noticeable when the size of the slit is comparable to the wavelength of the waves.

Changing the wavelength

Changing the wavelength of the waves also affects the diffraction pattern. Diffraction is more noticeable, that is, a more circular wave pattern is produced, if the wavelength is equal to or greater than the opening. It is the relative difference in the size of the wavelength and the size of the slit that is important (Figure 14.20).

A diffraction pattern can be observed in the ocean around boats, large rocks and buoys. However, the amount of diffraction depends on the size of the objects compared with the wavelength. If the object is large compared with the wavelength a significant diffraction pattern occurs around the edges of the object, producing a shadow zone.

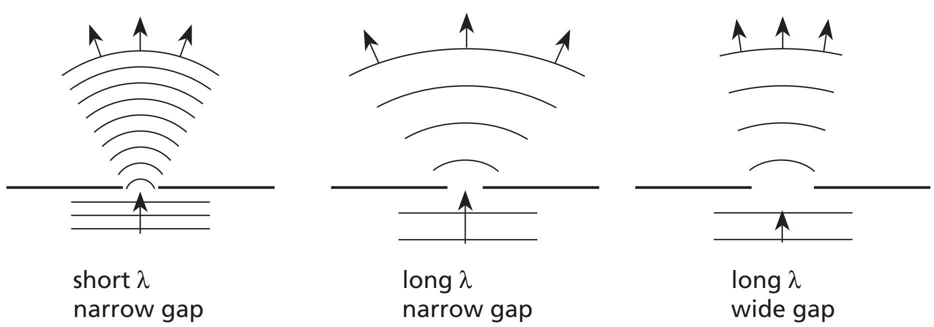


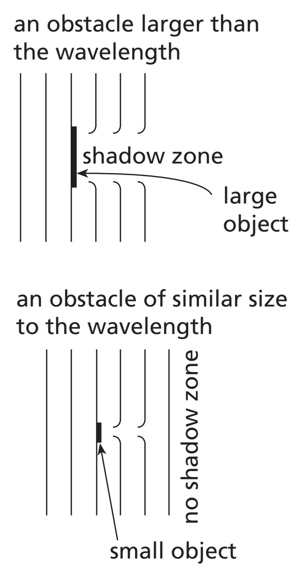
Figure 14.20
Diffraction is greater when the wavelength is large and the gap narrow.

Why don't body surfers produce diffraction patterns and shadow zones?

Diffraction of water waves in the ocean is partly responsible for the formation of many offshore islands. Consider a land formation consisting of a peninsula as shown in Figure 14.22. Straight waves parallel to the shoreline would diffract around the front of the peninsula, thus breaking on the sides of the peninsula and causing erosion. Over thousands of years the erosion would eat into the peninsula, cutting off the land and forming an island. Constant diffraction around these islands would also affect the shape of the islands.

It is advantageous to students who sit in the back of the classroom that the wavelength of common human speech, about 1 m, is larger than common classroom objects: tables, chairs, and even students. Remember, if the wavelength of waves is smaller than, or similar in size to, objects placed in their path, diffraction causes shadow zones. (See Figure 14.21.) However, if the object is much smaller than the wavelength, this effect is unnoticeable. Think what would happen if sound waves produced by the teacher were smaller than the student themselves, or their chairs, tables etc. What would the students in the back of the room hear? Can you use this as an excuse for why you cannot hear, particularly if you have a large student sitting in front of you?

Figure 14.21
An obstacle affects the waves if it is large compared with the wavelength of the waves, producing a shadow zone.



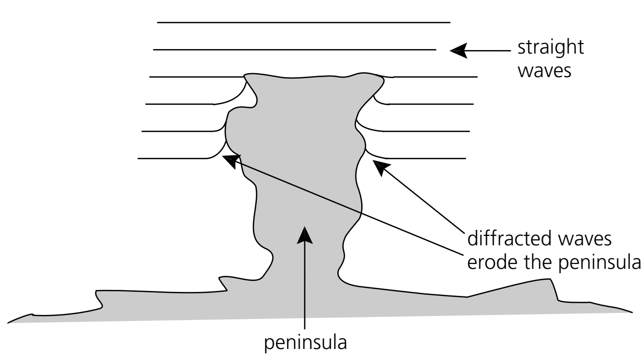


Figure 14.22
Straight waves in the ocean diffract around a peninsula, causing erosion on the sides, eventually cutting off the peninsula.

INTERFERENCE OF WATER WAVES 14.6

Photo 14.5
Interference produced by two sources of waves in a ripple tank.

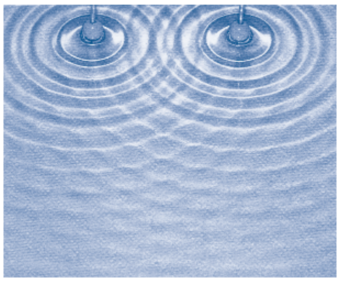


Figure 14.23
The interference pattern produced by constructive and destructive interference of waves from two sources.

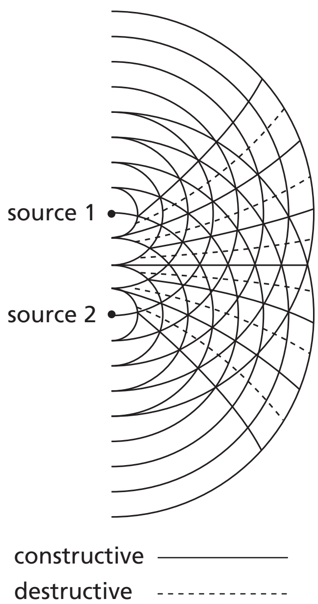


Figure 14.24
Interference pattern of waves from two sources.

As seen in Chapter 13, when waves meet they may reinforce or cancel each other out (superposition), but then continue on as though they had not met. If they are the same shape, and amplitude, and on the same side, they will **constructively interfere** to produce a wave of twice the amplitude, an **antinode**. If they are on opposite sides they will **destructively interfere**, cancelling each other out, producing a **node**. Two-dimensional water waves undergo the same phenomena but the pattern is more complicated as the waves from two sources radiate outward in all directions.

To produce waves in a ripple tank that are in phase and have the same amplitude involves the use of a wave generator that has two beads (dippers) attached to a straight rod. As the electric motor moves the rod up and down the two beads dip in and out of the water, producing two sets of circular radiating waves that are in phase and of the same amplitude. Another term for in-phase is **coherent**.

As the waves radiate outward, a crest from one dipper will meet a crest from the other, and constructive interference occurs, producing a larger crest. The same happens when a trough from one meets a trough from the other.

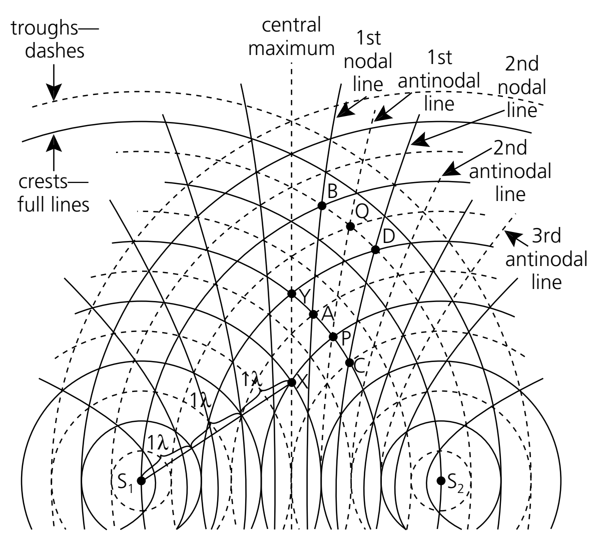
When a crest from one meets a trough from the other, destructive interference occurs and the waves cancel.

As the waves radiate outward they continue to add and cancel. The resulting pattern that occurs is the characteristic interference pattern produced by two sources in phase. (See Photo 14.5.)

It will be noticed in this photograph that there are regions that are bright, dark and grey. The grey or shadow areas are areas where cancellation has occurred, that is, where a trough from one source cancels a crest from the other. A bright area is where two crests meet, producing a larger crest. This acts like a convex lens and focuses the light on to the screen below the ripple tank, forming a bright area. When two troughs meet they produce a larger trough, which spreads the light out to form a dark region, as previously shown in Figure 14.1.

The resulting pattern can be drawn schematically with lines through the grey areas of undisturbed water, as shown in Figure 14.23. These lines are called **nodal lines**. Lines can also be drawn through the bright and dark regions — regions of constructive interference. These are **antinodal lines**.

You may understand this better when we consider some points on the pattern (Figure 14.24). Consider point X on the central antinodal line (**the central maximum**). This point is a distance of S_1X from source S_1 and a distance S_2X from source S_2. It will be noticed that X is 3λ from S_1 and is 3λ from S_2. The difference in distance from the two sources, the **path difference** (PD), is zero wavelengths.



Now look at a point Y further out but still on the central maximum. The difference in distance from the two sources S_1 and S_2 is $S_1Y - S_2Y = 4\lambda - 4\lambda = 0\lambda$.

If this is tried again it will be found that the path difference for all points on the central maximum is always zero wavelengths.

Now consider point P on the first antinodal line. The distance from S_1 to P (S_1P) is 4λ and the distance from S_2 to P (S_2P) is 3λ. The path difference is $S_1P - S_2P = 1\lambda$. If we find the path difference for point Q, further out on the first antinodal line, we will again find it to be 1λ. For all points on the first antinodal line the path difference is 1λ.

NOVEL CHALLENGE

To calculate the total number of nodal lines produced by two coherent point sources, let $\theta = 90°$, calculate 'n' and multiply by 4 (because there are four lots of 90° in a circle). Alternatively, calculate the angle between two nodal lines and divide into 360°. Each method gives a slightly different result. Why is this?

SR Activity 14.3 ANTINODAL LINES

Find the path difference for points on the second and third antinodal lines.

Antinodal line formula A general formula for antinodal lines becomes: **the path difference for points on the nth antinodal line = $n\lambda$.** Where n = 1, 2, 3, . . .

SR Activity 14.4 NODAL LINES

1 Repeat the above exercise to find the path difference for points A and B on the first nodal line and for points C and D on the second nodal line.

2 Derive a general formula for points on the 'nth' nodal line.

Nodal line formula The general formula obtained should be: **the path difference for points on the nth nodal line = $(n - \frac{1}{2})\lambda$.** Where n = 1, 2, 3, . . .

Example

Figure 14.25 shows the interference pattern produced by two coherent sources of waves in a ripple tank. The lines represent wave crests.

(a) Is constructive or destructive interference occurring at points **(i)** A; **(ii)** B; **(iii)** C?

(b) On what order nodal or antinodal line do the following points lie: **(i)** B; **(ii)** C?

(c) If the distance from S_1 to A is 4.0 cm and S_2 to A is 10 cm what is the wavelength of the waves?

(d) If the distances to an unknown point P are S_1P = 8 cm, S_2P = 16 cm, on which nodal or antinodal line does point P lie?

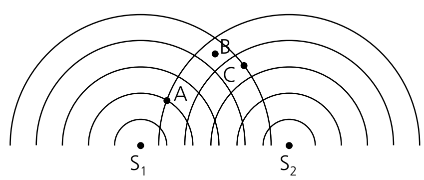


Figure 14.25
For example question.

Solution

(a) **(i)** Constructive, as A lies on a double crest.
(ii) Constructive, as B lies on a double trough.
(iii) Destructive, as C lies on a crest from S_1 and a trough from S_2.

(b) **(i)** B lies on the central maximum line as the path difference is zero wavelengths.
(ii) C lies on the second order nodal line as the path difference is one and a half wavelengths, $(n - \frac{1}{2})\lambda = 1\frac{1}{2}\lambda$.

(c) The distance from S_1 to A is 4 cm, which equals 2λ, then λ = 2 cm.

(d) The path difference is 16 cm – 8 cm = 8 cm, which equals 4λ. Therefore, this point lies on the 4th antinodal line.

These formulas can only be used when the two sources are in phase. If one source is generating waves at a different time from the other, that is, they are out of phase, the formula needs to be modified.

INTERFERENCE OF SOUND WAVES 14.7

Interference of sound waves can be observed very easily in the laboratory using two speakers connected to a single generator, as shown in Figure 14.26. It is important that the speakers are in phase. This can be checked by taking the grilles off the front of the speakers and checking to see that they both move in and out together when on a low frequency, say 1 Hz.

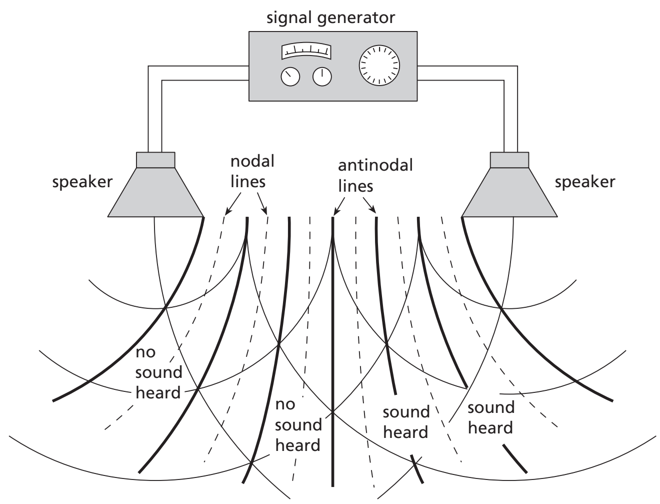


Figure 14.26
A single signal generator and two speakers produce coherent sound waves. There is an interference pattern containing nodal lines where no sound is heard.

Nodal lines are produced. If a person walks across the room in front of the speakers different degrees of loudness can be detected.

- Then why doesn't the interference of waves work at home?
- Why don't you observe this phenomenon when you move in front of the two speakers of your stereo?
- Why don't you see dark and bright light bands when you walk between two wall mounted lights in the living room at home?

The answer to the second question has already been discussed. Revise the previous section. The answer to the third question will be discussed in Chapter 15.

Questions

12 Figure 14.27 shows an interference pattern produced by two coherent sources.

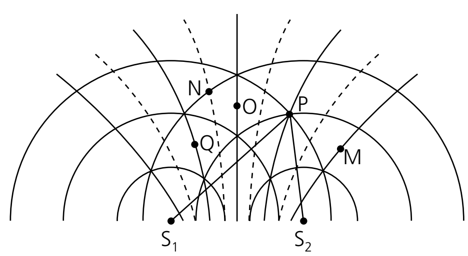


Figure 14.27
For question 12.

(a) Which of these sets of lines, the dotted lines or full lines, represents nodal lines? Why?

(b) If the distance from S_1 to point P is 6 cm and from S_2 to P is 4 cm, determine the wavelength of the waves.

(c) Is constructive or destructive interference occurring at points **(i)** M; **(ii)** N; **(iii)** O; **(iv)** Q?

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***13** What is meant by the terms 'wavelength' and 'direction of propagation'? Use diagrams to assist in the explanation of these terms with reference to straight and circular water waves.

***14** What is meant by the terms 'incident wavefront' and 'reflected wavefront'? Explain these terms with the use of a diagram.

***15** A dipper used in a ripple tank experiment dips into the water at the rate of 12 times in 2.0 seconds. At the end of that 2.0 seconds the outermost wavefront is found to be 15 cm from the source. Determine **(a)** the velocity of the waves; **(b)** the period of the waves; **(c)** the wavelength of the waves; **(d)** the frequency of these waves.

***16** A continuous set of straight waves is produced in a ripple tank by using a vibrator. After 1.2 s the furthest wavefront is found to have travelled a distance 6.0 cm from the source. The distance between successive crests is measured to be 1.2 cm. Calculate the speed and frequency of the waves.

***17** A set of straight waves strikes a straight barrier so that the wavefronts make an angle of 40° with the barrier. Show the incident wave and the reflected wave on a diagram.

***18** Periodic straight waves are generated in a ripple tank. They are seen through a stroboscope with 10 slits. The stroboscope is turned at its fastest speed to freeze the motion of the wave without changing the wave pattern. It is found that the stroboscope is turned at a speed of 25 revolutions in 10 seconds. Calculate **(a)** the frequency of the waves; **(b)** the speed of the waves if the distance between successive crests is 2.5 cm.

***19** A ripple tank is divided into a deep and a shallow region, by placing a thin sheet of glass to cover half the tank. In the deep region waves are found to have a velocity of 8 cm s^{-1}, and a wavelength of 6 cm. In the shallow region they have a speed of 6 cm s^{-1}.
(a) What is the frequency of the waves in the deep region?
(b) What is the wavelength of the waves in the shallow region?

***20** **(a)** Explain with the use of diagrams the meaning of diffraction.
(b) Give two examples of where diffraction occurs in nature.
(c) Show with the use of diagrams how the diffraction pattern depends on the size of the opening and the size of the wavelength of the waves.

***21** If a series of straight waves passes through a narrow slit in a ripple tank, what changes occur to the following properties of the waves as they pass through the slit: **(a)** wave pattern; **(b)** wave speed; **(c)** wavelength; **(d)** period?

***22** If noticeable diffraction of waves through an aperture is to occur, state the conditions that are necessary.

***23** How does the frequency of straight waves change as they are diffracted passing through a small opening?

***24** What is the difference between 'constructive' and 'destructive' interference?

***25** What conditions are necessary for points within an interference pattern to lie on **(a)** the 6th antinodal line; **(b)** the 3rd nodal line; **(c)** the *n*th antinodal line?

***26** Two point sources continually vibrate in phase in a ripple tank. An interference pattern with eight nodal lines is produced. What will happen to the number of nodal lines in each of the following cases?
(a) The frequency is halved.
(b) The distance between the sources is doubled.
(c) The wavelength is doubled.

NOVEL CHALLENGE

Make up five questions that would test a person's understanding of wave motion. Think about whether your questions are just about recall of facts or are really testing the person's understanding.

Ch14-Walding 4th 25/8/04 11:52 AM Page 317

Figure 14.28
For question 27.

***27** Draw the reflected wavefronts after reflection in the situations illustrated in Figure 14.28.

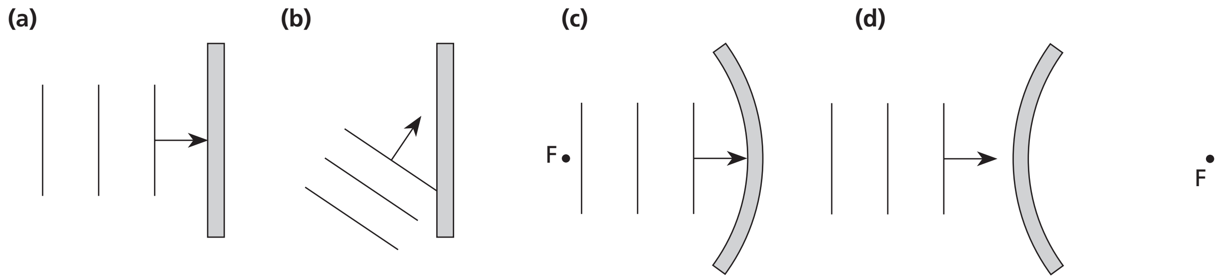
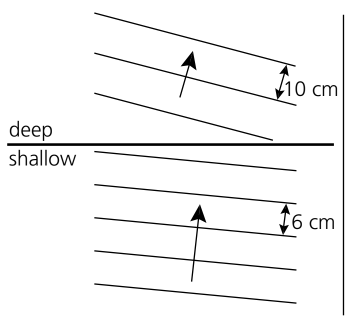


Figure 14.29
For Question 28.



***28** Figure 14.29 shows a set of waves going from shallow water to deep water. Find the ratio of:

(a) $\dfrac{\text{the frequency of waves in shallow water}}{\text{the frequency of waves in deep water}}$

(b) $\dfrac{\text{the velocity of waves in shallow water}}{\text{the velocity of waves in deep water}}$

***29** Straight waves of frequency 10 Hz are produced in the deep end of a ripple tank. These waves move from the deep end to the shallow end (Figure 14.30). Calculate the speed of the waves in the deep and the shallow regions of the ripple tank.

***30** Figure 14.31 represents three ripple tanks containing barriers to form apertures to straight waves. Redraw the diagrams to show the resulting wave pattern after the waves have passed through the openings.

Figure 14.30
For question 29.

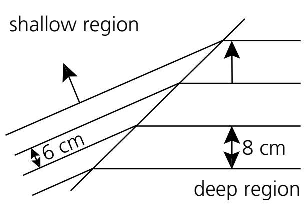
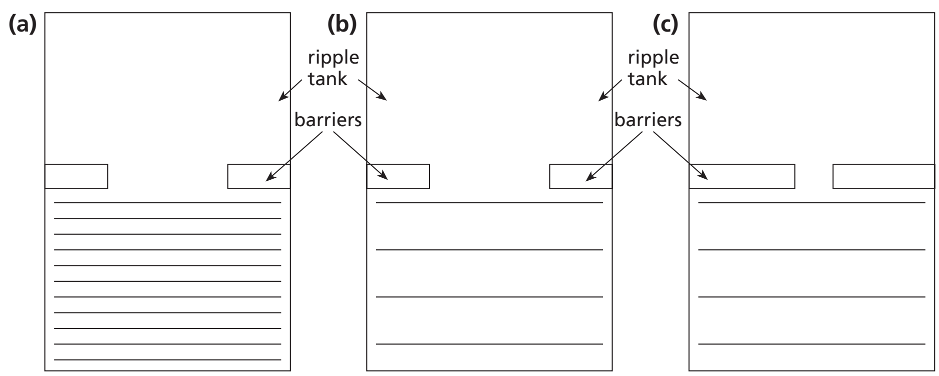
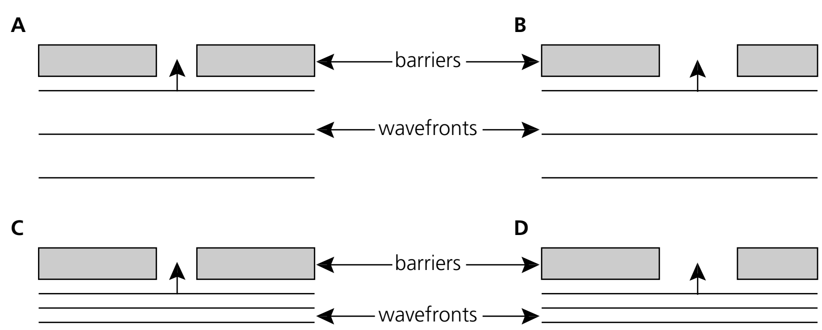


Figure 14.31
For question 30.



***31** Figure 14.32 shows sets of waves incident on openings between pairs of barriers. In which of the situations will diffraction most likely be more noticeable? State reasons.

Figure 14.32
For question 31.



***32** Straight waves in a ripple tank approach a straight barrier which is parallel to the wavefronts. There is a gap of width 'w' in the barrier. The wavelength of the waves is λ. In which of the cases in Table 14.3 will the waves be most strongly diffracted?

Table 14.3 DATA FOR QUESTION 32

	A	B	C	D	E
Wavelength (cm)	1.0	1.5	2.5	0.5	2.0
Width (cm)	2.0	2.0	2.0	4.0	4.0

***33** An interference pattern is shown in Figure 14.33. Are the points A, B, and C on nodal or antinodal lines?

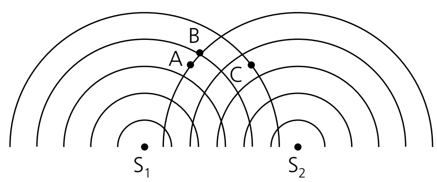


Figure 14.33
For question 33.

***34** Two dippers are separated by 8.0 cm in a ripple tank. The generator causes the dippers to oscillate 100 times in 10 seconds, and the circular waves produced travel at 20 cm s^{-1}.
(a) What is the frequency of the waves produced?
(b) What is the wavelength of the waves in the ripple tank?
(c) Draw a diagram of the resulting interference pattern labelling the nodal and antinodal lines.
(d) How many nodal lines are formed?
(e) How do the above answers change if the frequency of the dippers is doubled?

***35** Figure 14.34 shows the pattern produced when two sets of circular waves are produced by dippers in phase in a ripple tank.
(a) At point A, is constructive or destructive interference occurring?
(b) At point B, is constructive or destructive interference occurring?
(c) Is point C on a nodal line or an antinodal line? Which order line?
(d) Which nodal or antinodal line is point D on?
(e) If the frequency of the dippers is decreased what would happen to the interference pattern?
(f) If dipper X were producing waves half a period after dipper Y what would happen to the interference pattern?

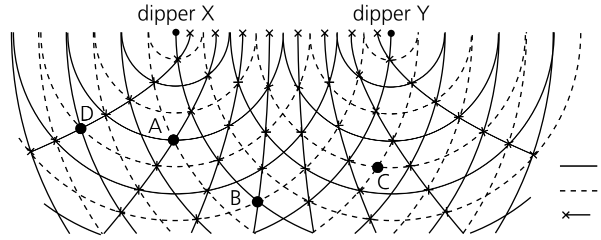
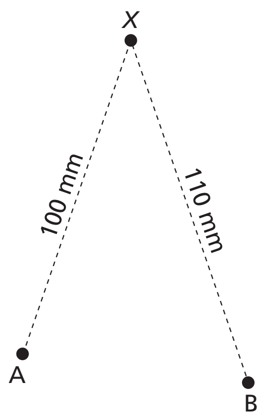


Figure 14.34
For question 35.

Figure 14.35
For question 36.



****36** Two point dippers A and B are driven by a vibrator to produce waves that are in phase and of the same frequency (10 Hz), in a ripple tank (Figure 14.35).

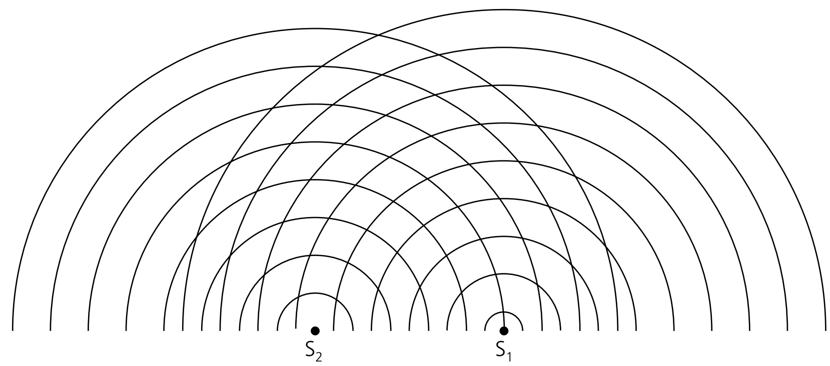
(a) If a wave takes 0.50 s to go from A to point X a distance of 100 mm, what is the speed of the waves?
(b) What will be the speed of the waves from source B?
(c) Determine the wavelength of the waves.
(d) Does constructive or destructive interference occur at point X?
(e) Which nodal or antinodal line does point X lie on?
(f) If the frequency of the sources is doubled indicate what would happen to the interference pattern. Which nodal or antinodal line does point X now lie on?

Extension — complex, challenging and novel

*****37** In Figure 14.36 S_2 is producing crests when S_1 is producing troughs. They are out of phase. The resulting wave pattern is shown.

(a) Draw the first three nodal and antinodal lines.
(b) By choosing a range of points on the first, second, and third antinodal lines and the first, second, and third nodal lines, develop a general formula for the path difference between points on the *n*th antinodal and the *n*th nodal line, for this out-of-phase situation.

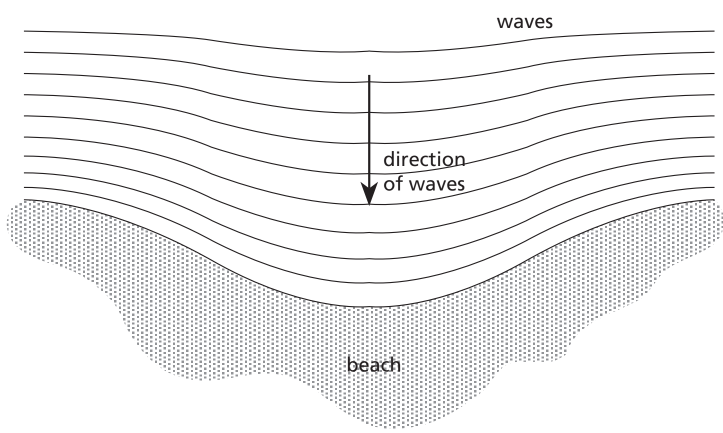
Figure 14.36
For question 37.



*****38** It is noticed in areas close to airports that television reception is distorted when aircraft fly overhead but radio reception remains unaffected. Explain how this might occur.
Hint: you might consider calculating approximate wavelengths of radio waves and television waves.

*****39** Wavefronts approaching a beach form a pattern similar to that shown in Figure 14.37. Analyse the wave pattern and deduce, with explanation, the structure of the beach that would produce such a pattern.

Figure 14.37
For question 39.



CHAPTER 15

Light — A Wave?

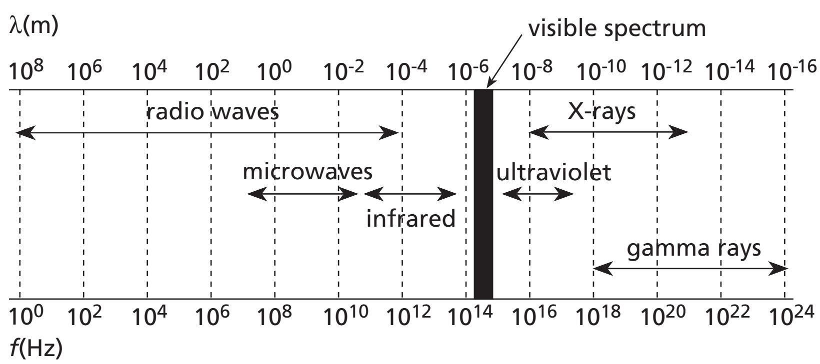
15.1 INTRODUCTION

Have you ever wondered about the following things concerning light?

- If a camera lens is made of clear glass, why does it look purple?
- Why do soap bubbles look so colourful?
- If you had a powerful enough microscope, could you see a single atom?
- How can light be both a wave and a particle? Surely it is one or the other?

In Chapter 14 we investigated the properties of two-dimensional waves; in particular, water waves. The reason for this was that they are easier to observe and investigate in the laboratory. It is now time to investigate the properties of light; in particular, visible light. However, visible light, the light that enables us to see objects, is just a small part of all the electromagnetic waves that are around us. Radio waves, microwaves, infrared waves, for example, all travel through space at the speed of 3×10^8 m s^{-1} and make up a part of that group of waves called electromagnetic waves.

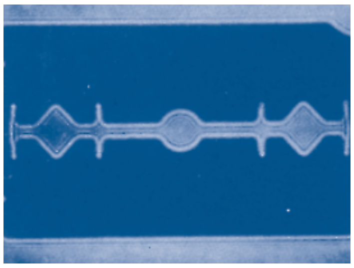
Figure 15.1
The electromagnetic spectrum.



But we seem to be jumping the gun. We are assuming that light travels through space by means of wave motion and not as particles. The particle nature of light is another issue and will be taken up in a later chapter. However, we can show that light does exhibit wave characteristics similar to those of water waves. Again, it is impossible to investigate the properties of all electromagnetic waves in a school laboratory, but we can investigate visible light waves as several of their effects can be seen with the unaided eye.

THE PROPERTIES OF LIGHT WAVES 15.2

Photo 15.1
Diffraction of light through a razor blade — the diffraction of light occurs around the edges of objects and through small apertures.



Light is a form of energy that propagates (travels) through empty space. The propagation of this energy does not require a medium, as proven by light energy from the Sun being able to reach us here on Earth where it can be converted to other forms such as heat used in solar hot water systems, or to electrical energy used by solar powered cars to race across the Northern Territory.

Several of those properties of water waves investigated in Chapter 14 can be applied to light and observed in the laboratory but this requires detailed observation as the wavelength of visible light waves ranges from 4×10^{-7} m to 7×10^{-7} m. These waves are much too small to be seen with the unaided eye.

Light waves or light rays can, like water waves, be reflected and refracted. This can very easily be seen in the laboratory using a laser or light boxes. However, these properties are not exclusive to wave characteristics and will be discussed in Chapters 17 and 18, where physical optics, such as the uses of mirrors, prisms, and lenses, are investigated.

In this chapter we will investigate the properties of light with respect to two distinct wave characteristics — **diffraction** and **interference**. If light has wave characteristics then the diffraction and interference of light waves should be observable.

DIFFRACTION OF LIGHT 15.3

Photo 15.2
The interference pattern produced by the diffraction of white light through a narrow slit.

Recall that diffraction is the bending of waves as they pass through an aperture or around the edge of an object in their path. This bending of waves is more noticeable if the wavelength of the waves is comparable to the size of the aperture. Also, if an object is placed in the path of the waves a 'shadow' is produced if the object is of the same size as the wavelength. (Revise Section 14.4.) Can this effect be observed with light? Remember, to observe this effect the slit or the object in front of the waves has to be of the same size as the wavelength of the waves, and light waves have very short wavelengths. However, this effect can be observed! Light does bend around the edges of objects to produce diffraction fringes. Objects seem to be blurred at the edges when light shone on them is focused on a screen. Photo 15.1 shows the diffraction fringes produced by white light passing the edges of a razor blade. The edges of the blade appear blurred and dark bands appear in the small apertures in the blade.

Diffraction of light can also be produced when light passes through a very narrow slit. Photo 15.2 shows the diffraction pattern produced on a screen when white light passes through a very narrow slit.

Photo 15.3
The diffraction pattern produced by a single slit using monochromatic (red) light from a laser.

Activity 15.1 FINGER FRINGES

1. Place your index and middle fingers very close together.
2. Put these fingers up close to one eye, close the other and look at a distant light.
3. Slowly start to separate these fingers and you will notice that black lines appear between your thinly separated two fingers. These are diffraction fringes.

You may have noticed at night in rainy weather how scratches on a car's windscreen produce long shafts of light, sometimes with black bands across them. This too is diffraction.

This pattern can be seen much more clearly if a laser and a commercially prepared narrow slit are used.

If different colours of light are used the pattern changes. If red light is used the pattern spreads out more than when blue light is used. Recall the diffraction of water waves. The larger the wavelength compared with the slit, the more the pattern spreads out and is noticeable. This would suggest that the wavelength of red light is larger than that of blue light. The reason these diffraction bands occur will be analysed in Section 15.5.

15.4 INTERFERENCE

One of the earliest reasons for suggesting that light did not behave like a wave and did not have wave characteristics was that it did not produce interference patterns normally associated with the interaction of waves from two sources.

Can you suggest why? Remember, for a stabilised pattern to be produced the two sources have to continuously generate waves in phase (coherent). Therefore they would have to have the same frequency. Also, for a reasonable pattern with a number of widely separated nodal lines to be produced requires the separation of the two sources to be small compared with the wavelength of the waves. (Refer to Figures 14.18 and 14.19.) Can you describe how an experiment could be designed for the interference of light to be observed?

Thomas Young (1773–1829), a brilliant English academic, made a place for himself in history with his investigations into the nature of light. Young studied medicine at university and later practised in London. He was always interested in sight and made large contributions to the understanding of the eye and eye defects, but his name is remembered in physics for his investigations into how light propagates.

Up to the nineteenth century, Sir Isaac Newton's reputation was enough to uphold the belief within the scientific community in the corpuscular theory of light developed 100 years earlier. Newton had suggested, with explanation, that light travelled as particles (corpuscles). This will be discussed more fully in Chapter 29, Quantum Physics.

Young raised the old debate on whether light travelled as waves or as particles in a paper, *Respecting Sound and Light*, published in 1800. His experimental work with light produced supporting evidence to suggest that light had wave properties. In 1801 Young proved the interference of light. In 1803 he gave a demonstration of the interference of light at a lecture called 'Experiments and Calculations Relative to Physical Optics'. The constructional design of his experiment allowed interference fringes to be produced, something that had not been done previously.

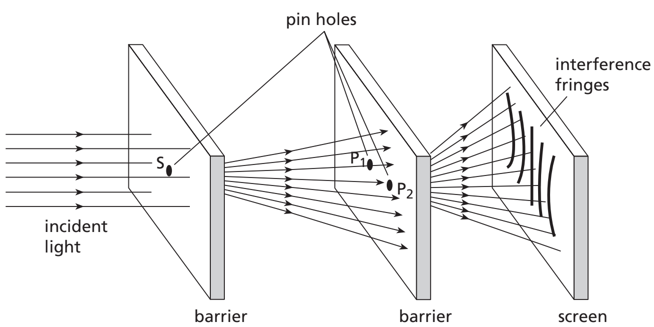


Figure 15.2
The set-up used by Young to produce interference of light waves.

To understand how this was done refer to Figure 15.2. Light from a source was directed onto an opaque sheet with a single pinhole. Light passed through this hole and was incident on two more pinholes, which were very close together, in another barrier. Light from these two pinholes was incident on a screen placed at a long distance from the barrier. Fringes as shown in Figure 15.3 appeared on the screen. These fringes disappeared when one of the pinholes was covered up — this is an indication that the phenomenon was the result of light from the two pinholes interfering.

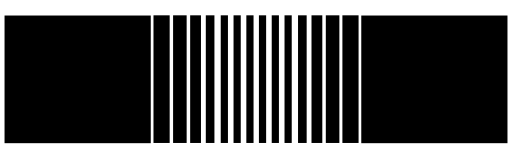


Figure 15.3
The interference pattern produced by light incident on a pair of closely spaced slits.

Young suggested, and demonstrated using Huygens's principle, that each pinhole acted like a point source of light, producing circular waves that radiated outward. The intersecting crests and crests, troughs and troughs, produced coloured fringes, and intersecting crests and troughs produced dark fringes. The interference pattern he drew resembled that of water waves. Again, it was Young's unique constructional design of the experiment that allowed these fringes to appear and remain stable — in the one place. Remember, to obtain a stable pattern the sources of the waves have to be continually in phase, that is, producing crests at the same time. This was the problem that previous exponents of the wave theory of light did not appreciate. Light is produced from the atoms of the light source, and since there are many atoms in the source and they do not produce light waves at the same time, we never have two light waves that are **coherent** (in phase and of the same frequency). Also, interference is most noticeable when the two sources are close together compared with the wavelength of the waves.

EI Activity 15.2 YOUNG'S EXPERIMENTAL DESIGN

Discuss how Young overcame the above difficulties.

By using a barrier with a single pinhole as the source of light, light that arrived at the second barrier with the two pinholes was essentially from the one source. Therefore crests arrived at the two pinholes at the same time. This made the light through these pinhole sources in phase and of the same frequency. The pinholes were also very close together.

The similarity between the pattern drawn by Young and the pattern produced by water waves from two sources added experimental evidence to the suggestion of the wave nature of light.

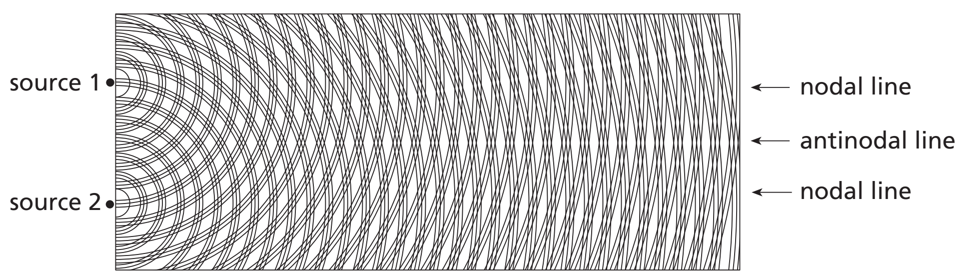
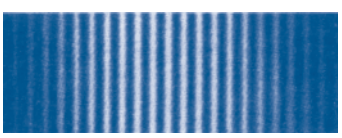


Figure 15.4
The interference of light waves. The intersection of crests and crests, troughs and troughs etc. produces a pattern similar to that of two-source interference of water waves.

Today Young's pattern is a great deal easier to produce in the laboratory. We use a laser, as the light produced by it is **monochromatic** (of one wavelength) and coherent. Using a monochromatic light source produces fringes of the one colour (see Photo 15.4) and the pattern is not complicated by different colours overlapping.

Photo 15.4
The interference pattern produced by monochromatic light (red) through a pair of thinly separated slits.



EI Activity 15.3 INTERFERENCE FRINGES

1. Shine a laser on a screen placed at the front of the room.
2. Place a prepared two-slit slide over the front of the laser making sure that the two slits are in front of the light.
3. Observe and draw the pattern produced on the screen.
4. What happens to the pattern on the screen if slides with different slit separations are used?

Mathematically, the relationship that exists between the positions of the fringes on the screen and the slit separations is similar to that found for water waves (Figure 15.5).

Waves from S are arriving at S_1 and S_2 in phase. These then act as two coherent sources of light.

Note: the distance between the two sources S_1 and S_2 is very small compared with the distance to the screen.

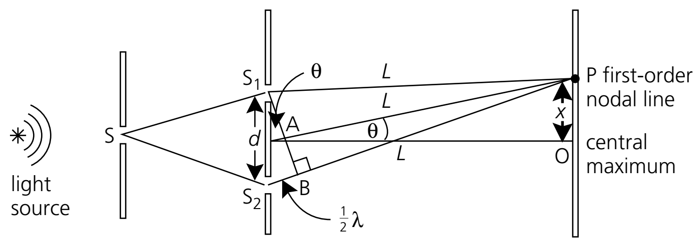


Figure 15.5
A schematic diagram of Young's set-up used to explain the positioning of the first-order node at P (not drawn to scale).

Let's look at a point P, a point on the first nodal line.

As already explained in Chapter 14, the path difference for all points on the first nodal line is $(n - \frac{1}{2})\lambda$. Therefore $S_2P - S_1P = (n - \frac{1}{2})\lambda$, or, since $n = 1$, $S_2P - S_1P = \frac{1}{2}\lambda$.

If we draw a line from S_1 to the line S_2P to meet S_2P at B so that $S_1P = BP$, then $S_2B = \frac{1}{2}\lambda$.

Since d, the distance between the slits, is very much smaller than L, the distance to the screen, the lines from S_1 to point P and S_2 to point P are approximately parallel and therefore S_1B is approximately perpendicular to S_2P.

Activity 15.4 THE INTERFERENCE ASSUMPTION

Just to make it clear in your mind that the above is a reasonable assumption:

1. Place two dots 1 mm apart on one side of your page and draw lines from these dots to another point on the opposite side of your page. You get the idea! These lines are close to being parallel.
2. Now if you could do this again making the dots 0.10 mm apart and draw lines from these dots to a point 3.0 m away what would you find?
3. Can you consider that the rays of light from the Sun are parallel? Explain.

In Figure 15.5 $\triangle$ S_1S_2B is similar to $\triangle$ APO, therefore angle S_2S_1B = angle OAP, which we will call θ.

From triangle S_1S_2B:

$$\sin\theta = \frac{S_2B}{S_1S_2} = \frac{\frac{1}{2}\lambda}{d}$$

From triangle APO:

$$\sin\theta = \frac{x}{L}$$

(*Note:* L is approximately equal to *AP* since *AO* is very large and *x* is very small, as seen in Activity 4.)

NOVEL CHALLENGE

Explain why a pair of car headlights does not produce an interference pattern. Justify your answer mathematically, stating what assumptions you have made about the data.

Therefore:

$$\sin\theta = \frac{\frac{1}{2}\lambda}{d} = \frac{x}{L}$$

If P is a point on the second nodal line, S_2B will be $1\frac{1}{2}\lambda$, then:

$$\sin\theta = \frac{1\frac{1}{2}\lambda}{d} = \frac{x}{L}$$

In general, for a point on the '*n*th' nodal line,

$$\sin\theta = \frac{(n - \frac{1}{2})\lambda}{d} = \frac{x}{L}$$

If P was a point on the first antinodal line then the path difference $S_1B = 1\lambda$. (Recall that the path difference for points on the *n*th antinodal line = $n\lambda$.)
Then, in $\triangle\, S_1S_2B$:

$$\sin\theta = \frac{1\lambda}{d}$$

For a point on the second antinodal line:

$$\sin\theta = \frac{2\lambda}{d}$$

In general, for all points on the '*n*th' antinodal line:

$$\sin\theta = \frac{n\lambda}{d} = \frac{x}{L}$$

where n is the order of the fringe, $n = 1, 2, 3, \ldots$; λ is the wavelength of the light used in metres; x is the distance from the central maximum in metres; L is the distance from the slits to the screen in metres; d is the distance between the slits in metres.

Using these equations Young was able to determine the wavelength for each colour of visible light. Young also used the principle of interference of light waves to explain the coloured pattern produced by reflected light from soap bubbles and thin films. (This will be explained in Section 15.6.)

Example

Monochromatic light from a laser was shone on to a pair of parallel slits 0.20 mm apart. The interference pattern produced was observed on a screen placed at the other end of the laboratory 3.0 m from the laser. It was observed that the first-order bright fringe was 9.0 mm from the central maximum.

(a) Draw a labelled diagram showing the set-up necessary to obtain these results.
(b) Determine the wavelength of the light used.
(c) Determine the distance to the third-order dark fringe.
(d) Determine the thickness of the central maximum.

Solution

(a) See Figure 15.6.

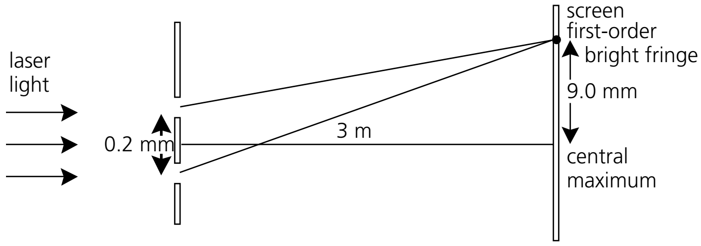


Figure 15.6
The answer to the example problem (previous page).

(b)

$$\sin\theta = \frac{n\lambda}{d} = \frac{x}{L}$$

$$\frac{1\lambda}{d} = \frac{x}{L}$$

$$\frac{1\lambda}{2\times 10^{-4}} = \frac{9.0\times 10^{-3}}{3.0}$$

$$\lambda = 6.0\times 10^{-7}\text{ m}$$

$$\lambda = 600\text{ nm}$$

(c) Destructive interference

$$\sin\theta = \frac{(n-\frac{1}{2})\lambda}{d} = \frac{x}{L}$$

$$\frac{(3-\frac{1}{2})\lambda}{2\times 10^{-4}} = \frac{x}{3.0}$$

$$x = 2.4\times 10^{-2}\text{ m}$$

(d) The thickness of the central maximum is the distance from the first nodal line on one side to the first nodal line on the other. That is, it is two times the distance from the middle of the central maximum to the first nodal line.

$$\sin\theta = \frac{(n-\frac{1}{2})\lambda}{d} = \frac{x}{L}$$

$$\frac{(1-\ )\lambda\times 640\times 10^{-9}}{2\times 10^{-4}} = \frac{x}{3.0}$$

$$x = 510\times 10^{-5}\text{ m}$$

$$= 5.1\text{ mm}$$

Therefore the thickness of the central maximum = 2 × 5.1 = 10.2 mm.

Questions

1 Monochromatic light of wavelength 580 nm is incident on a pair of slits 0.10 mm apart. An interference pattern is observed on a screen 2.8 m from the slits.
(a) What is the distance of the second-order dark fringe from the central maximum?
(b) Determine the distance from the central maximum to the fourth-order bright fringe.
(c) Determine the thickness of the central maximum.

2 Monochromatic light was shone on a pair of slits separated by a distance of 0.15 mm. The third-order dark fringe appeared 2.1 cm from the centre of the central maximum on a screen placed 1.8 m from the source. Determine the wavelength of the light used.

3 Blue light of wavelength 4.4×10^{-7} m was shone on a pair of slits that were separated by a distance of 0.20 mm. The resulting interference pattern was observed on a screen 2.4 m from the source.

(a) Determine the distance from the central maximum of the first-order and second-order bright fringes.

(b) If red light of wavelength 6.6×10^{-7} m was used instead of the blue light where would the first-order and second-order bright fringes now appear?

(c) Draw the fringes from part **(a)** and **(b)** showing their relative positions. (Use red and blue biros.)

(d) What would happen to the pattern obtained if the distance between the slits was doubled?

(e) Express the wavelength of the red and blue light in nanometres.

SINGLE-SLIT DIFFRACTION 15.5

Before Young's interference experiment rekindled the debate on the nature of light, an Italian Jesuit priest, **Francesco Grimaldi** (1618–63), demonstrated diffraction of light. He demonstrated, through a series of experiments in his monastery laboratory, that light can bend around objects. When objects were illuminated with narrow beams of light, bright lines appeared inside the shadow of objects where sharp shadows were expected. This is difficult to observe with normal sunlight as white light is made of a mixture of colours (**polychromatic**). However, it is more easily observed if monochromatic light is used. Grimaldi called this phenomenon 'diffraction' (from the Latin *fragere* = 'to break').

EI Activity 15.5 SINGLE-SLIT DIFFRACTION

Place a single slit in front of a laser and observe the pattern produced on a screen. A pattern similar to the earlier photo in Section 15.3 (p. 321) should be observed.

The pattern produced, even though similar to a double-slit interference pattern, has a number of differences:

- The bright bands are not regularly spaced.
- The central bright band is much wider than the others.
- The central bright band is much brighter than the next one.
- The brightness of each band decreases from the central maximum.

It was not until 1819 when a French engineer **Augustine Fresnel** (1788–1827), using his mathematical abilities, together with Huygens's wavefront principle, Young's interference explanation, and a relatively new mathematics (calculus), explained these observations.

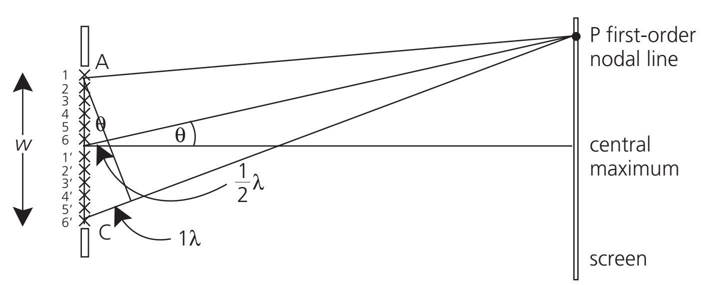


Figure 15.7
For single-slit diffraction the slit is divided into Fresnel zones to help to explain the interference pattern.

Fresnel used these tools to explain, for example, how the first dark fringe was produced (Figure 15.7). He divided the slit into two regions, known as **Fresnel zones**. He then paired up six points in the two zones, 1 and 1′, 2 and 2′ etc. Since the slit was very small, all of these 12 points were on the same wavefront arriving at the slit and therefore in phase. These 12 points then acted as point sources of secondary wavelets (Huygens's principle).

Now let us consider pairs of points within the slit. Point 1 and 1′ can be considered as small point sources of wavelets that are in phase, therefore as the path difference $1'P - 1P$ is $\frac{1}{2}\lambda$ waves arriving at P from these two points destructively interfere, producing a dark spot. Similarly $2'P - 2P = \frac{1}{2}\lambda$, and $3'P - 3P = \frac{1}{2}\lambda$. Light from these pairs of points destructively interferes, producing the first-order dark fringe at P.

Yes, you may say, but there is an infinite number of points between C and A and we could have paired up different points. This is true, which is why Fresnel needed the use of calculus to do a thorough mathematical interpretation of what was going on. However, the above analysis is enough for a reasonable understanding of how a dark fringe appears at point P, when the distance from C to P is 1λ further than from A to P.

Take another example (Figure 15.8) where we have a first-order bright fringe and $DP - AP = 1\frac{1}{2}\lambda$.

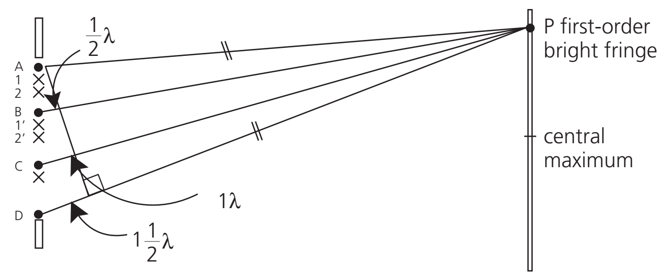


Figure 15.8
For the first-order bright fringe the slit is divided into three Fresnel zones with $\frac{1}{2}\lambda$ path difference between the top and bottom of each zone.

We now divide the slit into three 'Fresnel zones' (A to B, B to C, C to D), the bottom of each zone to point P being $\frac{1}{2}\lambda$ further than the top of the zone to point P. We again pair up six points within the zones.

Notice again that $1'P - 1P = \frac{1}{2}\lambda$ and $2'P - 2P = \frac{1}{2}\lambda$ etc., which makes wavelets produced in these two Fresnel zones, AB and BC, destructively interfere. However, light from the bottom third of the slit does not interfere with other parts of the slit, thus producing light at point P but only one-third as bright as the central region where light from the entire slit strikes.

Again you may say this only occurs because of the selective choice of points, but this is only to aid in the explanation. With a much more thorough mathematical interpretation the results would be the same — a bright fringe occurs at P, which is one-third as intense as the central maximum.

Activity 15.6 FURTHER SINGLE-SLIT INTERFERENCE INVESTIGATIONS

1 It would be a very worthwhile activity to carry out the above procedure to analyse:

(a) the second-order dark fringe (divide the slit into four zones);
(b) the second-order bright fringe.

2 What do you notice about the intensity of the second-order bright fringe?

Using the same assumptions as for double-slit interference we can obtain a mathematical relationship, similar to that of two-slit interference, between the positions of the fringes, the slit width, w, and the wavelength of the light used (Figure 15.9).

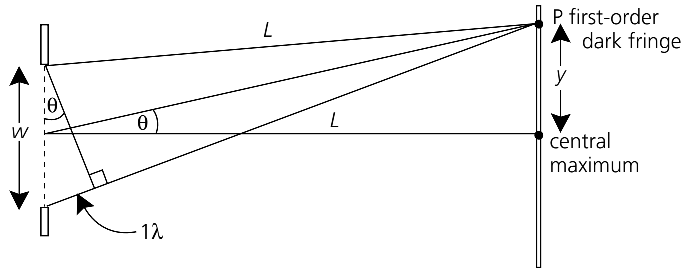


Figure 15.9
The first-order dark fringe is produced when the slit is divided into two Fresnel zones and the path difference between the top and the bottom of the slit is 1λ.

For destructive interference — dark fringes:

$$\sin\theta = \frac{1\lambda}{w} = \frac{y}{L}$$

$$\sin\theta = \frac{2\lambda}{w} = \frac{y}{L}$$

$$\therefore \sin\theta = \frac{n\lambda}{w} = \frac{y}{L} \text{ for the } n\text{th-order dark fringe.}$$

For constructive interference — bright fringes:

$$\sin\theta = \frac{1\frac{1}{2}\lambda}{w} = \frac{y}{L}$$

$$\sin\theta = \frac{2\frac{1}{2}\lambda}{w} = \frac{y}{L}$$

$$\therefore \sin\theta = \frac{(n+\frac{1}{2})\lambda}{w} = \frac{y}{L} \text{ for the } n\text{th-order bright fringe.}$$

Notice the difference between these equations and those obtained for two-slit interference. Two-slit destructive interference occurred when the path difference was an odd number of half wavelengths ($\frac{1}{2}\lambda$, $1\frac{1}{2}\lambda$, . . .) and constructive interference occurred when there was an even number of half wavelengths (1λ, 2λ . . .). *For single slits it is the opposite*.

The intensity pattern is also worthy of note: for two-slit interference the intensity of each bright band was fairly constant, producing an intensity pattern shown in Figure 15.10. The bands were also equally spaced, while for the single slit the central maximum is twice as wide and the intensity falls off as the band number increases. (Refer to Figure 15.11.) As an aid to your understanding, you should make a table setting out conditions for constructive and destructive interference from single and double slits.

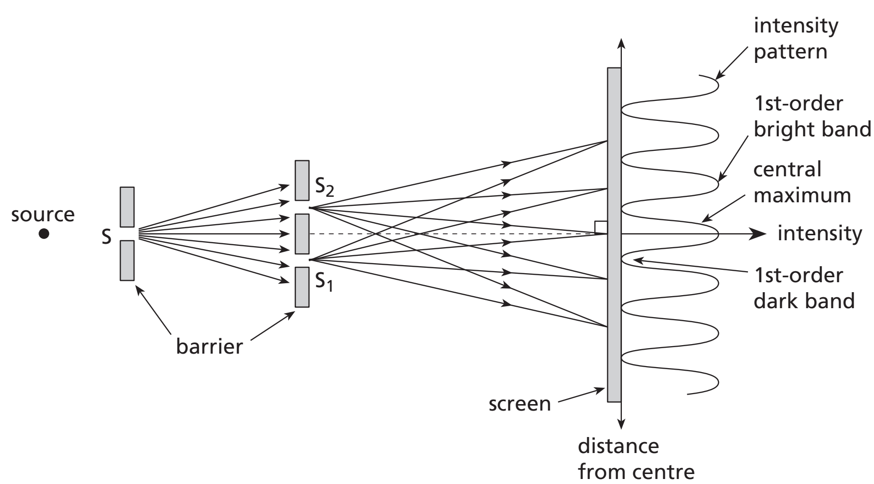


Figure 15.10
The intensity pattern produced by a pair of slits is equally spaced and the maximums are of about equal intensity.

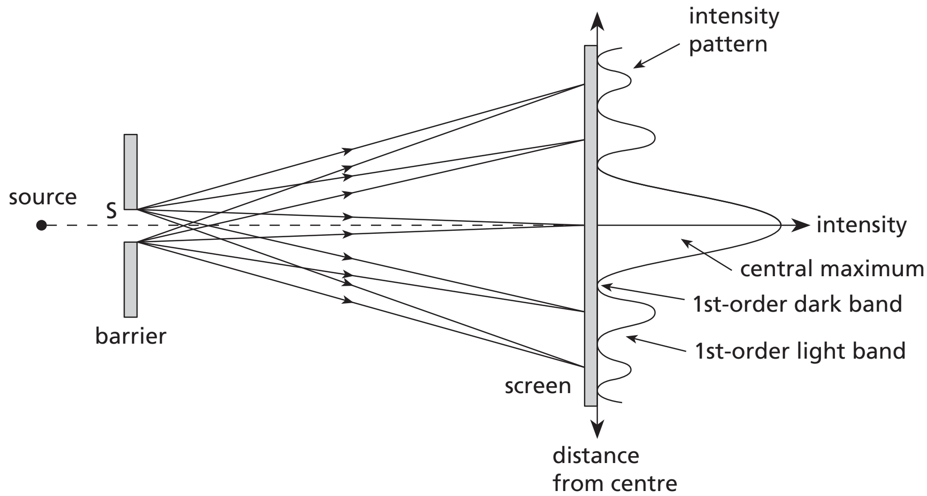


Figure 15.11
The intensity pattern produced by a single slit.

Fresnel's explanation supported the experimental evidence of Grimaldi.

Example

Monochromatic light of wavelength 5.2×10^{-7} m is shone onto a single slit of width 0.10 mm. This produces an interference pattern on a screen 3.0 m from the slit. Find the distance from the centre of the central maximum to the first-order bright fringe.

Solution

$$\sin\theta = \frac{(n + \frac{1}{2})\lambda}{w} = \frac{y}{L}$$

$n = 1$, then

$$\frac{1\frac{1}{2}\lambda}{w} = \frac{y}{L}$$

$$\frac{1\frac{1}{2} \times 5.2 \times 10^{-7}\text{ m}}{0.10 \times 10^{-3}} = \frac{y}{3.0}$$

$$y = 2.3 \times 10^{-2}\text{ m}$$

$$y = 2.3\text{ cm}$$

Questions

4 Find the width of the central maximum when light of wavelength 520 nm is shone on a single slit of width 0.050 mm and the interference pattern is observed on a screen 2.8 m from the slit.

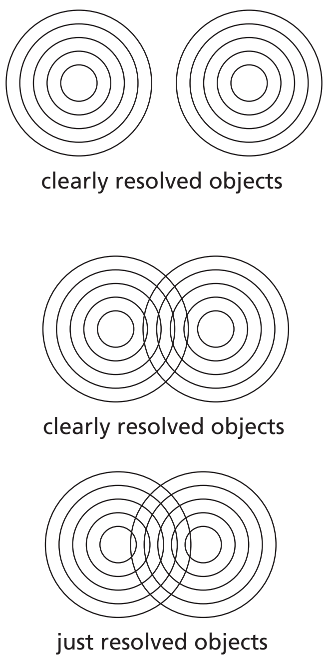
5 Monochromatic light of 585 nm is shone on a slit of width 8.0×10^{-2} mm and the interference pattern is produced on a screen 1.8 m from the source.
- **(a)** Find the distance from the centre of the pattern to the first-order dark fringe.
- **(b)** Find the distance to the second-order dark fringe.
- **(c)** What is the width of the central maximum?
- **(d)** What is the width of the first-order bright fringe?
- **(e)** What do you notice about the width of the central maximum and the other bright fringes?

6 A helium–neon (He–Ne) laser produces light of wavelength 632.8 nm. A single slit of unknown width was placed in front of the laser and the resulting pattern observed on a screen 2.8 m from the slit. The distance from the middle of the central maximum to the first dark band was 8.8 mm. How wide was the slit?

Ch15-Walding 4th 25/8/04 11:54 AM Page 331

Figure 15.12
The diffraction produced by a single circular aperture consists of concentric light and dark bands.

Figure 15.13
The diffraction patterns merge as the objects are moved closer, making resolution poor.



Resolving power of optical instruments

The interference and diffraction effects have consequences for the development of optical instruments. The diffraction pattern produced by light shone through a circular aperture consists of a circular bright centre with circular bright and dark bands, as shown in Figure 15.12.

This diffraction changes with the size of the aperture. If the aperture increases in size the width of the central maximum and the size of the diffraction pattern decrease. The reverse is also true. For light bands:

$$\sin\theta = \frac{(n - \frac{1}{2})\lambda}{w} = \frac{y}{L}$$

If w increases then y decreases.

This has consequences for optical instruments. If a telescope is used to observe light from stars in the sky, the pattern observed through the lens, a circular aperture, will consist of a central light spot with an alternating dark and bright band diffraction pattern. If two stars slightly separated in the sky are observed, overlapping diffraction patterns will be seen. If the sources of light move closer together the diffraction fringes confuse viewing of the sources. Eventually you may not be able to identify the two separate sources as the bright central maxima merge.

The **resolving power** of an instrument is the angular separation of the sources, which enables you to tell you are viewing two sources.
This is given by the following formula, known as the 'Rayleigh Criterion'.

$$\theta = 2.5 \times 10^5 \frac{\lambda}{d}$$

where θ is the angle subtended between the sources measured in seconds of arc; λ is the wavelength of the light used in metres; d is the diameter of the aperture in metres.

Example

What is the resolving power of the Hubble space telescope, which has an aperture of 1.8 m, when viewing two sources of red light of wavelength 650 nm?

Solution

$$\begin{aligned}\theta &= 2.5 \times 10^5 \frac{\lambda}{d} \\ &= 2.5 \times 10^5 \frac{650 \times 10^{-9}}{1.8} \\ &= 0.090 \text{ seconds of arc}\end{aligned}$$

Similar effects are produced when viewing two specimens using microscopes. If blue light is used to illuminate the specimens better resolving power is obtained.

Questions

7 Find the resolving power of a microscope whose objective lens is 0.50 cm in diameter **(a)** when blue light of 450 nm is used to illuminate the two sources; **(b)** when red light of 650 nm is used to illuminate the two sources. **(c)** Which colour of light would be best used to distinguish between two closely spaced specimens?

Diffraction gratings

A diffraction grating consists of a block of glass with many grooves cut into it by a diamond lathe. These grooves are very close together and, once cut, become opaque. Therefore the block of glass acts like many double slits. Typically these gratings have thousands of grooves per centimetre, which makes the distance between the slits very small. For example, if there are 10 000 grooves per centimetre then the distance between each pair of slits will be $\frac{1}{10\,000}$ or 10^{-4} cm.

This results in interference patterns being very spread out when incident on a screen.

Example

If light of 600 nm was shone on the above diffraction grating and the interference pattern was produced on a screen 2.0 m from the grating, what would be the distance from the central maximum to the first maximum?

Solution

$$\sin\theta = \frac{n\lambda}{d} = \frac{x}{L}$$

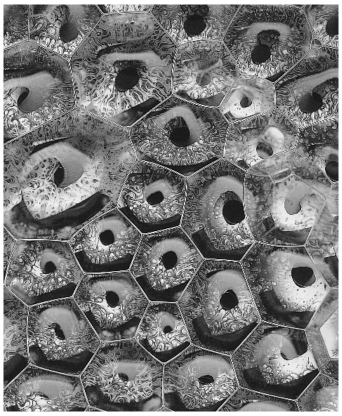
$$\frac{1 \times 600 \times 10^{-9}}{1 \times 10^{-6}} = \frac{x}{2}$$

$$x = 1200 \times 10^{-3}\ \text{m}$$

$$x = 1.2\ \text{m}$$

This is a large distance compared with those distances obtained in previous double-slit problems.

Photo 15.5
The colourful interferences effects produced by white light reflecting from soap bubbles (see also inside back cover).



Questions

8 Yellow light of 590 nm is shone on a diffraction grating that contains 10 000 lines per cm and the pattern is produced on a screen 2.2 m from the grating.
(a) What is the distance from the central maximum to the second-order bright fringe?
(b) What is the distance from the central maximum to the third-order bright fringe?

9 Violet light of 400 nm is shone on a diffraction grating that has 5000 lines per cm.
(a) What is the angular deviation of the second-order bright fringe?
(b) What is the angular deviation of the third-order bright fringe?
(c) What will be the order of the last bright spot that will be seen for this set-up? Explain why this will occur.

15.6 THIN FILMS

Very spectacular colourful interference effects may be observed when a light wave is reflected from a thin film, such as oil floating on water, or a soap bubble, as shown in Photo 15.5 (see also colour section).

EI Activity 15.7 SOAP BUBBLES

Dip a loop of wire into some washing-up detergent. Pull it out and gently blow into the loop to make soap bubbles. Observe the colours of light produced by light reflecting from the bubble.

To explain mathematically how this occurs we will use an example of a thin film of water on a piece of glass. When a light wave strikes the upper surface of the water (see

Figure 15.14
Interference effects occur when light reflects from two surfaces of a film. Rays A and B can constructively or destructively interfere.

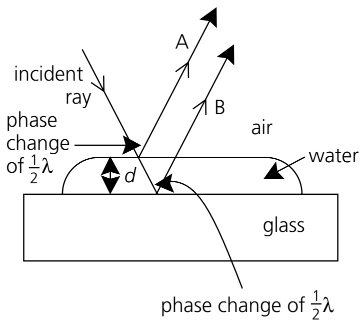


Figure 15.14) it will set up a multitude of reflected waves due to the reflection from the upper surface, the reflection from the lower surface, and multiple zigzags between the surfaces. In Figure 15.14, only the first two reflections are shown. These are the most important and are the strongest as other light will be absorbed.

To find the conditions for constructive and destructive interference between these two waves, let us make a simplifying assumption that the incident wave is nearly perpendicular to the surface. The two most intense waves are those that suffer only one reflection: the wave that reflects from the upper surface (wave A) and the wave that reflects on the lower surface (wave B). Now for a rhetorical question: under what condition will these two waves constructively interfere in the air above the film?

Obviously, the wave that reflects from the bottom surface has to travel further than the wave reflected from the top surface. If the thickness of the film is d, then the **path difference** is equal to $2d$. Provided that this distance is equal to one, two, three, etc. wavelengths, the wave reflected from the upper surface will meet crest to crest with the wave from the lower surface. They will be 'in phase' and will constructively interfere — be bright; or have a maximum amplitude. Note, however, that both waves are striking a 'fixed' boundary, that is, a boundary where the wave strikes a more dense medium than the one it is travelling in. In this case *both* waves will be reflected upside-down and will undergo a half wavelength ($\frac{1}{2}\lambda$) **phase change**. Wave A then is $\frac{1}{2}\lambda$ out of phase with the incident wave and, if the film is $\frac{1}{2}\lambda$ thick, wave B will be one-and-a-half wavelengths ($\frac{3}{2}\lambda$) out of phase with the incident wave — one wavelength due to the path difference and $\frac{1}{2}\lambda$ due to phase change on reflection. Hence, waves A and B will be in phase with each other and constructively interfere. If the film was $\frac{1}{4}\lambda$ thick, the path difference would be $\frac{1}{2}\lambda$ and waves A and B would be $\frac{1}{2}\lambda$ out of phase — B would be 1λ out of phase (or in phase) with the incident wave while A would be $\frac{1}{2}\lambda$ out of phase with the incident wave. Therefore A and B would destructively interfere, producing a dark region.

There are many situations that can be explained using interference of waves. The explanation depends on the optical density (refractive index) of the film and the bounding media. The following are a few examples.

1 *A film bounded on one side by a medium of* lower *optical density and on the other side by a medium of* higher *optical density* Examples are:
- a water film on glass (air–water–glass)
- an oil film on water (air–oil–water)
- a magnesium fluoride (MgF_2) coating on a glass lens (air–MgF_2–glass). (Refer to Figure 15.14.)

Table 15.1 is a summary of the interference effects of various thicknesses of the film.

Table 15.1

	THICKNESS OF FILM (d)			
	$\frac{0}{4}\lambda$	$\frac{1}{4}\lambda$	$\frac{2}{4}\lambda$	$\frac{3}{4}\lambda$
Path difference due to distance travelled ($2d$)	$\frac{0}{2}\lambda$	$\frac{1}{2}\lambda$	$\frac{2}{2}\lambda$	$\frac{3}{2}\lambda$
Phase difference due to reflection	0λ	0λ	0λ	0λ
Sum = total phase difference between A and B	0λ	$\frac{1}{2}\lambda$	$\frac{2}{2}\lambda$	$\frac{3}{2}\lambda$
Constructive (C) or destructive (D) interference	C	D	C	D

Summary: Destructive interference occurs at odd $\frac{1}{4}\lambda$ thicknesses.
Formulae: For *destructive* interference: $2d = (m + \frac{1}{2})\lambda$, where $m = 0, 1, 2, \ldots$
For *constructive* interference: $2d = m\lambda$, where $m = 0, 1, 2, \ldots$

2 *A film bounded on both sides by media of lower density or by media of higher density*
Examples are:

- *Type A* A low–high–low density distribution; for example, a water film or soap film in air (air–soap–air) (Figure 15.15).

Since the waves are reflected from a lower density material at X there will be no phase change, but wave A will still undergo a $\frac{1}{2}\lambda$ phase change at Y.

Table 15.2 is a summary of the resulting interference that occurs for various film thicknesses.

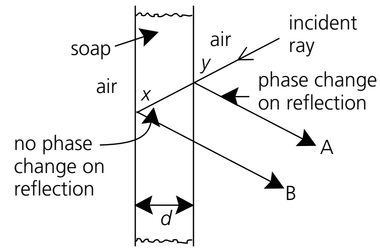


Figure 15.15
The different effects depend on the medium bounding the film. This is a low–high–low optical density situation.

Table 15.2

	THICKNESS OF FILM (d) $\frac{0}{4}\lambda$	$\frac{1}{4}\lambda$	$\frac{2}{4}\lambda$	$\frac{3}{4}\lambda$
Path difference due to distance travelled ($2d$)	$\frac{0}{2}\lambda$	$\frac{1}{2}\lambda$	$\frac{2}{2}\lambda$	$\frac{3}{2}\lambda$
Phase difference due to reflection	$\frac{1}{2}\lambda$	$\frac{1}{2}\lambda$	$\frac{1}{2}\lambda$	$\frac{1}{2}\lambda$
Sum = total phase difference between A and B	$\frac{1}{2}\lambda$	$\frac{2}{2}\lambda$	$\frac{3}{2}\lambda$	$\frac{4}{2}\lambda$
Constructive (C) or destructive (D) interference	D	C	D	C

Summary: Destructive interference occurs at even $\frac{1}{4}\lambda$ thicknesses.
Formulae: For *destructive* interference: $2d = m\lambda$, where $m = 0, 1, 2, \ldots$
For *constructive* interference: $2d = (m + \frac{1}{2})\lambda$, where $m = 0, 1, 2, \ldots$

- *Type B* A high–low–high distribution; for example, an air film between glass slabs (glass–air–glass) (Figure 15.16).

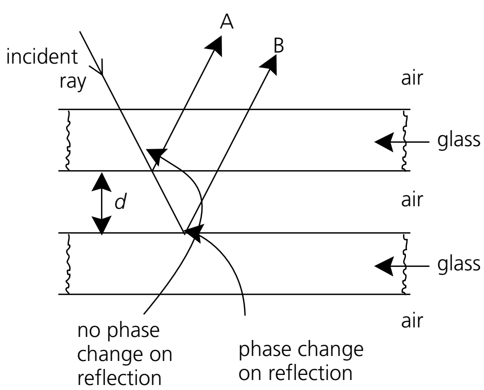


Figure 15.16
This is a high–low–high optical density situation.

Note: as one of the waves undergoes a phase change and the other does not, this situation is similar to that of 2: Type A, and the same table can be used.

A summary of the above situations is given in Table 15.3.

Table 15.3

SITUATION	DESTRUCTIVE INTERFERENCE	CONSTRUCTIVE INTERFERENCE
L–M–H	$2d = (m + \frac{1}{2})\lambda$	$2d = m\lambda$
L–M–L	$2d = m\lambda$	$2d = (m + \frac{1}{2})\lambda$
H–M–H	$2d = m\lambda$	$2d = (m + \frac{1}{2})\lambda$

Colours

The colours seen from oil slicks, Christmas beetles' wings and soap bubbles arise from such interference effects. Since the colours of light have different wavelengths the thickness of a film may be $\frac{1}{2}\lambda$ for one colour but it will not be $\frac{1}{2}\lambda$ for all colours. Therefore constructive interference will occur at different film thicknesses for each colour of light.

You can see that by analysing both the path difference between the two waves and the phase changes on reflection, it can be determined whether the two waves will constructively interfere (producing light) or destructively interfere (producing an absence of light). Different portions of an oil or soap film usually have different thicknesses, and they therefore give constructive interference for different wavelengths at different positions. This results in a pattern of bright coloured bands or fringes.

Note: the symbol λ in the above explanations stands for the wavelength of the light *in the film* and will be different from that in air as light travels more slowly in materials. Remember the frequency does not change. The value of λ in the material can be calculated from the wavelength in air by the formula:

$$\lambda_{\text{film}} = \frac{\lambda_a}{n_{\text{a–film}}}$$

where λ_{film} = the wavelength of the light in the film; λ_a = the wavelength of the light in air; $n_{\text{a–film}}$ or just n_{film} = the absolute refractive index of light going from air to the film. This will be discussed more in Chapter 18.

Example 1

What is the smallest thickness of water film that would produce constructive interference when viewed in reflected light of wavelength of 650 nm in air? ($n_{\text{a–water}}$ = 1.33.)

Solution

The wavelength of the light in water:

$$\lambda_w = \frac{\lambda_a}{n_{\text{a–w}}} = \frac{650}{1.33} = 489 \text{ nm}$$

For air–water–air constructive interference to occur $2d = (m + \frac{1}{2})\lambda$ (see the previous section) then:

$$2d = (0 + \tfrac{1}{2})489 \text{ nm}$$
$$d = 0.5 \times \frac{489}{2}$$
$$d = 122 \text{ nm}$$

Example 2

To reduce the reflection of light from a glass lens (n = 1.50), a coating of magnesium fluoride is added (n_{MgF_2} = 1.36). What minimum thickness should this film be to produce destructive interference and remove as much reflection as possible? Consider the average wavelength of light in air to be 520 nm.

Solution

$$\lambda_{MgF_2} = \frac{\lambda_a}{n_{a\text{-}MgF_2}} = \frac{520}{1.36} = 382 \text{ nm}$$

For this case (low–medium–high) we will have destructive interference at:

$$2d = (m + \tfrac{1}{2})\lambda$$

$$d = \frac{(m + \frac{1}{2})\lambda}{2} = \frac{1}{2} \times \frac{382}{2} = 95.6 \text{ nm}$$

Questions

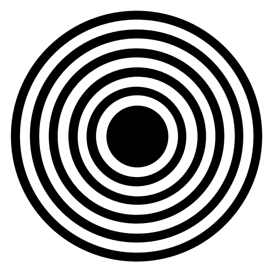
10 A soap film is 100 nm thick. What wavelength of light will be most strongly reflected by this film; that is, what colour will it appear? (n_{soap} = 1.33.)

11 A film of kerosene 4500 angstroms thick floats on water. White light, a mixture of all visible colours, is vertically incident on the film.
(a) Which of the wavelengths contained in the white light will give maximum intensity upon reflection?
(b) Which will give minimum intensity?
(*Note:* 1 angstrom (Å) = 10^{-10} m, n_{kero} = 1.2.)

12 The wall of a soap bubble floating in air has a thickness of 400 nm. If sunlight strikes the wall perpendicularly, what colours in the reflected light will be strongly enhanced as seen in air? The refractive index of the soap film is 1.35.

13 A thin oil slick of refractive index 1.3 floats on water. When a beam of white light strikes this thin film vertically, the only colours enhanced in the reflected beam seen in air are orange–red (650 nm) and violet (430 nm). What is the thickness of the oil slick?

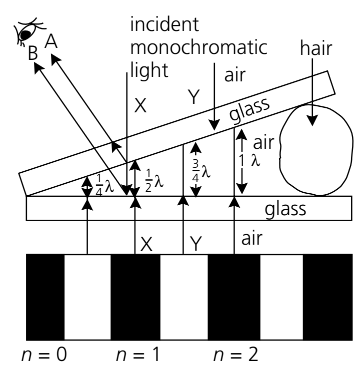
Figure 15.17
Newton's rings are produced by a plano-convex lens resting on a piece of glass.



Newton's rings

Interference effects can also arise in a narrow gap between a flat glass plate and a slightly curved glass plate. The convex surface of the curved glass will touch the plate at the centre but leave a gradually widening gap as the distance from the centre increases. At the bright rings, the width of the gap will produce constructive interference of the reflected light. The dark (destructive) rings are called **Newton's rings**.

Figure 15.18
The varying thickness of the film in the case of a wedge produces light and dark bands corresponding to the different thicknesses.



Wedges

The same result as Newton's rings occurs if two thin glass slides are placed together with a hair between one end. However, this time the bands will be regularly spaced.

Interference can be used to measure the thickness of the hair. At position X where the air gap is $\frac{1}{2}\lambda$ thick we will observe destructive interference between rays A and B (refer to example 2 type B) producing a dark fringe if a particular colour of light is used.

When the thickness is $\frac{3}{4}\lambda$ thick at Y we will observe a light fringe. We would thus observe dark and light fringes regularly spaced between the apex of the slides and the hair.

Example

Two glass slides are separated by a human hair. When viewed from above, dark interference bands are produced as a result of wedge interference. If the 50th dark band occurs above the hair when red light of wavelength 650 nm is reflected from the glass slides, what is the thickness of the hair?

Solution

As this is an example 2 type B problem (refer to thin films, section 15.6), destructive interference occurs when:

$$2d = m\lambda$$
$$2d = 50\lambda$$
$$d = 25 \times 650 \text{ nm}$$
$$d = 1.60 \times 10^4 \text{ nm}$$

NOVEL CHALLENGE

CD players will soon use a blue-green laser of wavelength 400 nm. *How high (or deep) will the pits be?*

Questions

14 The edge of a piece of paper is placed between the ends of two glass slides. The other ends of the slides are in contact. These slides are then illuminated from above with red light of wavelength 700 nm and the interference pattern produced by reflected light is observed. The 200th dark fringe is observed over the edge of the page. What is the thickness of the paper?

15 Students viewing the interference properties of soap bubbles dipped a wire loop into a soap solution and held it vertically for a few seconds. When this loop was illuminated with blue light of wavelength 450 nm and reflected light viewed, dark bands appeared. It was observed that the bottom of the loop contained the 50th blue band. How thick is the soap at the bottom of the loop of wire? Would you expect a regular, evenly spaced pattern of blue and dark bands to appear?

Compact discs

The digital on/off process used in compact discs relies on the destructive interference of reflected laser light. A CD player uses a gallium–arsenide (Ga–As) laser of wavelength 840 nm. The tiny pits in a CD are one-quarter of a wavelength deep so they produce a path difference of $2 \times \frac{1}{4}\lambda$ ($\frac{1}{2}\lambda$) compared with light that is reflected from the upper surface. (See Figure 15.19.) Destructive interference occurs and no light returns to the pick-up photodiode. This is read as 'off'. When light strikes the surface where there is no pit, there is no phase difference between reflected waves, hence the digital pulse is read as 'on'. The subsequent stream of digital on/off pulses is converted to the studio signal and then amplified to become the output sound.

Further discussion of the principles of CDs and DVDs as part of modern sound technology occurs in Chapter 16.10.

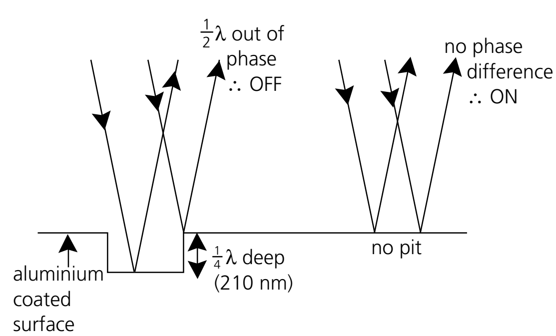


Figure 15.19
CD players and CD discs use interference effects in the replication of sound.

15.7 ELECTROMAGNETIC RADIATION

By the 1880s experimental evidence provided more support for the wave theory of light. In 1849 Armand Fizeau measured the speed of light in media other than air and showed that light travelled more slowly in dense materials. This supported the wave theory of refraction and opposed Newton's corpuscle or particle theory, which suggested that light particles travelled faster in more dense materials. However, the nature of light waves was still not understood.

A Scottish physicist, **James Clerk Maxwell** (1831–79), explained the nature of light waves based on electric and magnetic interactions, which is still the accepted theory today. This explanation was built on Oersted's theories that electric current produced magnetic fields, and Faraday's experiments that showed that changing magnetic fields induced electromotive forces and electric fields. Maxwell suggested that a changing electric field would result in a changing magnetic field, which in turn would produce a changing electric field — one inducing the other in a self-propagating process. He suggested the possibility of transverse **electromagnetic waves** propagating through space as changing electric and magnetic fields that are at right angles to each other, as shown in Figure 15.20. He developed general mathematical equations for these electromagnetic waves. The experimental value for the speed of light was found to be close to that predicted by his equations, suggesting that light was in fact electromagnetic in origin.

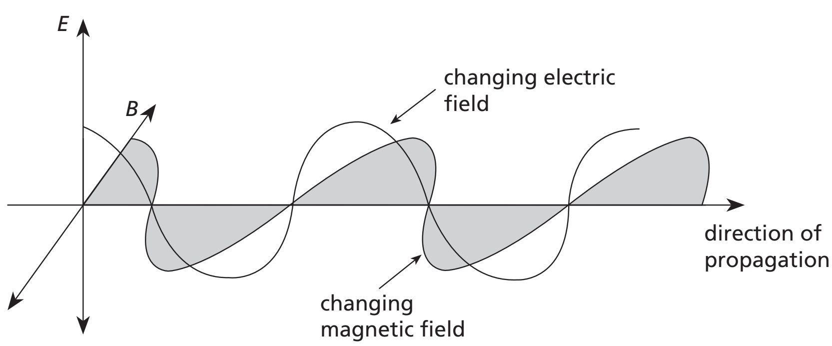


Figure 15.20
A diagrammatic representation of an electromagnetic wave with the changing magnetic and electric fields at right angles to each other and to the direction of propagation.

In general terms Maxwell suggested that electromagnetic waves had the following characteristics:

- They consisted of changing electric and magnetic fields.
- The electric and magnetic fields are at right angles to each other as well as to the direction of propagation — the waves are transverse in nature.
- The speed of the waves is dependent on the electric and magnetic properties of the material in which they are travelling. In air or a vacuum, the speed of light is 3×10^8 m s^{-1}.

In 1887 Heinrich Hertz (1853–94) produced experimental evidence to support Maxwell's explanation of the propagating electromagnetic fields. Using varying voltages, Hertz created a spark across the terminal of a primary loop of wire. The spark or discharge was the result of varying electric fields produced between the ends of the loop. Hertz was able to cause a spark to be produced across the ends of a second detector loop of wire placed some distance from the primary loop (Figure 15.21). This release of energy in the detector loop of wire suggested that energy had been carried by waves from the primary to the secondary loop. Hertz spent a great deal of time investigating the waves produced by the primary loop and discovered that they had all the normal characteristics of waves. After being produced by accelerating and decelerating charged particles the propagating waves travelled at a speed of 3×10^8 m s^{-1} in air and slowed down when they travelled through different media.

Ch15-Walding 4th 25/8/04 11:54 AM Page 339

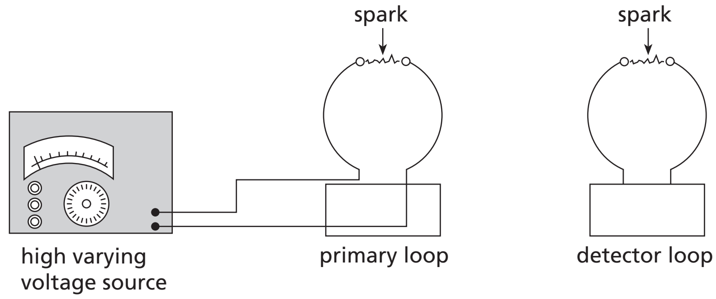


Figure 15.21 A schematic representation of Hertz's apparatus showing that changing voltages in the primary coil produce electromagnetic waves that can be detected in a detector loop of wire.

The frequency of the waves was controlled by the frequency of the changing speeds of the particles at the source. The frequency of generation of the waves in the source produced different wave frequencies, wavelengths, and visible light colours. The wavelengths associated with each colour of light are given in Table 15.4. (See Figure 15.22.) The relationship between the frequency, wavelength and speed of these waves is the same for all waves and is given by the general wave equation $c = f\lambda$.

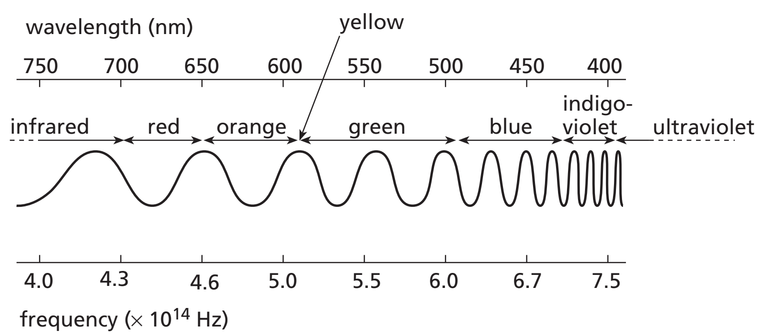


Figure 15.22 The visible light spectrum.

> **✔ TEST YOUR UNDERSTANDING**
>
> Stealth Bombers are invisible to radar because they reflect the waves up and down rather than back to the radar station. *How could you detect them with radar, then? A good answer could be worth millions.*

Table 15.4 THE RANGE OF WAVELENGTHS OF VARIOUS ELECTROMAGNETIC WAVES

COLOUR	WAVELENGTH (nm)
Ultraviolet	200–400
Indigo–violet	400–420
Blue	420–490
Green	490–580
Yellow	580–590
Orange	590–650
Red	650–700
Infrared	≥ 700

However, visible light is only a part of the entire electromagnetic spectrum, which ranges from radio waves of frequencies 15 kHz to gamma rays of frequencies 10^{24} Hz, all of which travel in air or a vacuum at a speed of 3×10^8 m s^{-1}.

THE ELECTROMAGNETIC SPECTRUM 15.8

> **NOVEL CHALLENGE**
>
> You are listening to a radio broadcast of a live orchestral concert in London 20 000 km away. *Would you hear it before or after a person at the rear of the concert hall 50 m away from the orchestra? (Sound travels at about 330 m s^{-1}.)*

Radio and television waves

Radio waves make up one of the biggest groups of waves in the electromagnetic spectrum. They are produced by the oscillations (the acceleration and deceleration) of electrons. The uses of radio waves are widespread: they are used in radio and television broadcasting as well as for communications. AM radio waves have a very long wavelength of several hundred metres and therefore are easily diffracted around buildings. FM radio waves are much shorter (several metres) and diffraction of these is noticeable as buildings are much larger than the wavelengths. Therefore FM reception would be weaker on the side of the building not facing the station. You may have noticed this difference as you drive through town. The good thing about FM radio waves, though, is that they are of much higher energy than AM waves and can penetrate into underground car parks better so you do not lose reception as easily.

UHF (ultra-high frequency, therefore short wavelength) TV wavelengths are only fractions of a metre long and therefore are not diffracted by buildings. TV antennae have to point towards the stations. TV waves, being very short, can also be reflected from objects; in particular, aircraft passing overhead. This could result in waves arriving at your home both directly from the station and reflected from the aircraft. Since this results in a path difference, interference could result in distortion or pulsating images on your TV set when an aircraft passes overhead. Reflection off nearby hills produces a weak delayed signal, which results in 'ghosting'.

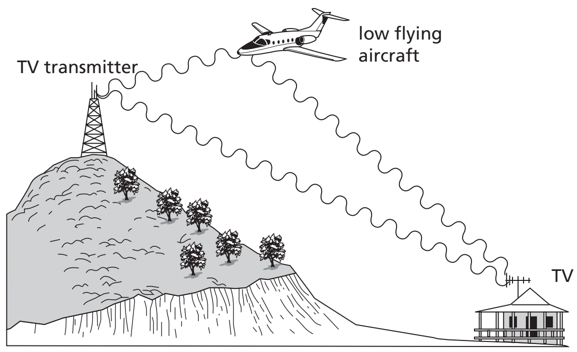


Figure 15.23
The interference of TV waves from the transmitter and those reflected from low flying aircraft can cause distortion on TV sets.

Short wave radio signals of about tens of metres long can be reflected from the ionosphere (a layer in the atmosphere). By bouncing these waves off the ionosphere messages can be sent around the world.

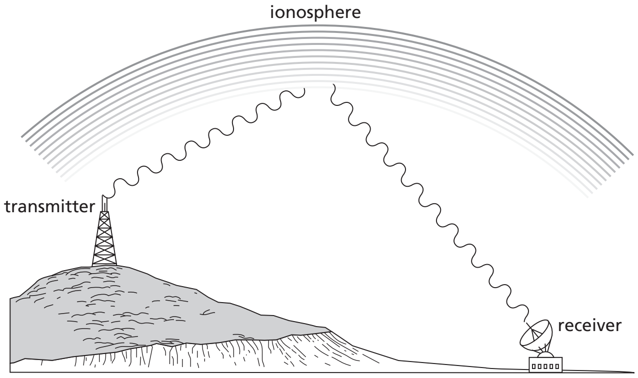


Figure 15.24
Short wave radio transmissions reflect from the ionosphere thus being able to be received around the world.

Table 15.5 gives a summary of the variety of radio waves and their uses.

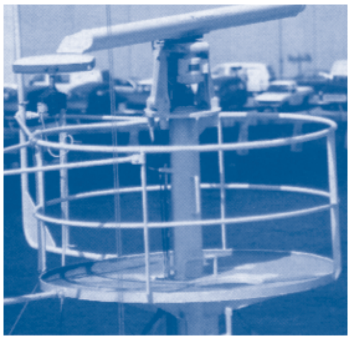
Table 15.5 THE FREQUENCY AND WAVELENGTH OF VARIOUS RADIO WAVES

TYPE OF RADIO WAVE	WAVELENGTH RANGE	FREQUENCY RANGE	USE
Long waves	600 m – 20 km	15 kHz – 500 kHz	communications
Medium waves	100 m – 600 m	500 kHz – 3 MHz	AM radio
Short waves	10 m – 100 m	3 MHz – 30 MHz	AM radio, communications
VHF (very high frequency)	1 m – 10 m	30 MHz – 300 MHz	FM radio
UHF (ultra-high frequency)	0.1 m – 1 m	300 MHz – 3000 MHz	television

INVESTIGATING

You can hear popcorn popping from the outside of a microwave oven. *If the sound can get out, why can't the dangerous microwaves get out?*

Photo 15.6
A microwave radar system.



Microwaves and radar

Microwaves have shorter wavelengths than radio waves. Because of this they can penetrate the ionosphere and can be reflected by smaller objects. They therefore have been used in communication and radar systems. They can be sent to satellites, which retransmit them to ground stations around the world.

Since they can be reflected by small objects they are used in radar systems. **Radar** was a term coined by the British in the mid-1930s as an abbreviation of 'radio direction and ranging'. The possibility of using radio waves (wavelength 10 cm to 10 m) reflected back from aircraft and other metal objects such as ships and submarines attracted much attention in Britain, America, Germany and France in the 1930s. However, it was the pioneering work of British government research physicist Professor Sir Robert Watson-Watt that led to its successful deployment in the defence of Britain's coastline. By March 1936, radar stations were being erected all along the south coast, using 10 m wavelength radio waves to detect German aircraft at distances of up to 100 km. It was found that the best reflections came from objects approximately equal in size to the wavelength.

Although 10 m radar could detect large metal objects such as planes, ships and submarines, it was useless for detecting submerged submarines where only the small *schnorkel* (air intake) was above the surface. The invention in 1940 of the cavity magnetron (now used in microwave ovens) changed all that. It could produce radar with a wavelength of 10 cm, making objects as small as 10 cm visible from as much as 10 km away. In a strange twist of fate, Watson-Watt was caught speeding in a police radar trap on a visit to the USA in 1954.

A radar system consists of a transmitter, an aerial, and a receiver. Pulses are transmitted via the aerial, which rotates on its axis to scan the surroundings for reflected signals, which are heard as an echo. The distance and direction of the objects reflecting the pulses can thus be calculated.

Shorter microwaves of about 0.1 mm produce considerable heat. They do this by causing the particles of matter they penetrate to vibrate faster, resulting in the matter heating up. They are thus used in microwave ovens where they are especially suitable for vibrating water molecules in foods. You may have noticed that when food is cooked in a microwave oven the food itself gets very hot but the plate does not get nearly as hot as in a conventional oven.

NOVEL CHALLENGE

A microwave oven doesn't heat evenly. It's hot in the centre, cooler a bit further out and hot again near the edge. This is due to standing waves being produced, and producing nodes and antinodes. *If the frequency of microwaves in an oven is 2.45 GHz, calculate the wavelength in centimetres and then draw a wave diagram to show this uneven heating.*

Infrared waves

Infrared waves have wavelengths between microwaves and visible light of approximately 0.010 mm. These cannot be detected by the human eye but can be detected by phototransistors and special infrared-sensitive photographic film. All hot bodies emit infrared radiation. Below 500°C, bodies emit only infrared radiation; above 500°C they emit some visible light as well. (Infrared waves and heat effects have been discussed in Chapter 12.) The infrared radiation heating effect is also used in infrared lamps to help overcome injury and to heat and dry objects such as paint on cars during manufacture.

As well as those applications mentioned in Chapter 12 — medical, military and satellite detection applications — infrared waves are also used in alarm systems. As infrared waves cannot be seen, intruders do not notice when they break a beam of the waves and set off the alarm. Self-opening doors of shops work on the same principle.

Ultraviolet waves

Ultraviolet (above violet) waves, or radiation, are those waves with shorter wavelengths than those of visible light — their wavelengths are between 100 nm and 400 nm. They are produced by very hot bodies such as the Sun as well as by electrical discharges through gases. Overdoses of UV radiation can cause sunburn, skin cancer and eye damage; however, luckily for us, most of this radiation from the Sun is absorbed by the ozone layer in the upper atmosphere. (Refer to Chapter 32.)

Ch15-Walding 4th 25/8/04 11:54 AM Page 342

UV radiation carries more energy than visible light and can cause electrons to be ejected from metals when shone on them. Also, when it strikes some substances it causes the substances to emit visible light in a process called **fluorescence**.

Ultraviolet light is also used to detect cracks in materials and to sterilise objects.

X-rays

X-rays are waves with even shorter wavelengths than ultraviolet waves. They have wavelengths of approximately 1.0 nm. They are produced by firing high-speed electrons at a metallic surface. The fast deceleration of these electrons produces X-rays. Because X-rays have great penetrating power through matter and affect photographic film they are used to 'see' through objects. X-ray photographs are a useful tool in medicine. They are also used to detect flaws in metallic structures and welds.

X-rays have high energy and are able to kill living cells. Because malignant cancerous tumours are more susceptible to X-rays than normal cells, controlled doses are used to kill the cancerous cells. Operators have to be careful not to give themselves too high a dose so they wear monitors to register the doses they receive and use lead aprons for protection.

Gamma rays

Gamma rays, of wavelength 0.01 nm, have the shortest wavelength of all forms of electromagnetic radiation. Because they have such a short wavelength they are often not considered as waves. Wave properties of gamma radiation cannot easily be observed or detected. They are the most energetic and penetrating of all forms of electromagnetic radiation and require a thick sheet of lead or a concrete wall to stop them. They are emitted from radioactive nuclei — this will be covered more fully in Chapter 28. They too are used to treat cancers and have numerous other applications.

Table 15.6 summarises the types of electromagnetic waves, their uses, and detection methods.

Table 15.6 TYPES, USES AND METHODS OF DETECTION OF ELECTROMAGNETIC WAVES

	RADIO	TV	MICROWAVES	INFRARED	VISIBLE	ULTRAVIOLET	X-RAYS	λ-RAYS
Wavelength	0.1–10 m	0.1–1 m	10^{-1}–10^{-3} m	10^{-3}–10^{-6} m	2×10^{-7}–7×10^{-7} m	10^{-7}–10^{-9} m	10^{-9}–10^{-12} m	10^{-11} m
Use	broadcasts	TV broadcasts	ovens, radar	scanning, drying paint, treating injuries, alarm systems	sight	flaw detection, sterilising objects	medicine, flaw detection	medicine, power stations
Source	radio transmitters	TV transmitters		warm and hot objects	hot objects, fluorescent substances	very hot objects	X-ray tubes	radioactive substances
Detector	aerial with radio	aerial with TV set		skin, thermometer, thermistor	eyes, photographic film, LDR	photographic film, skin	photographic film	Geiger–Muller tube

Activity 15.8 RADIO PHYSICS

Predict and justify the outcome of the following, and then try them. **(a)** Tune a small transistor radio in to a station and squat down with it in your lap (will the radio still pick up the broadcast?) **(b)** Instead, just wrap the radio in alfoil. **(c)** Tune a radio to a distant station at night and then turn it on in the morning without changing stations. If any of your predictions were wrong, explain the result.

Ch15-Walding 4th 25/8/04 11:54 AM Page 343

Activity 15.9 MICROWAVE COOKERY FOR PHYSICISTS

Ask your teacher to put the following into a microwave oven and to turn on 'high' for a few seconds (they are all pretty safe): **(a)** a small fluorescent bulb; **(b)** a neon pilot light; **(c)** two alfoil squares just touching at their corners; **(d)** a 100 W incandescent bulb; **(e)** some large squares of moist cobalt chloride paper (pink when wet, blue when dry); **(f)** a grape that you have prepared beforehand — cut it almost in half and peel it backwards so that each half is connected by a thin skin bridge. This is really spectacular. Don't do it at home; your mum won't be pleased.

POLARISATION 15.9

INVESTIGATING

Have you ever wondered what happens when you shine a laser through Polaroid? Is a laser polarised and so none gets through? If we had time we'd try it.

Electromagnetic waves, as suggested by Maxwell, are a type of transverse wave composed of oscillating electric and magnetic fields. These transverse waves have components of electric and magnetic fields in all directions. When all components of the electric fields except for one are blocked the wave is said to be **plane polarised**. It is usually the electric field vector that defines the direction of polarisation.

A device that allows only one component of the electric field through is called a **polariser**. An example may explain this more simply; if a slinky spring is threaded through a slit in a wall as shown in Figure 15.25 and shaken in all directions, only those pulses that are in the same plane as the slit will get through; the rest will be blocked.

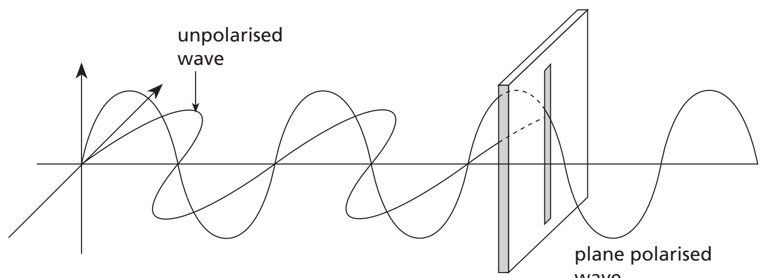


Figure 15.25
Polarisers allow only waves in the plane of the polariser to pass through.

'Polaroid' is the brand name of synthetic materials that have these polarising properties. They polarise light or allow light waves vibrating in one direction through. If a second piece of polaroid called an **analyser** is placed after the polariser, as shown in Figure 15.26, it is possible to block out all light from the source by rotating the analyser. If the plane of the analyser and the polariser line up then light will be seen. However, if they are at right angles no light will be transmitted through the analyser. Between these two extremes the amount of light transmitted will vary depending on the angle between the planes of the polariser and the analyser.

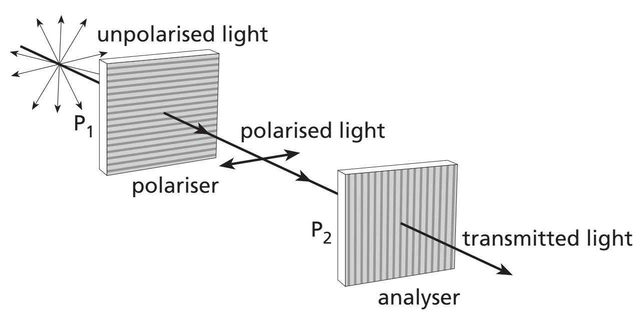


Figure 15.26
The angles between the polariser and the analyser determine the amount of light that is transmitted.

EI Activity 15.10 SUNGLASSES

If you have an old pair of Polaroid sunglasses, pop the lenses out and place them together. When they are rotated as shown in Figure 15.27 the amount of light passing through will vary.

1. Cut one of the lenses in half — now you have three pieces.
2. Cross one pair so that they go black (Figure 15.27 **(a)**).
3. Slide the third one in at an angle and note what happens (Figure 15.27 **(b)**). How on earth can this happen?
4. Try it in front and behind the crossed polarisers. What happens?

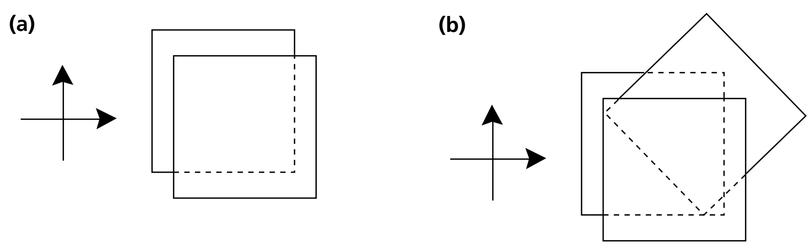


Figure 15.27
For Activity 15.10.

INVESTIGATING

The nematic crystals used in liquid crystal displays (e.g. your calculator) melt like all crystals. Put your calculator in the sun and watch the display go black. *At what temperature did this happen?*

Polaroid sunglasses

When wearing Polaroid sunglasses, annoying reflections from horizontal surfaces, such as shiny floors, wet roads, car bonnets, the ocean and the beach are eliminated. When sunlight reflects from these surfaces it becomes horizontally polarised because most of the other components are scattered. Polaroid sunglasses have their polarising plane vertical so as to block these reflections.

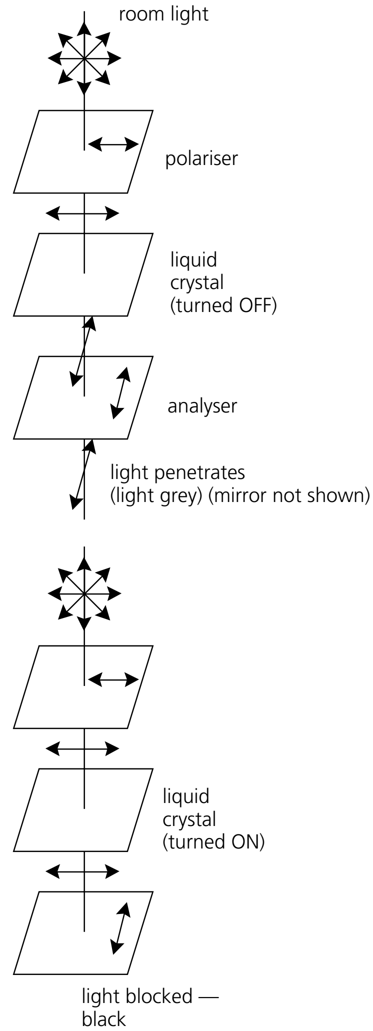
Camera filters

To reduce the brightness of light entering a camera, photographers sometimes use Polaroid filters which have polarisers that can be rotated. Light intensity can be reduced by rotating one of the polarisers. By doing this instead of closing the aperture down, the depth of field of the lens is not affected. The polarising filter is also used to reduce unwanted reflections from glass or water surfaces.

Liquid crystal display

The sort of display used in calculators and digital watches uses two pieces of Polaroid that are crossed. Room light passes through the top polariser, where it is then rotated through 90° by the liquid crystals before it strikes the bottom polariser. The bottom polariser is crossed with respect to the top polariser but because the liquid crystals have rotated the light through 90° the light passes through. Underneath the bottom layer is a mirror which reflects the room light back to the user. (See Figure 15.28.) When a voltage is applied to the crystals they stop rotating the light so it is blocked, and appears black. If you have a broken calculator check that it has Polaroid and mirrors in it.

Figure 15.28
Liquid crystal displays in calculators and digital watches use polarising filters. A voltage applied to the liquid controls the amount of light reflected back to the user.



Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***16** State two properties of waves that are considered to be strictly wave characteristics.

***17** State the important characteristics of Young's experimental design that allowed the interference fringes to be observed.

****18** Light from a red laser of wavelength 620 nm is shone on a pair of parallel slits 0.10 mm apart. The interference pattern produced is incident on a screen 2.8 m from the slits.

- **(a)** Calculate the distance from the central maximum to the first-order nodal line in this pattern.
- **(b)** Calculate the distance to the third-order antinodal line.
- **(c)** What would happen to the pattern if the distance between the slits was reduced?
- **(d)** What would happen to the pattern if the slits and laser were moved toward the screen?
- **(e)** What would happen to the pattern if yellow light was used instead of red light?

****19** If Young's experiment was used to produce an interference pattern with X-rays (wavelength of 1.0 nm), what slit separation would be needed to make the second antinode 2.0 mm from the central maximum on a screen placed 2.0 m from the slits? (Is this feasible?)

***20** Green light of wavelength 510 nm is shone on a pair of slits placed 2.0 m from a screen. The distance between the slits is 0.20 mm.

- **(a)** What is the distance from the central maximum to the third antinodal line?
- **(b)** What is the thickness of the central maximum?

***21** Students use a single slit to determine the wavelength of light produced by a laser. Light from the laser is incident on a slit of 1.0 mm width, and the interference pattern is observed on a screen 3.0 m from the laser. The first-order dark band appears 2.0 mm from the middle of the central maximum. What is the wavelength of light emitted by the laser?

***22** S_1 and S_2 are sources of VHF radio waves of wavelength 0.50 m. They are connected to the same generator and remain in phase.

- **(a)** If a detector of this radiation is moved from P to Q as shown in Figure 15.29 how far from P will the first minimum be detected?
- **(b)** Why does the detector register a maximum of intensity when it is 1.5 m from P?

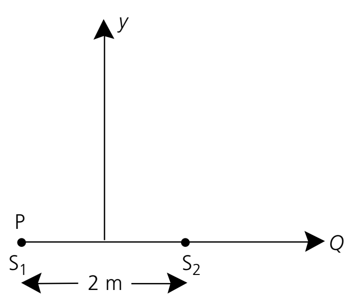


Figure 15.29
For question 22.

***23** Blue light of wavelength 450 nm is incident on a pinhole of 2.0×10^{-4} m diameter made in an opaque sheet and the resultant interference pattern is produced on a screen 2.5 m from the pinhole.

- **(a)** Describe the pattern produced on the screen.
- **(b)** What is the distance from the central maximum to the third bright line?
- **(c)** What is the diameter of the central maximum?

****24** A water film (n = 1.33) in air is 320 nm thick. It is illuminated by white light at normal incidence. What colour of light occurs in the reflected interference pattern produced above the film?

***25** A thin film of refractive index 1.5, and 4.0×10^{-5} cm thick, is surrounded by air and illuminated by white light normal to its surface. What wavelengths within the visible spectrum will be intensified in the reflected beam?

***26** Light of wavelength 680 nm in air illuminates at right angles two glass plates 12 cm long that touch at one end and are separated at the other end by a wire of 0.048 mm diameter. How many bright fringes will appear over the 12 cm length?

***27** **(a)** What is the speed in air of **(i)** red light of wavelength 620 nm; **(ii)** blue light of wavelength 470 nm; **(iii)** X-rays of wavelength 1.0 nm?
(b) Find the frequency of each of the above electromagnetic waves.
(c) What would be the frequency of these rays in water?

***28** List several pieces of evidence that support the theory that light is a wave.

***29** A radio station transmits at a frequency of 105 MHz. Calculate the wavelength of these radio waves.

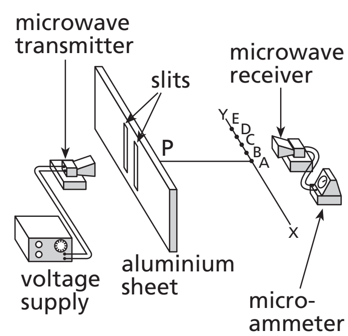
***30** A ship using a microwave radar system to detect distant aircraft receives 'echoes' back 5.0×10^{-4} s after transmission of a radar pulse. Determine the distance to the aircraft.

***31** Weather satellites use infra-red detectors rather than visible light detectors. Explain the advantages of this.

****32** Microwaves can be used in the laboratory to show the interference of waves. The microwaves are transmitted by a microwave transmitter and received by a detector.

The microwave transmitter transmits on a frequency of 2.0×10^{10} Hz. It is found that the maximum intensities occur at points A, C, and E, while minimum intensities occur at points D and B. The distance from A to D is found to be 1.5 m while the distance from P to A is 3 m. (Refer to Figure 15.30.)
(a) Calculate the wavelength of the microwaves used.
(b) What is the path difference from the two slits to point **(i)** D; **(ii)** C?
(c) Calculate the separation between the slits produced by the aluminium plates to produce this pattern.

Figure 15.30
For question 32.



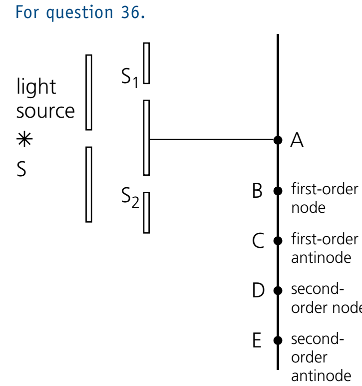
***33** Which category of wave is used for the following:
(a) reflecting from the ionosphere;
(b) communicating by retransmitting from satellites;
(c) producing heat for drying paint on cars;
(d) treating cancers;
(e) seeing objects;
(f) detecting cracks in welds;
(g) 'seeing' internal structures;
(h) operating radar systems;
(i) causing skin cancers?

***34** A radio station is transmitting carrier waves of 5.0 m wavelength. Is this likely to be an FM or an AM station?

***35** The categories of waves in the electromagnetic spectrum overlap. There is no distinct division between, say, X-rays and gamma rays. However, they are produced differently. Explain how they differ in their production.

****36** Figure 15.31 is a schematic diagram of Young's double-slit experiment.
(a) What are the distances **(i)** $S_1A–S_2A$; **(ii)** $S_1B–S_2B$; **(iii)** $S_1E–S_2E$?
(b) What kind of interference occurs at **(i)** D; **(ii)** A?
(c) If the distance from A to D is 4.0 mm, the distance to the screen is 90 cm and the distance between the slits S_1, S_2 is 0.15 mm, find the wavelength of the light used.

Figure 15.31
For question 36.



***37** Choose two AM and two FM radio stations. On what frequencies do they transmit? Calculate the wavelength of these radio waves.

***38** The new generation of compact discs can fit over two hours of music and video onto one disc but they need a much higher frequency laser to read the CD pits. Why do you think this is?

Extension — complex, challenging and novel

***39 A parallel beam of light containing red light of wavelength 650 nm and orange light of wavelength 580 nm is shone on a diffraction grating containing 1000 lines per centimetre. Calculate the angular deviation of these two wavelengths in the second-order bright fringe.

***40 White light reflected at normal incidence from a soap film has, in the visible spectrum, an interference maximum that occurs for light of 600 nm wavelength and a minimum that occurs for light of 450 nm wavelength, with no minimums between these wavelengths. If n = 1.33, what is the film thickness, assumed uniform?

***41 Monochromatic light falls normally on a thin film of oil that covers a glass plate. The wavelength of the source can be varied continuously. Complete destructive interference of the reflected light is observed for wavelengths of 500 and 700 nm and for no other wavelengths in between. If the index of refraction of the oil is 1.3 and that of glass is 1.5, find the thickness of the oil film.

***42 An oil tanker accidentally discharges oil onto the Great Barrier Reef. If the thickness of the oil film produced on the water was 5.0×10^{-7} m what colour would the film appear? (n_{oil} = 1.45.)

***43 An air wedge is made by placing a thin sheet of paper between the ends of a pair of glass slides. Red light of 650 nm is shone at right angles onto the surface of the slides. The bright bands in the interference pattern of reflected light are observed to be 0.40 mm apart. If the paper is 5.0 cm from the point of contact of the slides, what is the thickness of the paper?

***44 One method of measuring the thickness of a razor blade is by using double-slit interference. Design an experiment that would enable you to do this. (*Hint:* use two blades.)

**45 Answer true or false:

(a) Light is a mixture of particles and waves.
(b) Light waves and radio waves are not the same thing.
(c) Microwaves have an extremely short wavelength.
(d) The addition of all colours produces black.
(e) Light exists in the crest of a wave and darkness in the trough.
(f) Rays and wavefronts are the same thing.
(g) Photons are just neutral electrons.

CHAPTER 16

Sound, Music and Audio Technology

16.1 INTRODUCTION

What is sound? One frequently-played TV advertisement for a popular drink showed people whistling and causing glasses and bottles to shatter. A measure of greatness of singers is that they have the ability to shatter crystal glasses by reaching a certain high-pitched note. The renowned opera singer Maria Callas (1923–77) was reputed to be able to do this.

- Why does this happen? Does it happen or is it one of those exaggerated myths?

After completing this chapter you will be able to answer this question and others such as these:

- Can you hear space ships explode in space?
- How is sound used by dentists, doctors, bats, the blind, and fishermen?
- Why do you hear a siren differently as a police car comes toward you and goes away? How does an understanding of this allow you to measure the speed of a cricket ball?
- Why can you hear so well in the 'bush' at night?
- Do you know how insects can hear bats? Can any animal 'jam' a bat's sonar?

16.2 WHAT IS SOUND AND HOW IS IT PRODUCED?

In Chapter 13 we suggested that if a tuning fork is tapped and held beside another one of the same frequency, the second fork also begins to vibrate. Sound waves also cause our ear drums to vibrate and microphones to produce small alternating voltages.

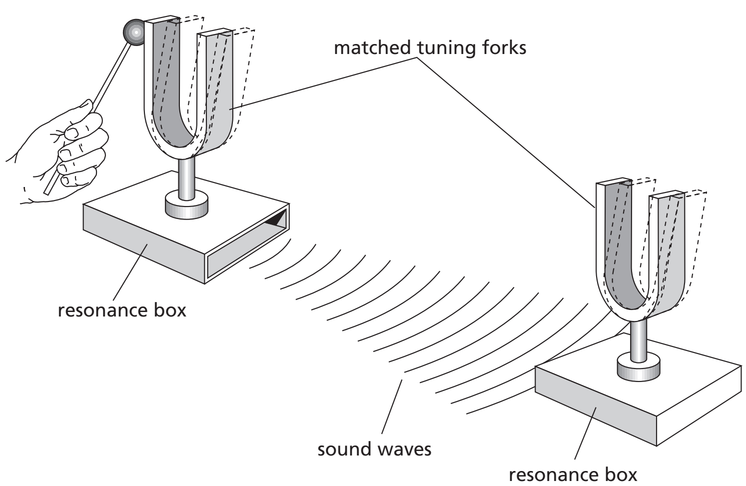


Figure 16.1
One vibrating tuning fork will cause another close by to vibrate.

Sound is a form of energy — one that travels from the source to the receiver by means of waves. In Chapter 13 it was indicated that sound waves were longitudinal mechanical waves — waves that require a medium for transmission.

Activity 16.1 SOUND ENERGY

Research an area that supports the proposition that sound is a form of energy, and be prepared to explain your evidence to the class.

Activity 16.2 SOUND IN A VACUUM

If your school possesses a vacuum pump, the following demonstration should indicate that sound requires a medium for its propagation.

1. Place a ringing electric bell in a bell jar and extract the air from the jar using the vacuum pump.
2. What do you notice as the amount of air in the jar is reduced?
3. Predict what you would notice if the air pressure was increased.

Activity 16.3 SOUND VIBRATIONS

Part A

1. Place a metal ruler on a desk with about three-quarters of its length overhanging the edge of the desk. Pull the overhanging end down and let it vibrate.
2. What happens? What do you hear?
3. Place your ear on the desk and you will hear the sound with astonishing clarity.
4. What do you notice when you rest the bone behind your ear on the desk?
5. As the ruler is vibrating move it in so less overhangs. What do you hear now?
6. What happens if you use a softer surface?

Part B

It is not easy to see a tuning fork vibrating but it can be felt. Touch the stem to your lips, your teeth, your head. Touch the prongs on to the surface of water. What do you notice?

Photo 16.1
Siren disc.

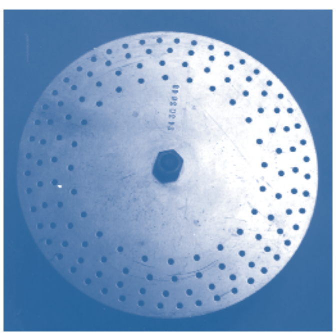
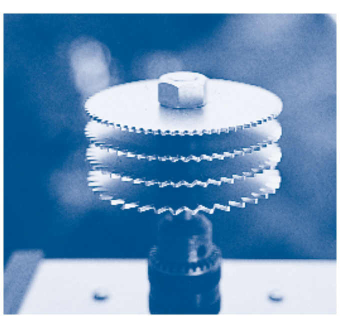


Photo 16.2
Savart's disc.



It can easily be seen that the vibrations cause the sound. Music students in the class may be able to tell you why guitars, flutes or trombones produce sound. Sound is produced by something vibrating — a string, a reed or an air column. The different sounds are produced by the different frequencies of vibrations. This can be shown by tapping two tuning forks of different frequencies. They produce different sounds due to the different rates of vibration of the arms of the forks.

We speak and hear due to the vibration of membranes. The vocal cords in our throats vibrate at different rates to produce sounds. The tension of the vocal cords controls the rate of vibrations. The energy carried by the sound waves causes the ear drum to vibrate. These vibrations are transposed into discernible noises by the brain. When you get a throat infection your vocal cords become inflamed and swollen. This gives you a husky voice.

Sound is a form of energy produced by the vibrations of objects and carried by longitudinal mechanical waves.

A siren disc (see Photo 16.1) has concentric circles of holes in an aluminium disk. When it is spun at high speed and air blown through the holes, the chopped jet of air produces a sound wave.

Savart's toothed wheel can also be used to produce 'musical' sounds. By holding a piece of card against the spinning toothed wheel, different frequencies can be produced. The number of teeth per wheel are in a set ratio and, when sounded at the right speed, produce what music students call the notes of the major triad: C_4, E_4, G_4, and C_5.

Activity 16.4 WAVE MOTION

1. Mount a tiny mirror on one prong of a tuning fork held upright in a stand. Shine a laser beam onto the mirror so that it reflects onto a wall. Set the tuning fork vibrating and slowly rotate it back and forth.
2. What pattern appears on the wall?

16.3 PROPAGATION OF SOUND WAVES

Recall from Chapter 13 that sound waves are longitudinal waves producing compressions and rarefactions of the air particles in the direction the wave propagates. Therefore without air (in a vacuum) no sound can propagate. So do not believe those space movies in which spaceships explode with a large noise.

- Would you be able to hear the explosion?
- Would you be able to see it?

Then how does a Space Shuttle communicate with Earth? This is done by means of radio waves — a form of electromagnetic waves that do not require a medium for their propagation. Therefore sound can propagate in all media that have particles — air, water, wood and the ground. You may have seen Indian scouts in old Westerns fall to their knees and press their ears to the ground to detect distant and unseen riders. They relied on the fact that sound travels through the ground very well and doesn't get scattered as it does in the air.

NOVEL CHALLENGE

If you put your head underwater while having a bath, you can hear sounds from all over the house that you wouldn't normally hear. Why is that?

Speed of sound

The more rigid the particles in a medium, the faster the sound will travel through it. This is shown in Table 16.1.

Table 16.1 THE SPEED OF SOUND IN VARIOUS MEDIUMS

MEDIUM	SPEED OF SOUND (m s^{-1})
Air	342
Water	1410
Copper	3560
Aluminium	5100

In air, where the particles are loosely connected, the speed of sound is approximately 340 m s^{-1}. However, this varies with the atmospheric conditions such as temperature, humidity, and air movement. As air temperature rises, the speed of sound increases by about 0.6 m s^{-1} for each degree. At 0°C sound travels at 331 m s^{-1}, whereas at 20°C it travels at $331 + 20 \times 0.6 = 343$ m s^{-1}.

As an orchestra warms up, the pitch of wind instruments becomes higher because the speed of sound in air increases and this affects the frequency of the standing waves inside the instruments. String instruments, on the other hand, go lower in pitch because the friction of the fingers rubbing over the strings heats them up and causes them to lengthen.

In other gases, the speed is different, as Table 16.1 shows. You may have seen people take a lung-full of helium gas from party balloons and when they speak they sound like Donald Duck. The speed of sound in helium is about 965 m s^{-1}, which causes the resonant frequency of the throat and mouth cavity to rise.

NOVEL CHALLENGE

In the *Song of the White Horse* by David Belford, the lead soprano is required to breathe in helium to reach the extremely high top note.
Question: if you released some of the helium into the middle of the orchestra, what would happen to the pitch of the following instruments: wind, brass, strings, percussion?

Range of hearing

The frequency and therefore wavelength of sound waves is controlled by the frequency of the vibrating source. The human ear can detect frequencies from 20 Hz to 20 000 Hz; however, the ear does not respond in the same way to all frequencies. The ear is most sensitive to frequencies of about 3000 Hz (Figure 16.2). Age also affects our hearing. As people age, they find high frequencies more difficult to hear. At about 65 years old, the highest frequency heard is about 5000 Hz. Human speech produces frequencies from 600 Hz to 4800 Hz.

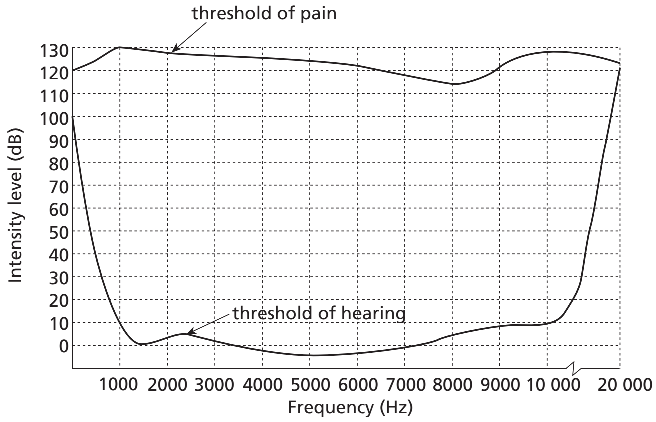


Figure 16.2
The graph shows that the human ear does not respond equally to all frequencies. Some frequencies need to be more intense to be heard.

NOVEL CHALLENGE

The air bladder of a fish serves another purpose besides buoyancy: it enhances the fish's hearing. How does it do that?

Infrasound

Our ears are not the only detectors of sound. Other body organs can detect sound waves, especially vibrations of low frequencies. For example, the intestines and stomach are susceptible to low vibrations, with a maximum response at about 3 Hz. Your stomach wall actually moves in and out in response to such vibrations (Figure 16.3). This property is used in cinemas to produce an effect called 'sensurround' in which low frequencies are generated by banks of large 'woofer' speakers to make the effects of bomb blasts and earthquakes more realistic as you feel the effect as well as hear it. Physicists call such low frequencies **infrasound** (Latin *infra* = 'below') and while it may be safe under controlled conditions, infrasound can also cause nausea and dizziness, such as in car sickness. Death can occur in extreme cases when internal organs rub against each other and haemorrhage (rupture). Enjoy your movie.

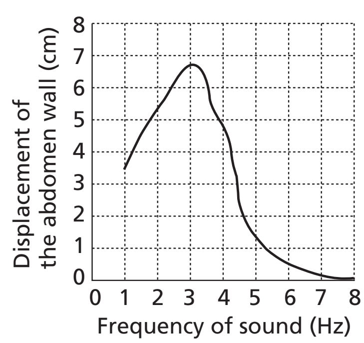


Figure 16.3
The abdomen responds to low frequencies. The abdominal wall can move an incredible 6.5 cm in response to frequencies of about 3 Hz.

NOVEL CHALLENGE

The frequencies of insects' wing beats has been measured: butterfly 12 Hz, bumblebee 130 Hz, honeybee 225 Hz, mosquito 600 Hz. Which can you hear, and why? Spiders know when an insect has been caught in their web by the vibrations emitted. Spiders will run to a 256 Hz tuning fork held in their web, but not if you just stick your finger in it. Don't you want to have a go?

Questions

1. Sound travels through media containing particles. Explain how this occurs.
2. People can not only hear sound travelling in materials, they can also feel the sound vibrations. Apart from the above situations can you think where this is used?
3. Describe what is meant by the terms 'compression', and 'rarefaction'.
4. Use Figure 16.3 to determine what range of frequencies is most suitable for 'sensurround' speakers to produce.

16.4 PROPERTIES OF SOUND WAVES

As sound is a wave, those properties common to all waves apply to sound. Sound waves can be reflected, refracted, diffracted and can interfere. The wave equation $\boldsymbol{v} = f\lambda$ that you have used for water waves and light waves also applies to sound waves.

The following are a few examples of where properties of sound waves are experienced every day. You can probably think of many more.

Reflection

When you hear an echo you are hearing the reflected sound from a distant mountain, cliff or wall. This is very noticeable when you shout in an empty room as there are no materials to absorb the wave energy. Careful design and placement of curtains and furniture is essential in concert halls to absorb sound and stop reflected waves interfering with sounds produced by the artists. Theatre design and construction utilises computer simulations to show where reflections will interfere and where absorbing materials are essential. People's bodies also assist in absorbing sounds. Modern theatre acoustics now include designs that allow unoccupied seats to retain the same sound-absorbing qualities as human beings, so that the sound reinforcement is similar whether the theatre is full or empty.

Reflection of sound can be used to measure the speed of sound.

EI Activity 16.5 THE SPEED OF SOUND BY REFLECTION

Clap your hands hard (or hit a piece of wood with a mallet) at a distance from a wall in an open area. When you hear the echo clap your hands again. Continue to do this until 10 echoes are heard. The speed of sound can be calculated by dividing the total distance the sound covered in the 10 'trips' by the time taken from the first clap till the last echo.

1. How closely did your measurement match the stated value for the speed of sound?
2. Discuss the sources of error in your measurement, and how the experiment could be improved.

NOVEL CHALLENGE

Thunder from lightning refracts upwards, and at distances greater than 24 km you can't hear it. Draw a diagram to illustrate this phenomenon.

Refraction

Have you ever wondered why it is easier to hear sound at night than in the day? It is particularly noticeable in open spaces, where fewer reflections occur. This is due to the refraction of sound, which results from the fact that the speed of sound changes with temperature — sound travels faster in warmer air. During the day the air directly above the ground is warm whereas higher up it is cooler. Therefore sound waves, instead of propagating parallel to the ground, are refracted upward. This refraction will not be an abrupt change but a gradual change, as there is no distinct boundary between the warm and the cool air. At night the ground is cooler than the air and the sound waves will be refracted downward. So during the day sound waves are refracted upward away from observers while at night they are refracted downward towards the observer. In the city this effect is not as noticeable due to reflections from buildings.

Air movement can also cause the refraction of sound, as wind affects the motion of the particles that are necessary for the propagation of the sound waves. (See Figure 16.4.)

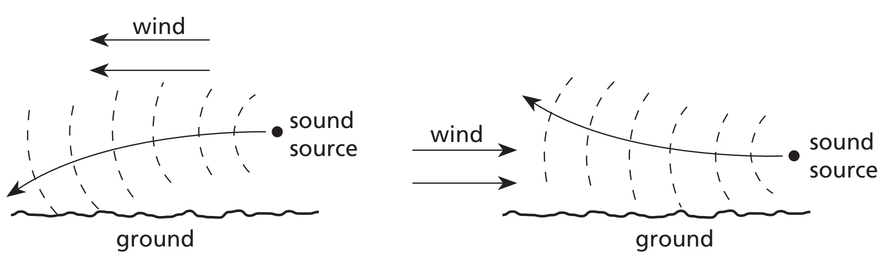


Figure 16.4
Air movement also causes the refraction of sound waves.

Diffraction

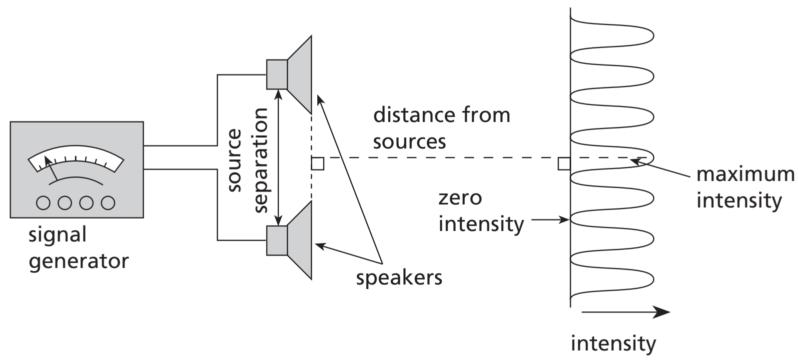
Imagine if sound waves were not diffracted around corners of objects but travelled in straight lines. It would mean that when you spoke, sound waves produced by your vocal cords would come out of your mouth and would go straight ahead. You would have to stand directly in front of a person to be heard. However, we know that people at your side can hear you because the waves diffract around the edges of your mouth. For the same reason, sound can be heard around edges of open doors. If you are put outside the classroom you can still hear the teacher's voice even though you are around the edge of the door and cannot see the teacher. Why is this?

You may also have noticed how echoes have a higher pitch than the original sound. Recall that long wavelengths are diffracted more than short ones, so the low frequencies (large wavelength) are diffracted more whereas the higher ones are reflected back to you.

Interference

As sound is a wave, interference abides by the same rules as for all waves. This can be demonstrated in the classroom by using a signal generator connected to two speakers, as shown in Figure 16.5.

Figure 16.5
The interference of sound can be observed using two speakers connected to the same signal generator making sound waves from the speakers in phase.



If you walk across the room in front of the speakers, constructive and destructive interference producing maximums and minimums of loudness will be detected. Constructive interference (loud) will produce sound when the path difference from the two speakers to the detector is $n\lambda$. Destructive interference will be detected, and therefore no sound heard, when there is a path difference of $(n - \frac{1}{2})\lambda$.

Pitch of sound

Pitch is our perception of whether a musical note is high like a soprano or low like a bass singer. The pitch of a sound refers to its frequency. If the sound has a high frequency it is said to be of a high pitch. The pitch or frequency of sounds emitted by humans is controlled by the tension in the vocal cords.

NEI

Activity 16.6 PRE-PUBESCENT BOYS' VOICES

1. It would be interesting to find out why boys, before puberty, have higher-pitched voices than after puberty. What is happening to cause this 'cracking'? Do some research to discover what is happening to their vocal cords.
2. Why doesn't this happen to girls?
3. At what age, if any, does a dog's bark 'crack' and become deeper?

EI

The 'snickometer'

Here's an idea for an experiment. TV broadcasts of test cricket are often accompanied by the use of a 'snickometer' to study the waveform of various noises, so that the commentators and third umpire can judge whether a ball made contact with the bat, pads, gloves etc. How do the following waveforms differ: bat on ball, bat on pads, bat on pad buckle? Propose a justifiable hypothesis before you begin. Taping the sounds and studying the waveforms on a CRO may be the way to go.

Loudness and energy

The loudness of a sound is related to the intensity of the sound, which depends on the energy carried by the wave. The energy is a function of the amplitude of the wave. Loud sounds carry large amounts of energy and cause large vibrations in the particles of the medium in which they are moving. This causes large vibrations in detecting equipment — ear drum, or microphone membranes. Refer to Section 16.9 for further discussion.

Sound quality

The quality of a sound produced by a musical instrument is dependent on the waveform associated with that instrument. The waveform is made up of a combination of frequencies and not just one frequency like that emitted by a 256 Hz tuning fork. The waveform produced by a musical instrument consists of a combination of the **fundamental frequency**, the lowest natural one, and a number of other less intense frequencies called **overtones** or **harmonics**. Overtones are whole number multiples of the fundamental frequency. This will be discussed in Section 16.6 where diagrammatic representations will help to explain the production of fundamental frequencies and overtones. Figure 16.6 shows examples of the waveforms produced by some common musical instruments.

Notes are said to differ in pitch by one **octave** when the frequency of one is double that of the other. In the frequency table of musical notes (Table 16.2), doh′ is one octave above middle C. In Latin, *octa* means 'eight', referring to the fact that there are eight notes from C to C′ inclusive.

Figure 16.6
The quality of the sound from an instrument depends on the mixture of frequencies emitted by the instrument. A tuning fork emits only one frequency.

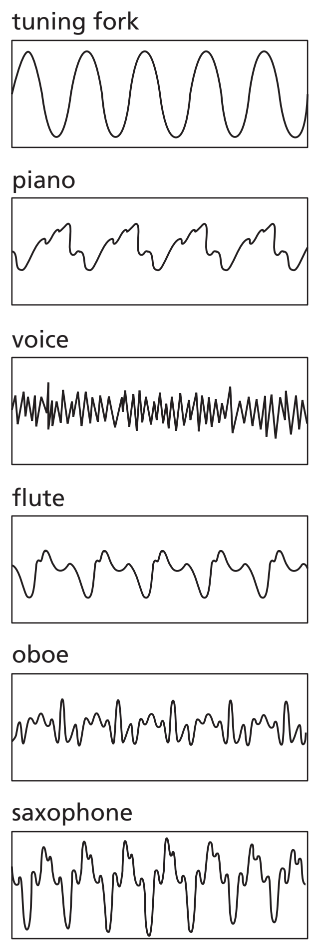


Table 16.2 MUSICAL NOTES AND THEIR FREQUENCIES

C	doh	256 Hz
D	ray	288 Hz
E	me	320 Hz
F	fah	340 Hz
G	soh	384 Hz
A	lah	427 Hz
B	te	480 Hz
C′	doh′	512 Hz

Questions

5 A student makes a noise in the back of the classroom with a frequency of 1000 Hz. Calculate how long it will take to reach the teacher in the front of the room 3.0 m from the student. What will be the wavelength of this sound?

6 A scuba diver taps one end of a 50 m copper pipe under water. Who will be first to know this has occurred — a scuba diver 50 m from the source or another diver holding the other end of the pipe? How much earlier will it be recorded?

7 Calculate the speed of a sound wave that has a frequency of 800 Hz and a wavelength of 42 cm. Can this sound be heard? Explain your answer.

8 When watching the fireworks at the local show you observe that you see the flashes from the exploding fireworks in the sky 0.50 s before you hear them. How far are they exploding from you?

9 A person standing on an observation platform in the mountains shouts and hears the echo off a cliff 1.5 s later. How far is the cliff from the 'outlook'?

10 A signal generator produces sound waves of frequency 1700 Hz, which are fed into two speakers placed 1.0 m apart. A student walks across the room 5.0 m from, and parallel to, the speakers. (Refer back to Figure 16.5.) How far apart will the student hear minimums of intensity? (The speed of sound is 340 m s^{-1}.)

11 Students sitting in the stands at an athletic competition 200 m from the start see the smoke from the starting pistol 0.6 s before hearing the sound of the gun. Find the speed of sound in air on this day.

ULTRASOUND 16.5

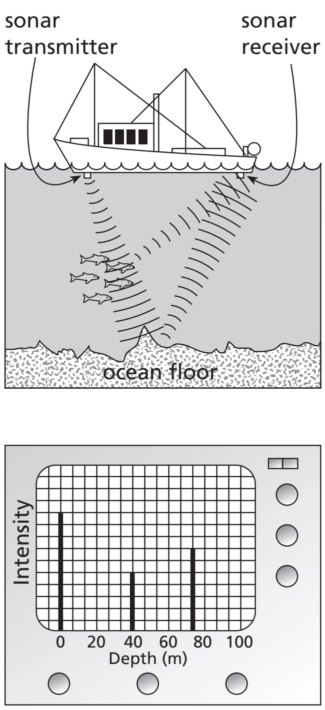
> **NOVEL CHALLENGE**
>
> A bat emits a frequency of 120 kHz. What size insect could it best detect? You may need to read the section on radar in Chapter 15 (Section 15.8) to help you answer this.

Ultrasound is sound at frequencies above that of human hearing range, that is, above 20 000 Hz. In Latin *ultra* means 'beyond'. Artificial ultrasounds are produced by vibrating quartz crystals, which are induced to vibrate by high-frequency alternating currents.

The uses of ultrasound waves or **ultrasonics** are increasing very rapidly especially in the medical profession. Reflection and refraction of ultrasound waves are used to see unborn babies, tumours and body organs. High-frequency ultrasound is used to make particles of objects vibrate at such a rate as to make them shatter. This is used by dentists to remove plaque from teeth and by doctors to break up kidney stones. (Refer to Chapter 33, Medical Physics.) Ultrasound is also used to heal muscular injuries; high-frequency ultrasound causes a muscle fibre to vibrate, thus generating heat and increasing the blood flow to the area, improving the rate of repair to damaged muscle.

In industry, ultrasound is used for cleaning small parts, for welding plastics and metals, for driving piles, and for drilling holes in glass. In most of these high-power applications the action is caused not by the direct agitation of the sound wave but by heating and bubble formation.

Figure 16.7 Sonar equipment makes use of ultrasound waves. Reflected waves are recorded on a ship's sonar equipment as echoes.



Ultrasound has been used for a number of years in sonar equipment developed during the Second World War to detect enemy submarines. Sonar comes from the term *so*und *n*avigation *a*nd *r*anging. Because ultrasound waves have shorter wavelengths, they are less diffracted by water than sound waves are and they are not absorbed by sea water as much as microwaves are. They can therefore penetrate to great depths in water. Objects in water, the ocean floor, a school of fish, submerged ships, or enemy submarines, can be detected by the reflection of ultrasound waves.

The use of ultrasound waves to find objects is thousands of years old. Can you think of an animal that uses ultrasound to find its way around?

This principle is also used in guidance systems to help the blind. Ultrasound transmitters emit waves that are reflected from objects and the reflected waves picked up by a detector are heard through an earpiece. Blind people are trained to make sense of the reflected sounds, thus enabling them to identify obstacles.

Bats

Bats use ultrasound to assist their poor eyesight. They are able to produce ultrasound pulses 0.010 s apart and of approximately 100 000 Hz with approximately 3.0 mm wavelength. Because the wavelength is so small these waves can reflect from small objects. Bats can also determine the nature of the reflecting surface — a hard surface gives a hard reflection; a soft powdery surface gives softer sounds. Some bats emit a short constant-frequency signal and can analyse the return signal for frequency changes (see Doppler effect later in this chapter). Bats' ears are also concave to concentrate the reflected sound waves.

Most insects' hearing extends into the ultrasonic region. Because insects are common prey for bats their ears have become sensitive to bat frequencies. On hearing these frequencies the insects fly the other way. Ingenious scientists make use of this in developing insect repelling devices that emit ultrasound waves at bat frequencies. Some species of moths have evolved very clever methods of evading bats. They are able to produce sounds at the bat ultrasound frequencies. When they detect bat ultrasonic waves they emit bursts of ultrasonic waves at bat frequencies but of a lower intensity to confuse the bats. This causes the bats to swerve or stop and listen. While this is going on the moth folds its wings and falls to the ground undetected. Other animals such as whales and dolphins use ultrasound but the method is not fully understood.

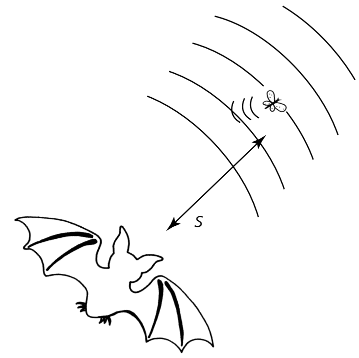


Figure 16.8
Bats use the reflection of ultrasound waves to find their way around as well as to find their food.

Questions

12 Why do bats use ultrasonic waves rather than audible sound waves?

13 A fishing vessel looking for schools of fish sends out pulses of ultrasonic waves and finds the echo returns 1 s and 3 s later.
(a) What is the explanation for this?
(b) At what depth would you expect to find fish?
(The velocity of sound waves in sea water is 1450 m s^{-1}.)

14 The minimum wavelength of ultrasound waves bats can emit is 3.3 mm. What is the highest frequency of sound that bats can emit?

15 Explain the technique of submarine captains resting their vessels on the bottom to prevent them being detected by a surface vessel's sonar.

16.6 THE PRODUCTION OF MUSICAL SOUNDS

Musical instruments produce sounds due to the standing waves set up in three different media:

- Strings; for example, the guitar, piano, violin, viola, cello.
- Air columns; for example, the flute, clarinet, recorder, organ, trumpet.
- Membranes; for example, drums, bongos, cymbals.

Strings

When a guitar string is plucked a standing wave is formed between the ends of the string. The frequency of vibration of the string produces the resulting sound. A number of different standing wave patterns can form in a string. (Recall from Chapter 13 that there are nodes formed at the fixed ends.)

The simplest standing wave pattern established in a guitar string is shown in Figure 16.9**(a)**.

The length of the string is equal to $\frac{1}{2}\lambda$. This string length produces the **fundamental frequency** (f_0).

The second possible standing wave established is shown in Figure 16.9**(b)**. Here the length of the string is equal to 1λ. This is the first **overtone** and because it is twice the fundamental it is also the second **harmonic**. Musicians often refer to overtones as **partials**.

The third possibility is shown in Figure 16.9**(c)**. This time the length is equal to $1\frac{1}{2}\lambda$. This is the second overtone and as it is equal to three times f_0 it is the third harmonic. Stringed instruments — guitars, banjos, violins, and pianos — as well as percussion instruments such as drums and bongos, rely on the above principle.

NOVEL CHALLENGE

If you told a violinist that you are a physicist and she should play the strings about one-seventh of their length from the end what would she say? Measure where she plays — is it one-seventh?

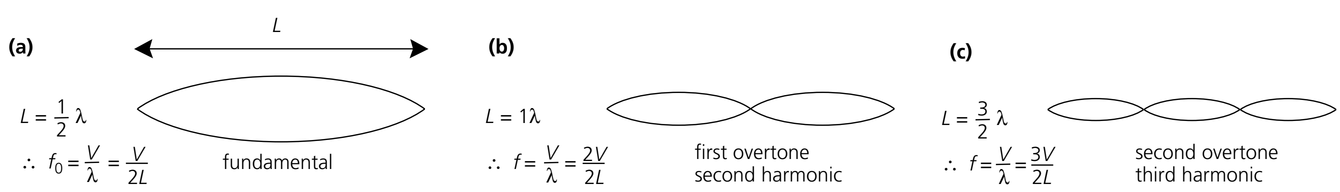


Figure 16.9
The types of standing waves set up in guitar strings.

The frequency of sounds produced by string instruments has been shown to depend on the tension of the string, the length of the string and the mass per unit length of the string. For example, the six strings of a guitar are all of the same length but they produce different notes because they have different tensions and masses per unit length. The relationship between the tension, length, and mass per unit length of the string is expressed by the mathematical formula:

$$f = \frac{1}{2L}\sqrt{\frac{T}{M}}$$

where f is the frequency of the note produced in Hz; L is the length of the string in metres; T is the tension in the string in newtons; M is the mass per unit length in kg m^{-1}.

Example

Find the fundamental frequency produced by a 48 cm wire of mass 1.0 g under a tension of 85 N.

Solution

- $L = 0.48$ m.
- $T = 85$ N.
- $M = 1.0 \times 10^{-3}/0.48$ m $= 2.1 \times 10^{-3}$ kg m^{-1}.

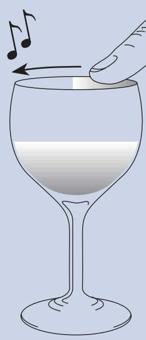
$$\begin{aligned} f &= \frac{1}{2L}\sqrt{\frac{T}{M}} \\ &= \frac{1}{2 \times 0.48}\sqrt{\frac{85}{0.00\,21}} \\ &= 209.6 \text{ Hz} \end{aligned}$$

NOVEL CHALLENGE

What is the squeaky sound when you are washing your hair with shampoo and what is the similarity with the shrill sound beginners produce on a violin?

NOVEL CHALLENGE

You have dipped your finger in some wine (or metho) and run it around the rim of a wineglass. A loud sound is produced. *Why doesn't it work if your finger is not degreased by the wine?*



Activity 16.7 RUBBER BAND GUITAR

EI

1. Place a rubber band between the index fingers of each hand. Stretch it and pluck it with your thumb.
2. Listen to the frequency of the sound produced. Stretch it more and listen to the sound again. Let it get less taut and listen to the sound again.
3. You may be surprised with the result. Try to explain it using the law above. *Hint:* as you increase the tension what happens to the mass per unit length?

Wind instruments

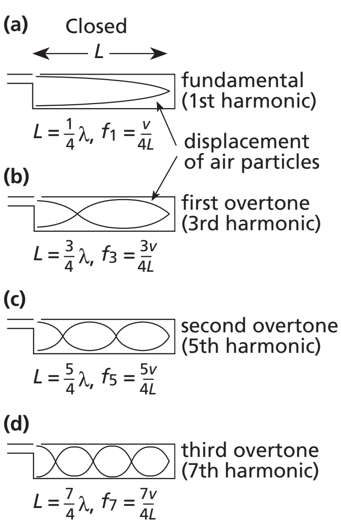
Wind instruments such as flutes also produce standing waves but in an air column. Again, as with stringed instruments, the air column can vibrate in a number of ways. The way it vibrates also depends on the nature of the instrument, whether it is closed-ended or open-ended.

Closed-ended pipes

The most common examples of closed pipes are the clarinet family and stopped flue pipes in the organ. In such pipes, the simplest standing wave pattern set up is shown in Figure 16.10**(a)**.

Notice that a node occurs at the fixed or closed end and an antinode occurs at the open or free end. For this standing wave the length of the pipe is equal to one-quarter of a wavelength, that is, $L = \frac{1}{4}\lambda$.

Figure 16.10
The first three harmonics produced by a closed-ended pipe. Notice nodes form at fixed ends and antinodes form at open ends. The different standing wave patterns produce the harmonics. (The lines show the displacement of air particles with time.)



Ch16-Walding 4th 25/8/04 11:57 AM Page 358

$$\therefore\ \lambda = 4L$$

Since

$$v = f\lambda$$
$$f = \frac{v}{\lambda}$$
$$f_0 = \frac{v}{4L}$$

This is the fundamental frequency (f_0), also called the first harmonic.

The second possible standing wave pattern produced is shown in Figure 16.10**(b)**. Here the length $L = \frac{3}{4}\lambda$.

$$\therefore\ 3\lambda = 4L$$
$$\lambda = \frac{4L}{3}$$
$$f = \frac{3v}{4L}$$

This is the first overtone but since its frequency is equal to 3 times f_0 it is the third harmonic.

Figure 16.10**(c)** shows the third possible standing wave. This is the second overtone or the fifth harmonic.

NEI Activity 16.8 OPEN-ENDED PIPES

Using the information in Figure 16.11, draw a diagram showing the next two waveforms for the 5th and 6th harmonics of the open-ended pipe. Also label these with correct overtone names.

You may also wish to research a quite famous mathematical relationship which is used in the physics of vibrating strings. It is called Mersenne's law, and relates the vibration frequency of the string or wire to its linear mass, tension and length. Present your findings in a research report including a discussion of the formula.
Find out Mersenne's full name, his nationality, and how Galileo fits into the story.

Open-ended pipes

The major difference between open-ended and closed-ended wind pipes is that the standing waves formed in open-ended pipes have antinodes at both ends. Apart from the clarinet family and stopped flue pipes in the organ, most other wind instruments are of the open-ended variety.

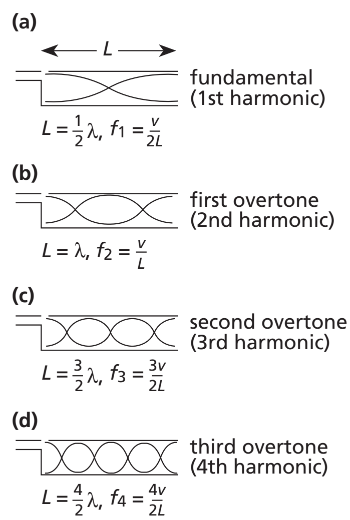
The simplest possible standing wave formed in an open-ended pipe is shown in Figure 16.11**(a)**.

Here

$$L = \frac{1\lambda}{2}$$
$$\therefore\ \lambda = 2L$$
$$\therefore\ f = \frac{v}{2L}$$

This is the fundamental or the first harmonic.

Figure 16.11
In open-ended pipes antinodes form at the open ends.



The second possibility in shown in Figure 16.11**(b)**.

Here $$L = 1\lambda = \frac{2\lambda}{2}$$

$$\therefore f = \frac{2v}{2L}$$

Therefore this is the first overtone or the second harmonic.

The third possible standing wave is shown in Figure 16.11**(c)**. This is the second overtone or the third harmonic.

Changes in length

In pipes and strings, you have seen that changes in length of the string or air column alter the frequency of the standing waves. Here are some examples with musical instruments:

Guitars When a string is pressed down on to a 'fret' (the wood or metal bars on the fingerboard), the length of the string is shortened and produces the next semitone higher. When the bass E string is pressed down at the fifth fret its frequency corresponds to that of the second string (A). A combination of finger placements on the different strings produces harmonic sounds called 'chords'.

Flutes and reeds (recorder, clarinet, organ, bassoon) Fixed-length pipes have been around since the beginning of human history but they can only produce set harmonics as shown in the discussion above. In about the tenth century, to fill in the missing tones of the musical scale, finger holes were cut into musical pipes. This effectively shortens the length of the air column, making an antinode at the position of the uncovered hole. In the eighteenth and nineteenth centuries, makers covered the holes of trumpets and horns with keys to enable them to play the complete scale.

Trombone Another way of providing missing tones was to increase the sounding length of the tube using a telescoping slide. Hand-stopping the 'bell' (the open end) was discovered in the eighteenth century to fill in more gaps in the instrument's harmonic series.

NOVEL CHALLENGE

An orchestra tunes up at the start of a concert but as the theatre warms up the musicians have to retune their instruments.

Do they find that the pitch of their instruments rises or falls as the theatre warms up? String musicians can change the tension of the strings. What do wind musicians do?

End correction

The wavelength formulas of closed-ended and open-ended pipes discussed above are not strictly correct. The particles of air at the open end of the tube do not strictly vibrate in one dimension, and a small correction, which depends on the diameter of the pipe and takes account of the motion of the particles in other dimensions, is needed.

To be strictly correct, the fundamental wavelength produced in a closed-ended pipe is:

$$\lambda = 4(L + 0.4d)$$

where λ is the fundamental wavelength; L is the length of the pipe; d is the diameter of the pipe.

For an open-ended pipe this correction factor is required for both ends. Thus the fundamental wavelength produced in an open-ended pipe is:

$$\lambda = 2(L + 0.4d + 0.4d)$$

or $$\lambda = 2(L + 0.8d)$$

Only apply end correction in problems when you are told to do so.

Activity 16.9 OPEN-ENDED PIPES

Part A

1 Draw the next two possible standing wave patterns. (Refer to Figure 16.11.)

2 Determine the overtones and harmonics of these two.

Part B

1 Blow into a recorder gently and listen to the sound. If you don't have a recorder, use a test tube half filled with water.

2 Blow harder and see if you can produce the first overtone. Musicians call this 'overblowing'.

3 Blow harder still and if you are lucky you may create the next overtone. It sounds awful but this is physics, not culture.

TEST YOUR UNDERSTANDING

Have you ever heard someone with a very loud and expensive car stereo driving down the road? Why do you just hear a thumping bass sound and no high frequencies if it is supposed to be such a good stereo?

Activity 16.10 A USEFUL SUMMARY

Prepare a data table as shown (Table 16.3) and fill it out for the first four standing wave patterns for **(a)** strings; **(b)** closed pipes; **(c)** open pipes. Keep this for revision.

Table 16.3

DIAGRAM	L	λ	f	OVERTONE NUMBER	HARMONIC NUMBER

Questions

16 **(a)** Is a higher frequency note produced in a long guitar string or a short string of the same tension?

(b) Find the frequency of the note produced by a 51 cm guitar string of mass 0.50 g under a tension of 90 N.

17 A student blows across the mouth of a closed-ended plastic pipe 0.25 m long. Calculate the fundamental frequency and the frequency of the third harmonic.

18 The distance between the reed and a hole in a recorder is 15 cm. Calculate the fundamental frequency and the next three harmonics that are heard when this instrument is played.

19 Draw the standing wave patterns set up in an open-ended pipe whose length is 20 cm and where the pipe's length equals **(a)** $1\frac{1}{2}\lambda$; **(b)** 3λ; **(c)** $\frac{1}{2}\lambda$. **(d)** Calculate the resultant frequencies emitted by the pipe in the above situations.

Ch16-Walding 4th 25/8/04 11:57 AM Page 361

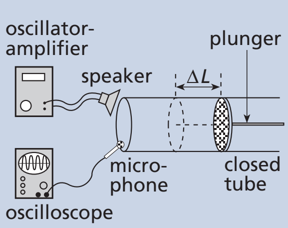
RESONANCE AND FORCED VIBRATIONS 16.7

PHYSICS UPDATE

The ‘trombone duckbill’ was a 10 m high dinosaur known as Parasaurolophus. It had a 1.5 m long skull (see figure), which had a large cavity consisting of tubes and chambers. Scientists at the New Mexico Museum of Natural History deduced that this was a resonance chamber used for mating and warning calls. The sound produced would have been about the same frequency as the lowest note on the piano. You can sample this sound on the Internet — just search for the museum (nmmnh + trombone).

TUBE RESONANCE

A closed tube with a variable length can be made to resonate to particular sound frequencies. The change in length of the tube between successive resonant positions is equal to one-half a wavelength. Thus $v = \Delta L\ .\ 2\,f$ This fact can be used to measure the speed of sound in the laboratory using the apparatus below.



Forced vibrations

When a tuning fork is struck with a mallet it vibrates at its natural fundamental frequency as well as emitting a few less intense lower order harmonics. This fundamental frequency depends on the length, thickness and composition of the fork. The intensity of the sound produced can be increased by placing the end of the fork on a table top. The table top is forced to vibrate at the same frequency as the fork thus intensifying the sound produced. The same occurs for a guitar string. When it is held between two clamps and plucked it does not produce a very intense sound, but when attached across the bridge of a guitar and plucked the sound is more intense because the wood of the guitar (the ‘soundboard’) is forced to vibrate in response to the vibrating string. Violins, basses and other musical instruments use the same principle of forced vibrations to intensify the sound produced.

Activity 16.11 FORCED VIBRATIONS

1. Excite a tuning fork and time its ‘life’ — that is, how long you can hear it.
2. Repeat the above but this time hold its stem on a desk.
3. What do you hear now? Why is this?

Resonance

Resonance is the effect that occurs when a body vibrates at its natural frequency. All bodies possess a natural resonance frequency. They can be made to start to resonate by another body touching them or being in close proximity to another vibrating object at the correct frequency. For example, a vibrating tuning fork can cause another one of the same frequency to resonate if it is close by. Here are some practical examples of resonance:

- Resonance vibration is why soldiers do not march in step when crossing older bridges as they might cause the bridge to resonate and possibly collapse.
- Mechanical resonance caused the collapse of the Tacoma Narrows Bridge in the USA in 1940. On a windy day four months after the completion of this suspension bridge it began to vibrate at its resonant frequency, causing it to collapse.
- Opera singers can shatter glasses because they can sing notes at frequencies that hit the resonance frequency of the glass.
- Singing in the shower sounds so good because the column of air in the shower enclosure is about the right dimension to resonate as a closed pipe to amplify many singing notes in your voice.
- If you run your moist finger around the lip of a crystal glass you hear a high-pitched squeal. This is the resonant frequency of the glass. You may first have to dip your finger in wine or metho to remove traces of oil.
- You may also have noticed that your hair gets squeaky clean after you wash it. This too is a resonance effect of longitudinal vibrations in the hair strands.
- Much louder sounds occur in musical instruments when standing waves in the tubes resonate in harmony with the vibration in the mouthpiece or reed, etc. This can be demonstrated if a tuning fork is attached to a resonance box. The sound of the tuning fork becomes much louder.
- Good loudspeakers are designed such that no one (or more) of their resonance frequencies (of which there are many) is dominant. Otherwise you would hear some frequencies louder than others.
- A seashell acts like a closed-ended pipe or resonator. The surrounding soft background noise provides sounds containing all frequencies the ear can hear. However, the shell increases the intensity of the frequency that is the same as its resonance frequency, thus creating a louder, almost pure, frequency or tone.

Ch16-Walding 4th 25/8/04 11:57 AM Page 362

Activity 16.12 RESONANCE

Activities to demonstrate resonance:

Part A

1 Set up apparatus as shown in Figure 16.12.

2 Set one pendulum swinging and observe what happens to the other.

3 Shorten one to 15 cm and repeat the experiment.

4 What do you notice?

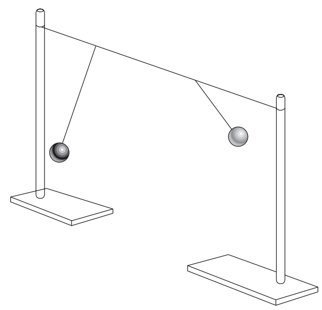


Figure 16.12 The apparatus used to demonstrate resonance. If one weight is made to oscillate the other will begin to oscillate.

Part B

1 Excite a tuning fork and hold it near your open mouth.

2 If resonance does not occur explain why and try another fork.

3 What distinction can you find between boys' and girls' mouth cavities?

Part C

1 Suspend a ping-pong ball on a piece of thread so that it hangs just touching the prong of a tuning fork.

2 Excite another matching fork nearby and see if the ball moves.

3 Explain this in terms of resonance.

NOVEL CHALLENGE

If you hold the right tuning fork up to your mouth cavity you can cause the cavity to resonate. *Would you expect boys or girls to have the lower resonant frequency? Why?*

16.8 BEATS

Beats are heard when sound waves of slightly different frequencies occur together. For example, when frequencies of 320 Hz and 322 Hz occur together, the constructive and destructive interference of these sound waves causes sounds that get louder and softer at regular intervals. The above example has a **beat frequency** of 2 beats per second, which is the difference in the two frequencies: $f_B = f_2 - f_1$.

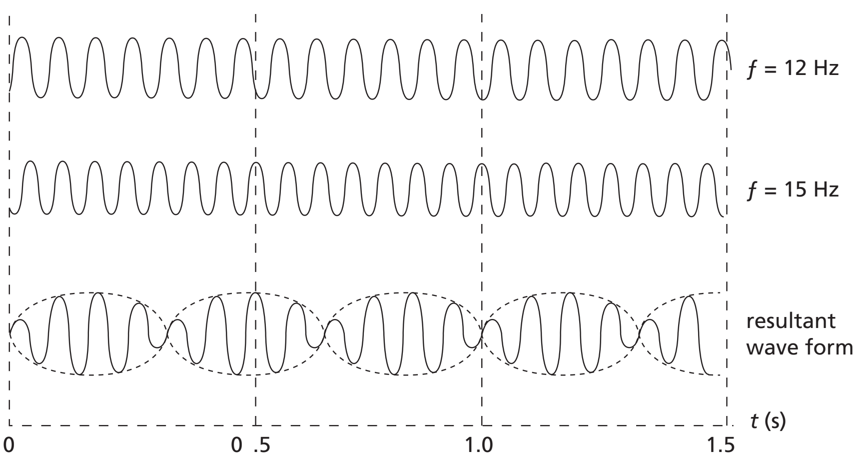


Figure 16.13 Beats are formed by the constructive and destructive interference of waves that are produced together but have slightly different frequencies.

Beats are used to tune pianos and other musical instruments. If a note on the piano and a tuning fork are sounded at the same time beats may be heard. The tension in the piano string can be adjusted until no beats are heard. At this time the piano is producing the same frequency as the tuning fork. Do you wonder why it is called a tuning fork?

EI Activity 16.13 BEATS

1 Sound a pair of matched tuning forks together and listen for beats. There should be none.

2 Add a small lump of Blu-tack or a rubber band on one prong of one fork and excite both again.

3 What do you hear?

4 Does the beat frequency increase or decrease as the Blu-tack 'load' is moved down the prong? Explain why.

5 Take a look at Figure 16.5 again and redraw the diagram to illustrate how a similar set-up could be used in the laboratory to produce 'sound wave beats'. You will need two signal generators. Try to set up the system with your teacher.

Questions

20 **(a)** Explain what resonance is and how it is produced.
(b) Explain why some older cars and buses start to vibrate when their engines reach certain revs.

21 State the beat frequency when the following pairs of tuning forks are sounded together: **(a)** 220 Hz and 217 Hz; **(b)** 340 Hz and 336 Hz; **(c)** 682 Hz and 688 Hz.

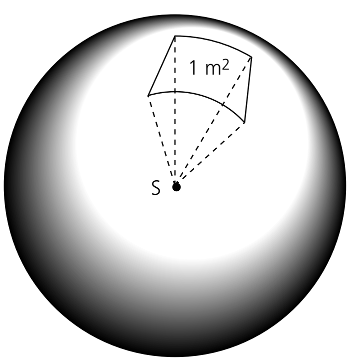
INTENSITY OF SOUND 16.9

One of the commonest properties of sound discussed in newspapers or on TV is sound intensity. There are ongoing protests about noise levels by people who live near airports or near outdoor concert venues. People living near an airport have to cope with noise levels of about 100 dB. The level of noise produced by rock bands may be as high as 120 dB. To safeguard the health and safety of workers, governments pass legislation that places restrictions on the noise levels in working environments. However, what do these noise levels 100 dB, 120 dB, etc. mean?

Sound waves, like all waves, carry energy, which spreads out from the source like a pebble being dropped into a pond. Suppose we take a 1 metre square area at some distance from a source, as shown in Figure 16.14.

The intensity of sound at this distance from the source is defined as the amount of sound energy that passes through this area per second.

Figure 16.14
The intensity of sound is measured by the sound energy passing through a square metre square area per second. It is measured in W m^{-2}. The intensity level is measured in dB.



The absolute sound intensity (I) is the energy carried by the waves per second through an area of 1.0 m^2

Since power is energy per unit time, then absolute sound intensity is the power passing through a unit area perpendicular to the propagation of the wave.

$$I = \text{power/area}$$

Therefore the units of absolute sound intensity are W m^{-2}.

A range of intensities of various sound sources is given in Table 16.4. Notice the large range of intensities the human ear can detect.

Table 16.4 SOUND INTENSITIES OF VARIOUS COMMON SOURCES

NOISE	ABSOLUTE INTENSITY, ($W\ m^{-2}$)	RELATIVE INTENSITY LEVEL (dB)
Jet plane taking off	10^3	150
Pain-producing	1	120
Rock concert	1	120
Chain-saw	$10^{0.5}$	115
Power mower	10^{-2}	100
Jackhammer	10^{-2}	100
Noisy restaurant	10^{-4}	80
Vacuum cleaner	$10^{-4.5}$	75
Ordinary conversation	10^{-6}	60
Average home	10^{-7}	50
Purring cat	$10^{-9.5}$	25
Whisper	10^{-10}	20
Rustling leaves	$10^{-10.5}$	15
Faintest sound that can be heard	10^{-12}	0

The terms sound intensity and sound loudness do not mean the same thing, although they are related. The absolute intensity of any sound wave of a given frequency depends on its amplitude. The greater the wave amplitude, the greater the sound intensity. This is reasonable since the greater the amplitude the greater the energy initially expended in setting up the sound wave. The sensation of loudness is related to the measurable quantity, absolute intensity. In general, sound waves of higher intensity are louder to the ear, but the ear is not equally sensitive to all frequencies. Consequently a high-frequency sound may not seem as loud as a low-frequency sound of the same intensity. The relationship between intensity and loudness is not linear. For example, for a sound of a given frequency to be twice as loud to the ear it must have ten times the absolute intensity. The relationship between the two is therefore logarithmic.

Because of the great range it is often convenient to use a comparative scale to express relative sound intensity levels — the **decibel scale**. This is a logarithmic scale.

$$\beta = 10 \log (I/I_0)$$

where I is the intensity of the sound in $W\ m^{-2}$; I_0 is the reference level, taken as the least audible sound ($10^{-12}\ W\ m^{-2}$); β is the relative intensity level in dB.
Therefore the relative intensity level of a whisper in dB is:

$$\begin{aligned} \beta &= 10 \log \frac{10^{-10}}{10^{-12}} \\ &= 10 \log 10^2 \\ &= 20 \text{ dB} \end{aligned}$$

The conversion of absolute sound intensities to relative intensity levels in dB for many common sounds is also shown in Table 16.4.

The decibel scale was named in honour of Alexander Graham Bell (1847–1922), the inventor of the telephone. One bel (1 B) is equal to 10 decibels (10 dB). The bel is too big for general use.

NOVEL CHALLENGE

When the physics laboratory is quiet, a dropped pin can be heard clearly at the back of the room. Calculate the energy arriving to the ear of a person at the back. If you want second-hand data, assume the pin has a mass of 0.2 g and it is dropped from a height of 1.0 m (use GPE = mgh). Assume all GPE is transformed into sound energy that radiates outwards as a large spherical surface ($A_{sphere} = 4\pi r^2$). *Calculate the amount of energy per square centimetre at the back of the room (say r = 5 m). Is it more or less than 10^{-9} J/cm^2? Not much huh?*

NOVEL CHALLENGE

The Tacoma Narrows suspension bridge in the USA collapsed in a mild windstorm (67 km/h) on 7 November 1940. It started oscillating up and down 30 times per minute with an initial amplitude of 1 m, which later increased to 8 m prior to collapse. You can simulate the effect by directing an air dryer jet sideways onto a strip of paper held at both ends. *What different motions can you achieve? Did you get twisting and bending? How does tautness affect it?*

NOVEL CHALLENGE

A Think of sound waves radiating outwards in a spherical shell. Calculate the intensity of a rock band's sound at 100 m if it has an intensity of 120 dB close up. Note: the surface area of a sphere = $4\pi r^2$.

B Humans' hearing is pretty sensitive. If a pin is dropped from 1 m high on to a desktop, you can hear it from the back of the classroom (try it!). Make some estimations to calculate the sound energy received by your ear. Recall that the sound radiates out as a concentric spherical surface and this has an area of $4\pi r^2$. How far is it to the back of a room? How big is your ear hole? What does a pin weigh?

Questions

22 Convert 10^{-7} W m^{-2} to decibels.
23 Convert 5×10^{-8} W m^{-2} to decibels.
24 Convert 85 dB to W m^{-2}.

Noise pollution

Sounds from cars, aircraft, lawn mowers, and rock bands can be more than just annoying. **Noises** — sounds that are not periodic and lack order — can cause damage to the ear, and may cause temporary or permanent deafness. They can cause other physical ailments such as tiredness and lack of concentration as well as stress. Some controversies concerning noise:

- A great deal of debate occurs when changes to runways of airports result in houses being in flight paths.
- Councils in residential areas place bans on the operation of noisy machinery between certain hours, and especially on Sundays. See if you can find the local regulation with regard to this.
- Many complaints are received by police and councils about barking dogs and even squawking birds.
- Workplace safety regulations require operators of noisy machinery — tractors, pneumatic drills, jackhammers, etc. — to wear ear protection.
- Police also have a duty to control the noise emissions from cars.

All the above, combined with the emission of smoke, odours, etc., affect people's environment. People have a right to not have to suffer from activities that pollute their environment.

NEI Activity 16.14 NOISE POLLUTION

Use one of the five bullet points above to form the basis of a research report. Select one of the controversies listed and discuss fully your research into the laws of physics that are relevant to their solutions. You should present your own ideas to help solve the problem topic you choose.

MODERN SOUND TECHNOLOGY 16.10

Basic components

Photo 16.3
A typical home theatre system.

Modern audio engineering involves the recording, amplifying and playback of a range of sounds for music recording, public address systems and special effects such as in movie soundtracks. This process utilises both analog and digital signals and the aim is to represent and reproduce original sounds without distortion. Audio engineers need to be able to use musical instrument transducers, microphones, amplifiers, digital signal processors, audio and video cassette decks, CD–DVD players and speaker systems. These can be found in professional theatres and movie complexes, home theatre sound systems or even in modern car audio systems. Photo 16.3 shows a typical home theatre set-up with multiple components and speakers. We will briefly discuss some of these components here, but they are also further discussed in Chapters 24, 25, 26 and 31.

Microphones A microphone detects sound and converts the waves into electrical signals which can be further amplified and/or recorded. Dynamic microphones use a thin diaphragm attached to a voice coil which vibrates inside a magnetic field when struck by sound waves, producing a continuous AC micro-voltage. Crystal or piezo-microphones rely on a movable diaphragm that distorts a small piezo-crystal when sound waves strike and again produce AC micro-voltages. Capacitive type or Electret microphones rely on the capacitive effect of two small thin metallic plates separated by a dielectric gap, across which is a larger DC voltage. Sound waves striking the capacitor's movable plate cause a small AC micro-voltage variation in the DC voltage applied. (Refer to Figures 16.15 and 16.16.)

Especially in sound stage work, microphones have been developed that are based on these principles but are further enhanced. Shot-gun mikes provide a very narrow response beam; pressure-zone PZM mikes produce an omni-directional wide field response to sound-wave pressure variations. The microphones that most rock band singers use are dynamic types because they need to be used close to the mouth; they are fairly insensitive and do not pick up sounds from a wide direction pattern.

Amplifiers These devices electronically increase the voltage amplitude of the output signals from microphones, instrument transducers such as guitar pick-ups, or other auxiliary devices such as CD players or digital effects processors (synthesisers), in order to make the voltage and electrical power levels high enough to drive loudspeaker systems directly. Many electronic audio devices will also have many stages of internal pre-amplification before the output signal is sent to a separate multi-channel instrument or system power amplifier.

Many amplifiers rely on integrated circuit chip modules to carry out the voltage and power amplification, with output power measured in watts for a typical radio receiver amplifier, up to a few hundred watts per channel for a home theatre amplifier, and finally up to several thousand watts for large amplifiers used by the major rock bands. Amplifiers dissipate a large amount of heat and need to be protected from overheating with appropriate installation housings and cooling fans. Often amplifiers can sense an appropriate input signal level and switch into standby mode when not in use, to assist in heat control.

Loudspeaker systems These electrical devices are at the end of the audio chain and convert electrical AC signals from the amplifier outputs back into sound waves. There are many designs for speakers, and it is very complex to produce a design involving multiple drivers because the behaviour of sound waves inside closed or open boxes, as well as the nature of the listening environment, needs to be taken into account. The term **speaker** usually refers to the electromagnetic driver as well as its box housing, while the term **driver** refers to the actual device reproducing the sound. All drivers will eventually move air backwards and forwards by the use of a paper or polymer diaphragm or cone, producing rarefactions and compressions which are the sound pressure waves that our ears respond to. In many speaker driver designs the cone is attached via an assembly called a 'spider' to the voice coil, which carries the AC current waves from the amplifier. The voice coil moves as a result of being within the strong magnetic field of the speaker magnet assembly. When AC current flows through the voice coil the magnet induces a force on the coil which causes it to move laterally. This magnetic induction effect is discussed in more detail in Chapter 25. (Refer to Figure 16.17.)

Figure 16.15
A crystal microphone uses piezoelectric crystals, which are able to produce currents when subjected to pressure.

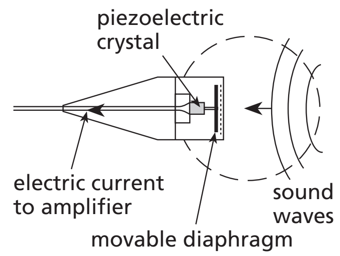
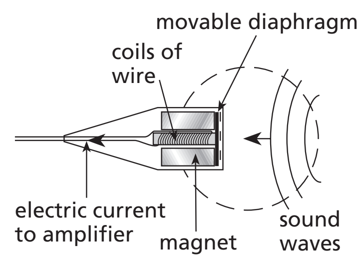


Figure 16.16
A moving coil microphone produces currents when the pressure of sound waves causes coils to move in magnetic fields.



NOVEL CHALLENGE

It is common to state amplifier output power in a variety of ways. Sometimes manufacturers use misleading terminology when describing output power in order to make their product seem better than it actually is. Find out the meaning of the measurement units called Peak, Peak to Peak, RMS or even PMPO; and find out which is the industry standard.
Why can the use of these terms be misleading to the consumer?

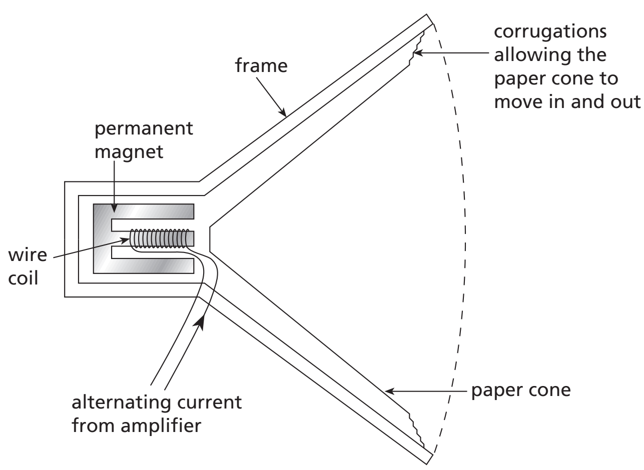


Figure 16.17
A schematic diagram of a speaker. Variations in the current in the coils cause variations in the movement of the paper cone, which results in the production of different sound waves.

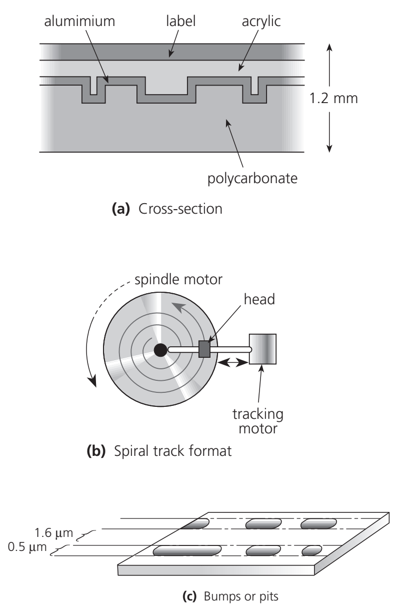
Modern speaker driver research is concentrating on the materials that make the magnet assembly. Australian CSIRO researchers are among the world's leaders, producing exotic magnetic alloys that provide very strong magnetic field densities. This enables speakers of high power-handling capability to be made smaller and smaller. A visit to the local hi-fi component retailer will confirm the incredibly small size of some modern speakers for the sound output they provide.

Speaker drivers are often designed to reproduce only a small part of the audio spectrum efficiently (from 20 Hz to 20 kHz); so when used in combination they need to be separated with electronic cross-over networks involving capacitor–inductor frequency filters. Some interesting names are given to various frequency component speaker drivers. For example, the **Subwoofer** is the high-power driver used to reproduce low-frequency effects, LFE signals such as explosions and deep bass music notes, usually below about 150 Hz. These are usually housed in large boxes. **Woofers** are used to reproduce low to mid-range frequencies up to about 1.0 kHz. **Squawkers or mid-range** drivers can handle a wide range of frequencies in the audio spectrum and are often used as the centre speech-reproducing driver, typically from several hundred hertz to about 50 kHz. **Tweeters** are the specialist high-frequency drivers that reproduce top-end sounds such as hisses and chirps from several thousand hertz right up to 20 kHz. Often they only need to be quite low in output power as the large amount of air movement is not necessary.

A well-designed speaker system may contain multiple drivers, with woofer–squawker–tweeter in the one box cabinet, and be able to handle simply one output channel from the amplifier. Speaker resistance or impedance to the flow of AC electric current is measured in ohms. Typical home theatre speaker systems are rated at 8 ohms, while typical car audio speaker systems may be rated at only 4 ohms.

The home theatre digital sound revolution

Figure 16.18
A compact disc.



Audiophiles of the past usually raved over their quadraphonic or stereophonic systems, which normally consisted of analog components such as amplifier, record player, AM–FM radio receiver, and one or more magnetic tape decks. This was usually coupled with an inefficient stereo left–right speaker system which, if the system was a good one, reproduced original music with minimal distortion and noise levels. With the coming of the digital revolution, audio and videophiles now have components such as digital AM–FM stereo PLL (phase locked loop) tuners, CD and DVD players, hi-fi stereo video recorders and DAT (digital audiotape) decks coupled with HD (high-definition) digital televisions, multi-channel and decoder amplifiers, which reproduce audio and video with incredible fidelity and without discernible electronic noise. The computer revolution has allowed this equipment to be used with ease to produce audio and video source material in the home without serious sound studios.

Let's now take a look some of the technology behind this digital audio–video revolution and find out how it works. The process of analog to digital conversion, which underlies this technology, is discussed in Chapter 24 (Section 24.2).

AM–FM radio tuners A radio tuner will receive transmitted amplitude or frequency modulated waves from a local radio station, usually in stereo (tuning principles are discussed in Chapter 31) and convert small microvolt signals into a form ready to be amplified.

Cassette or digital audio tape decks These contain electromagnetic induction heads that decode magnetic patterns stored on magnetic tape and convert them into an electronic signal stream, again allowing amplification.

Audio compact discs (CDs) An audio music (CD-DA format) or computer data (CD-ROM format) device is a simple injection-moulded circular disc of polycarbonate plastic about 12 cm in diameter and about 1.2 mm thick, onto which is evaporated a thin layer of reflective aluminium. This is then covered with a layer of acrylic for protection and marketing labels are attached to the top of the acrylic layer. When the CD is recorded or written to by a laser beam, small indentations that represent the 1s and 0s encoded digital bitstream are made into the reflective aluminium–plastic layer. (Refer to Figure 16.18, which of course is not to scale.) These indentations become raised bumps when read from the scanning pickup laser head side of the disc.

The track written onto a CD is one long continuous spiral 0.5 μm wide (1 μm = 1×10^{-6} m), laid down with separations of only about 1.5 μm. Typical data bumps on the track are at least 0.83 μm long and 126 nm high (1 nm = 1×10^{-9} m). For a typical 700 MB CD the recording track may be about 5.0 km in total length. The CD player has the difficult job of reading the data track and has four major functions to perform:

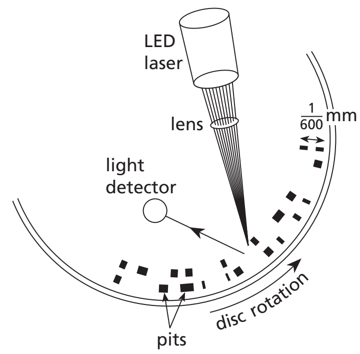
1 *The drive or spindle motor* must spin the disc at different speeds depending on what part of the data track is being read. Speeds of between 200 and 500 rpm are common, but CD-ROM discs in a computer drive may spin even more quickly.

2 *The tracking motor* moves the laser head linearly over the disc surface from inside to outside. The tracking motor position also determines the spindle motor speed because the reflective bumps towards the outer rim of the CD will be travelling more rapidly past the pickup head, and the spindle motor has to be slowed down so that data comes off the disc at a constant rate.

3 *The lens and laser assembly*. This is why all drives have the LASER CLASS device warning symbol attached to them, but in reality they are not dangerous at all when used normally. The lens system must focus the laser beam onto the lands and bumps of reflective aluminium. The laser beam passes through the polycarbonate layer and reflects off the bumps and lands, producing a variable light signal to the opto-electronic sensor. Further electronic digital gates in the drive provide the digital bitstream to the DAC and finally to the internal signal amplifier. Figures 16.19 and 16.20 show these processes.

4 *The error-correcting and subcode-data-reading electronics.* This system monitors any laser head misreads due to dust or scratches on the disc surface (called burst errors) and allows the bitstream to be recovered. The subcode data encodes position information on the disc so that operations like finding and skipping to a particular music track can be accomplished. These systems are most important on a computer data CD-ROM. Can you see why?

Figure 16.19
CDs contain pits that result in the interference of light to produce electrical currents.



tracks on a digital disc

It is interesting to further consider the differences between a normal CD and the recordable or writable **CD-R** format, or even the rewritable **CD-RW** format. Engineers are constantly improving the capabilities of these digital formats and player-readers. Figure 16.18 shows the layers of a normal CD. In the CD-R format there are no bumps or lands, just a smooth reflective aluminium layer that rests on top of a layer of photosensitive dye. On a blank CD-R disc this dye layer is translucent, but when the burner laser heats it up the dye becomes opaque. The data track thus becomes a series of dark and reflective spots. One disadvantage of this is that, once burnt, the CD-R layers are permanent.

To produce the re-recordable CD-RW format disc even more layers are required. Sitting below the aluminium layer is a special chemical layer containing a crystalline phase-changing compound of metallic alloy. Under normal conditions this compound's form is crystalline and translucent, but again if the burner write laser heats it above 600°C for an instant it melts and becomes amorphous and opaque even after it is cooled. This process allows the digital bitstream to be encoded as before. The CD-RW drive also contains an erase laser setting that can hold the phase-change layer long enough at its melting temperature to allow the compound to revert to its translucent state, thus allowing data to be rewritten over and over again.

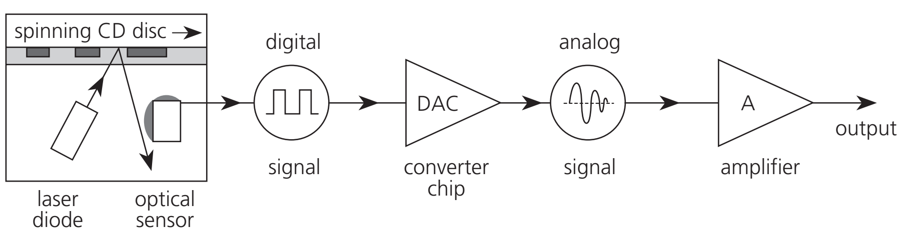


Figure 16.20
Schematic diagram of CD player electronics.

Digital versatile discs (DVDs) An early product of analog technology called the laser-disc provided better quality than magnetic videotape and had many advantages over that format, but did not last long in the industry. It has now been superseded by the digital versatile disc (DVD). A DVD is similar to a CD but can hold up to seven times as much data. Even the drive, tracking and read electronics is much the same as a standard CD. DVDs, because of their increased data-holding ability, are used for storing huge amounts of digital data such as complete multi-volume encyclopaedias or full-length MPEG-2 encoded (a data compression technique) movies and associated extras.

Compared to the 0.5 μm wide tracks of CDs, the DVD tracks are just 320 nm wide, separated by only 740 nm. The bumps encoded on the metal layers are just 120 nm high and have a minimum length of 400 nm. Not only is an aluminium layer used in DVDs but also semi-reflective gold layers are used to allow the laser beam to access multiple layers. Most DVD movies are encoded using 96 kHz 24-bit digital information, so this means that the DVD player must have at least a 96 kHz 24-bit DAC for replay. Table 16.5 lists the DVD layering process formats available.

Table 16.5 COMPARISON OF THE DVD LAYERING FORMATS

DVD LAYER FORMAT	DIGITAL CAPACITY	PLAYING TIME AVAILABLE
Single-sided/single-layer	4.38 GB	Approx 2 hours
Single-sided/double-layer	7.95 GB	4 hours
Double-sided/single-layer	8.75 GB	4.5 hours
Double-sided/double-layer	15.9 GB	Over 8 hours

To really enjoy recorded music at its brilliant digital best the latest DVD format is DVD-audio. This format requires the use of a 192 kHz 24-bit digital-to-analog converter DAC, which is higher than a typical DVD player. The following table compares the older CD-DA audio compact disc with the latest DVD-audio standard.

Table 16.6 COMPARISON OF CD-DA AND DVD-AUDIO COMPACT DISCS

AUDIO SPECIFICATION	CD-DA STANDARD	DVD-AUDIO STANDARD
Sampling rate	44.1 kHz	192 kHz
Sampling resolution	16-bit	24-bit
DAC Output levels	65 536	16 777 216

Home theatre digital sound formats As we said at the beginning of this section, one of the greatest advances in sound technology has been in the use of multi-channel sound systems and the improvements in video output. A typical DVD player and digital home theatre amplifier (receiver-amplifier if it also includes an AM-FM stereo radio-tuner) provides signal outputs for up to eight independent sound channels, commonly referred to as 'surround sound'. **Dolby Digital 5.1** or DD surround sound is one of the best known and is developed by Dolby Laboratories, famous for noise reduction technology in previous generations of sound equipment.

In general, to produce surround sound modes, electronic equipment needs to be able either to synthesise new channel information from the originally recorded simple stereo left–right (L–R) pair of signals, or to handle the already encoded multi-channel information coming from the recorded medium such as a DVD. The amplifier is said to contain sound system '**decoders**' if it is able to perform this function. Let's take a look at some of the more common digitally enhanced surround sound modes.

1 Three-channel stereo — an analog electronic simulation of an extra front centre channel: L–C–R.
2 Dolby Pro-Logic — an analog electronic simulation of a front centre channel, as well as the simulation of a mono rear channel, sent most often to two separate surround speakers: L–C–R–[SR–SL]. A most recent version called Dolby Pro-Logic 2 is available that increases the signal level amplitude of both the C and rear S channels.
3 Dolby AC-3 or Dolby Digital 5.1 — Full processing of a digital bitstream at either 384 or 448 kilobits per second of up to full discrete full bandwidth channels (20 Hz–20kHz) plus a low frequency effects LFE channel at a bandwidth of 120 Hz (this is the so-called 0.1 channel or subwoofer output): 6 speakers L–C–R–LS–RS–LFE.

Digital compression of the bitstream is used in all digital sound formats. DD 5.1 uses a compression ratio of about 12:1. Most amplifiers will provide only a low level signal output for the LFE channel, so it will often require extra amplification. Figure 16.21 illustrates the home theatre speaker set-up for this type of system compared to the audience position. It is important to remember that all speakers except for the subwoofer should ideally be at typical ear height, or just above, for the audience.

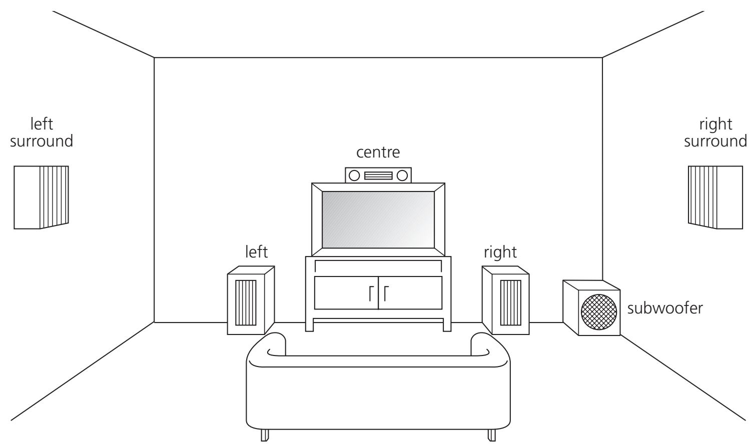


Figure 16.21
Dolby Digital 5.1 speaker set-up.

4 Digital Theatre System DTS — very similar to DD 5.1, with the advantage that digital compression of only 4:1 is used for supposedly greater fidelity. Typical DTS soundtracks are encoded with a 1.4 megabits per second bitstream, but of course require a separate decoder. The first DTS encoded DVD available in Australia was the hugely successful *Gladiator*, starring Russell Crowe as Maximus.
5 Dolby Digital Surround EX — again basically the same as DD 5.1 except that it includes an extra discrete sixth channel designed to be placed immediately behind the audience. It is like DD 5.1 enhanced with Dolby Pro-Logic: L–C–R–LS–RS–LFE–SC. The first DD-EX encoded movie was George Lucas's *Star Wars: Episode 1 — the Phantom Menace*.
6 Remaining systems worth mentioning in this context are the Sony Dynamic Digital Sound (SDDS) decoding, which is used primarily for large movie cinema sound reproduction, and the Motion Picture Experts Group (MPEG-2) encoding, which also can be used for DVDs.

You may also notice on recently released movies and DVDs the term **THX certification**. What this means is not another surround sound encoding system, but rather a set of performance standards established by the LucasFilm company and called the Tomlinson-Holman Experiment! This calls for particular functional and performance requirements from the audio equipment, such as decoders, equalisers, DVD players, amplifiers and speaker systems, both in professional cinemas (THX-Ultra for over 85 m^3 spaces) and home theatres (THX-Select for typical spaces of around 57 m^3). You know you have the ultimate in a home theatre system if you are able to achieve 'Home THX certification'. Now that would be something to brag about, but it would be very, very expensive.

The rear panel of a modern DVD player or home theatre amplifier contains a large number of connector jacks to allow for signals into and out of the device. Taking a typical DVD player, we might find at least the following:

- *Audio outputs:* 2 × analog stereo L–R, RCA connectors colour coded as red for right and white for left. 6 × analog 5.1 channel outputs, a single coaxial RCA digital output as well as a single optical digital output for the fibre-optic cable that is capable of carrying all six DD channels.
- *Video outputs:* single or double C-V (composite video) connectors, which provide lowest quality video signals directly to a television monitor; these are usually colour-coded yellow. One or two S-video outputs which provide good quality video signals to an S-video equipped monitor or data projector. The highest-quality signals are provided by the three separate RCA component video connectors, colour-coded red, green and blue. These provide RGB signals directly to the electron guns of the monitors or high-definition display devices.

Digital television, data projectors and plasma screens As well as surround sound systems, any good home theatre set-up will also provide the best in video displays, whether that be by normal large-screen television monitors, high-definition digital TV, wide-screen plasma panels or digital data projectors.

In Australia as from 2001 most TV stations began broadcasting some programs in digital format (wide-screen digital) as well as the normal analog PAL format. A fully digital TV monitor, or at least an electronic set-top box (STB decoder, similar to the currently available analog cable TV), is required to view these transmissions. The Australian government will phase out all PAL analog television by about 2008, so by then we will all be watching it. The main advantages of this digital technology will be greatly improved picture quality, which will be about the same as current PAL DVDs for the standard form (**SDTV**) and even better for high definition (**HDTV**). Television reception will be improved, as will the availability of surround-encoded sound formats such as DD 5.1 and the ability to include caption, subtitles and multi-angle viewing for sporting events etc.

The video or picture information is encoded using the MPEG-2 digital compression format at two resolution settings, being either SDTV = 576i = interlaced scanning at 50Hz, 576 active scanning lines and 720 pixels per line (720 × 570 pixels screens) or HDTV = 1080i = interlaced scanning at 50 Hz, 1080 lines at 1280 pixels per line (1080 × 1280 pixel screens). SDTV can be broadcast in either normal 4 × 3 ratio screen dimensions (letterbox format) or the 16 × 9 ratio (wide-screen); HDTV will be available only in wide-screen. Consumers will start to notice SDTV, HDTV and STB devices appearing in retail stores over the next few years. Good websites from which to discover more about the Australian digital TV scene are either Robert Simons at <http://www.digitaltv.com.au/index.html>, or the Digital Broadcasting Authority at <http://www.dba.org.au/>, or even the newsgroup aus.tv.digital. Take a look.

Another method of displaying video information is with **video or data projectors**. These are typically used where larger image sizes are required. Two main techniques are used, called LCD (liquid crystal display technology, better for smaller screens) and DLP (digital light processing technology, more suitable for larger cinema screens). George Lucas, in premiering his digitally produced *Star Wars: Episode 1 — The Phantom Menace*, called for DLP projectors to enhance the visual experience.

LCD projectors work by splitting the light from the projector lamp into three primary colour beams — red, green and blue (R-G-B). Each beam then passes through a small LCD panel which acts like an electronic slide. Each LCD panel typically has 800 × 600 pixel elements, and can be switched on and off according to the video recorded digital signal, to provide a correct full-colour image when the beams are recombined and passed through the main projector lens onto a screen. LCD panels and the transistors used to switch them on and off can produce quite a lot of heat, and the final image can look very pixellated on the screen due to the spacing between pixel elements on the LCD panels.

DLP projectors developed by the Texas Instrument semiconductor company are based on TI's digital micro-mirror devices (DMDs invented by Dr Larry Hornbeck in 1987), which are basically large-scale integrated circuit chips that contain a huge array of micro-miniature aluminium-coated mirror elements that pivot back and forth under the control of a separate digital signal processor (DSP). One mirror exists for each pixel element, which has an area of only 16 μm^2 at gaps of 1.0 μm; hence a 1024×758 (XGA) resolution DMD will contain 786 432 mirrors. Also, micro-torsion bar tilting of the mirror pivots is done with electric fields so that each mirror element is pointing (or not pointing) at the screen. DLP projectors can use one, two or three DMD devices. The three-chip approach (each separately handling R-G-B information) is the best, and the large expensive HDTV DLP cinema projectors contain DMDs that provide 1 310 720 mirror elements, giving S-XGA resolutions of 1280×1024 pixels. Texas Instrument specifications suggest that the DMD mirror elements switch at a rate of 5000 times per second. Not bad? Figure 16.22 indicates the DMD device structure as well as the two-chip DLP set-up.

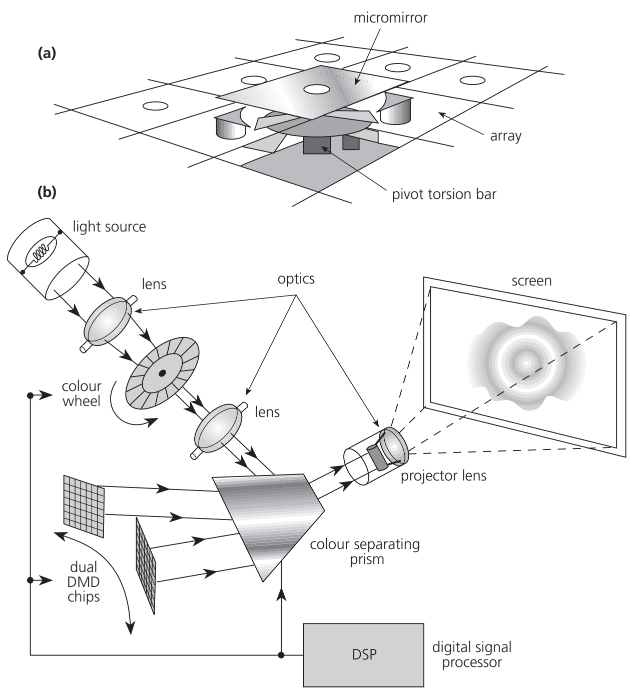


Figure 16.22
(a) DMD device.
(b) Two-chip DLP projection system.

A recent addition to the armoury of possible display devices is the plasma screen displays. These devices can be very thin in design and can easily accommodate large wide-screen video formats. Plasma screens do not use cathode ray tubes (CRTs) or LCD-DLP projectors, instead creating images by using an array of cells that receive a constant flow of low-pressure neon and xenon gas (hence the use of the word plasma). The cells are arranged in a rectangular matrix between sheets of thin glass and are covered with electrodes. When the electrodes are fired, the voltage stimulates the gas to emit UV light in a similar way to a fluorescent light tube. This UV light is then converted to visible coloured light by hitting phosphor coatings on another layer. Each cell is restricted to one particular R-G-B and each pixel that makes up the display image has three different cells, one for each colour. Plasma screen displays are still quite expensive, just like DLP projectors, but as with normal consumer market forces their prices will come down as manufactured numbers increase and more consumers start using the technology.

One final point worth noting concerning the home theatre sound technology improvements is that your lounge or living-room design might not be up to scratch for the best viewing and audio experience. You may very well research this point, making use of acoustics ideas from the next section. We're sure, though, that your parents will be quite willing to pay for an upgrade to your own study room so that you may have the best conditions when studying for your next physics paper. Right?

Activity 16.15 RESEARCH ON HOME THEATRE TECHNIQUES

1 The electrical signal sent from the amplifier to the speakers is sent through a filter to select the correct range of frequencies for each of the different speaker types. Discuss the construction of a filter and explain the physics principles involved. Compare and contrast a low-pass and a high-pass filter.

2 The circuit for a Dolby Digital decoder is not available, because the exact decoding process is strictly an industrial secret and needs to be added by manufacturers without alteration. Try to report on circuits for other types of surround sound decoders that are more freely available.

3 Three characteristics of a sub-woofer are: **(a)** you only need one for a stereo system, whereas other speakers handle left- and right-channels individually; **(b)** it doesn't matter where you put it; **(c)** it usually has its own power supply. Explain the physics behind each of these design features.

ROOM ACOUSTICS 16.11

Acoustics, the study of sound, sound technology and its effect on humans, plays an important part in the design of rooms, auditoriums and theatres, especially those used for high-quality performances. Walls, floors, ceilings all cause reflections of sound. These reflections may cause deterioration of the performance. However, some reflection is required — a listener would have trouble hearing if sounds were received from the source alone. These reflections affect the reverberation time and thus the acoustical quality. The **reverberation time** is the time it takes for the sound intensity to fall to one-millionth of its original intensity; that is, to fall by 60 dB. In lecture theatres, concert halls, etc. it is an important consideration. Multiple reflections are undesirable in lecture theatres because they obscure the spoken word. It is therefore desirable that the reverberation time be less than 1.0 s. Reflections are more desirable in concert halls as we want the listener to be totally immersed in the sound. It is therefore desirable that the reverberation time be of the order of 2.0 s. The reflections and reverberations lift the intensity of the sounds, but no single reflection should arrive at an ear later than one-twentieth of a second after the original sound or it will be heard as an echo.

The reverberations in a room depend on the size and shape of the room and the way the sound-absorbing linings of the room reflect or absorb sound. The absorbing quality of materials varies. Table 16.7 indicates the absorbing quality of several common materials. However, these qualities also depend on the frequency of the sound.

Table 16.7 SOUND ABSORPTION QUALITIES OF SOME COMMON MATERIALS

MATERIAL	SOUND ABSORPTION QUALITIES (PERCENTAGE OF INCIDENT ENERGY ABSORBED)
Glass window	4
Plasterboard	10
Carpet	25
Thick wool over brick	70

THE DOPPLER EFFECT 16.12

Everyone has observed the variation in frequency of sound from a police car, ambulance or fire engine as it rushes past. As the vehicle is approaching, the frequency of the sound of the siren is higher and at the moment it passes the frequency drops. This apparent change in frequency due to the object's motion is called the **Doppler effect**, and is attributed to an Austrian physicist and mathematician, Christian Doppler (1803–1853), who first investigated this phenomenon. This frequency change can be noticed if either the source of the waves is moving toward or away from a stationary observer, or the observer is moving toward or away from the stationary source of the waves.

Figure 16.23 will help to explain this effect.

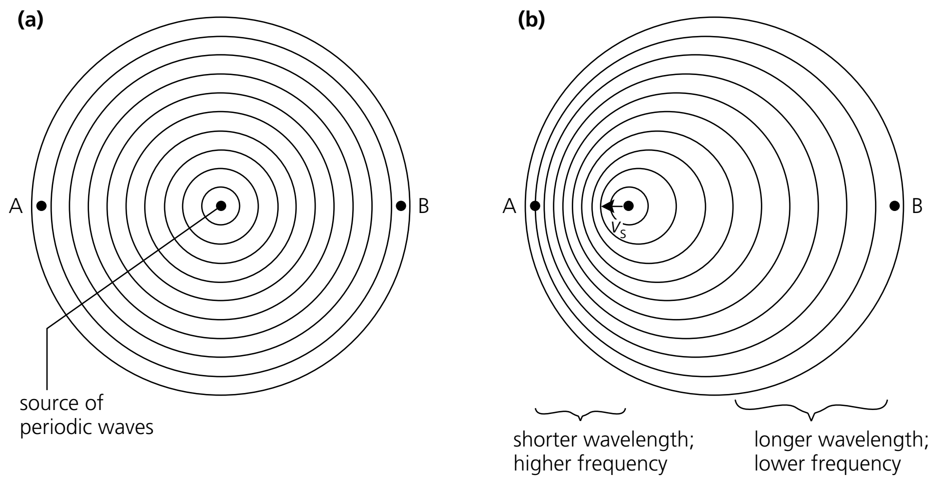


Figure 16.23
The Doppler effect. When the source of sound is moving toward the observer the waves are compressed.

Figure 16.23**(a)** shows the wave pattern produced by a stationary source. The waves are equally spaced and will arrive at points A and B at regular intervals. Now if the source is moving toward A (Figure 16.23**(b)**), the waves will be closer together in the direction of motion than if the source was not moving.

For example: Let the velocity of the waves be 10 m s^{-1}, the velocity of the source be 5.0 m s^{-1}, and the frequency of generation of the waves be 10 per second (10 Hz). If the source was not moving, after 1 second 10 waves would be produced and the furthest one would be 10 m from the source. The wavelength would be 1 m. However, if the source was moving at 5.0 m s^{-1}, the source would have moved 5.0 m and the 10 waves would exist in the 5.0 m between the source and A. The wavelength would be 0.50 m. The first wave would still have moved 10 m to point A. The waves would still have the same velocity as they would still be in the same medium, but the wavelength would be shorter and therefore the frequency increased.

For an observer at point B: After 1 s the distance between the source and B would now be 15 m with 10 waves between the source and B. The wavelength would be 1.5 m. This is a greater wavelength therefore the frequency would be lower.

For Figure 16.23**(a)** the wavelength is given by the equation: $v = f\lambda$ or $\lambda = v/f$, where λ is the wavelength; f is the frequency; v is the velocity of the wave.

Now for Figure 16.23**(b)**: the period of the wave $T = 1/f$ and the velocity of the source is v_s. After a certain time (t), after the production of n waves, $t = nT$: the distance between the source and point A is the difference between the distance the waves have travelled and the distance the source has moved.

This distance is:

$$\begin{aligned} d - d_s &= vt - v_s t \\ &= (v - v_s)t \\ &= (v - v_s)nT \\ &= (v - v_s)n/f \end{aligned}$$

NOVEL CHALLENGE

The frequency shift effect was first proposed by Doppler in 1842 but the first experiment was not done until French scientist Buys-Ballot had a go in 1845. He arranged for a carriage full of brass musicians to go past him in a train as they blew a steady note. *To study this effect quantitatively, what sort of measuring devices would be needed?*

NOVEL CHALLENGE

A motorcycle horn emits a note of 400 Hz when stationary. If a motorcyclist approached a wall emitting a 400 Hz sound and this was reflected back, what pitch would a stationary observer hear (higher, lower, the same)? What would the cyclist hear? If the motorcycle was moving at 20 m s^{-1}, calculate both of these frequencies.

The new wavelength is this distance divided by the number of waves:

$$\frac{d - d_s}{n} = \frac{(v - v_s)n}{fn}$$
$$\lambda' = \frac{v - v_s}{f}$$

Since $v = f\lambda$
Then:

$$\lambda' = \frac{v - v_s}{f}$$
$$\frac{v}{f'} = \frac{v - v_s}{f}$$
$$f' = f\frac{v}{(v - v_s)}$$

For our previous example:

$$\begin{aligned} f' &= f\frac{v}{(v - v_s)} \\ &= 10 \times \frac{10}{10 - 5} \\ &= \frac{100}{5} \\ &= 20 \text{ Hz} \end{aligned}$$

This is a higher frequency.

Similarly for point B, as the source is moving away the new distance becomes $d + d_s$. Therefore:

$$f' = f\frac{v}{(v + v_s)}$$

A similar analysis can be carried out with the observer moving instead of the source, producing an apparent frequency of:

$$f' = f\frac{(v + v_o)}{v}$$

if the observer is moving towards the source, and:

$$f' = f\frac{(v - v_o)}{v}$$

if the observer is moving away from the source.

One equation can be used for all situations, using a positive or negative to take account of the relative motion of the source or observer. This equation is:

$$f' = f\frac{(v \pm v_0)}{(v \pm v_0)}$$

where f' is the apparent frequency; f is the frequency of waves produced by the stationary source; v is the velocity of the waves; v_s is the velocity of the source; v_0 is the velocity of the object.

Example

A super-train moving past a station at a speed of 180 km h^{-1} (50 m s^{-1}) sounds its whistle as it comes into the station. If the frequency of the whistle on a stationary train is 320 Hz, what would be the frequency heard by the station-master standing on the platform if: **(a)** the train was approaching the platform; **(b)** the train was moving away from the platform? (The velocity of sound in still air is 341 m s^{-1}.)

Solution

(a) The observer is stationary ($v_0 = 0$). The source is moving toward the observer, therefore:

$$f' = f\frac{(v \pm v_0)}{(v \pm v_s)}$$

$$f' = f\frac{v}{(v - v_s)}$$

$$= 320\ \frac{341}{341 - 50}$$

$$= 375\ \text{Hz}$$

(b) The observer is stationary ($v_0 = 0$). The source is moving away from the observer, therefore:

$$f' = f\frac{v}{(v + v_s)}$$

$$= 320\ \frac{341}{341 + 50}$$

$$= 279\ \text{Hz}$$

Rule:

Source moving, observer stationary When the source moves toward the observer, the frequency is greater (toward equals greater) which requires a negative (–) sign in the denominator. If the source moves away from the observer, a plus (+) is used in the denominator.

Observer moving, source stationary When the observer moves toward the source, the frequency is greater (toward equals greater) which requires a plus (+) sign in the numerator. If the observer moves away from the source, a negative (–) is used in the numerator.

Some practical examples of the Doppler effect

- Astronomers use the Doppler shift of light frequencies to measure speeds of distant galaxies.
- Physicians can detect heartbeats of a foetus by means of a Doppler shift of ultrasound.
- Police radar units use the Doppler effect to measure the speed of cars, baseballs and cricket balls.
- Soldiers can tell if a rocket is coming toward them or going away by listening to the Doppler shift. The loudness enables them to also estimate this distance.

NOVEL CHALLENGE

Subsonic bullets don't emit a loud 'crack' when fired. What advantage would this be? Do any hand guns have subsonic bullets?

Questions

25 State what a listener observes about the apparent frequency of a sound source when the following occurs. Complete a copy of Table 16.8 but do not write in this book.

Table 16.8

LISTENER	SOURCE	APPARENT CHANGE IN FREQUENCY
Still	Approaching	
Still	Receding	
Moving away	Stationary	
Moving toward	Stationary	

26 Students celebrating the finish of Year 12 drive along the street at a speed of 60 km h^{-1} while sounding a whistle that has a frequency of 1200 Hz. Other students standing on the side of the road hear the noise as the car approaches and goes away. What is the apparent frequency of the whistle: **(a)** as the car approaches; **(b)** as the car goes away? (The speed of sound is 330 m s^{-1}.)

27 A police car's siren emits sound waves of 1000 Hz. If this car is involved in a car chase and is travelling at 120 km h^{-1} what frequency will a person on the side of the road hear: **(a)** as the car is approaching; **(b)** as the car is going away? **(c)** What frequency will the person driving the pursued car hear? (Assume this car is travelling at the same speed as the police car.) (The speed of sound is 340 m s^{-1}.)

The sound barrier

The speed of sound — Mach 1 (named after Austrian physicist and philosopher Ernst Mach) — is approximately 1200 km h^{-1} at sea-level and decreases as altitude increases. It is about 1050 km h^{-1} at a height of 11 000 m.

When planes fly faster than the speed of sound they are said to break the sound barrier. As they approach the speed of sound, sound waves compress in front of the plane (Doppler effect), which results in the formation of a shock wave.

As the plane breaks through the sound barrier the shock wave is left behind within a 'noise cone'. Within this noise cone the waves emitted in a forward direction accumulate and constructively interfere to make a very large amplitude disturbance. When this noise cone reaches an observer a loud 'bang' is heard. This is known as a **sonic boom** and results in large acoustic pressures.

In 1947, the first experimental piloted aircraft to break the sound barrier was the Bell X-1 powered by a four-chambered liquid-rocket engine and launched in the stratosphere from the underbelly of a flying bomber.

Photo 16.4
The Concorde was the only supersonic commercial aircraft in operation. Notice its sleek design, which allows it to safely break the sound barrier.

As planes began to break the sound barrier in the 1940s their design changed. As planes approach the sound barrier the air surrounding the plane becomes 'harder' to fly through and the resistance increases greatly, making planes unstable. This resulted in the deaths of many test pilots in the 1940s when the ambition to break the sound barrier was paramount. Supersonic aircraft need to be much sleeker with pointed noses; for example, the Concorde, which could fly at 2000 km/h. Although the Concorde was supersonic, it was also expensive to operate and passenger confidence never recovered after its crash in France in 2000. The last flight of the 12 remaining Concordes was in 2003. New commercial airlines of today focus on fuel economy, quietness and automation, instead of speed. Greater safety, increased reliability, less noise and pollution, better passenger comfort, more navigational aids and less room for pilot error are all guidelines for the commercial airplanes of tomorrow. Supersonic planes are no longer on the drawing boards of any major manufacturer.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

(For all questions unless specified use v_{sound} = 340 m s^{-1}.)

***28** When sound waves travel through a medium, in which direction do the particles of the medium vibrate?

***29** A tuning fork produces 2.4×10^4 compressions and rarefactions in the air particles around it in 10 s. The distance between each compression is 0.14 m.
(a) Find the frequency of the tuning fork.
(b) What is the velocity of sound in air?

***30** In a thunder storm the lightning is seen before the thunder is heard, as the velocity of light is much greater than the velocity of sound. If the thunder is heard 10 s after the lightning is seen how far away is the storm? (The velocity of sound is 340 m s^{-1} at the current temperature.)

***31** Explain how sounds can have the same pitch but different qualities.

****32** Two speakers are placed 2 m apart and produce sound of the same frequency and in phase. A person walks across the front of the speakers as shown in Figure 16.24. In doing so she notes that the intensity of the sound goes down, then up again, then down again at point B 0.5 m from the centre point A. What is the wavelength of the sound emitted by the speakers?

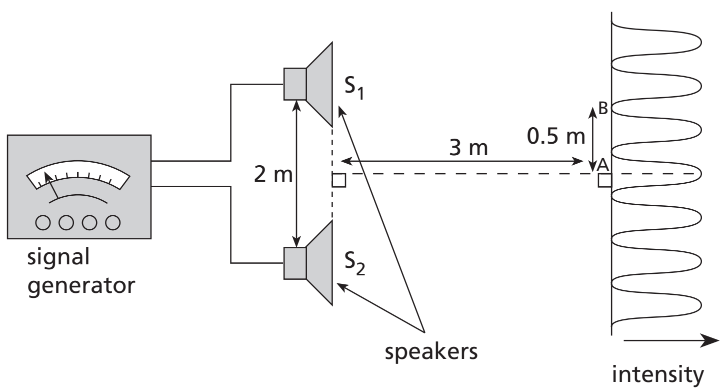


Figure 16.24
For question 32.

***33** A marine survey vessel plotting the contours of the ocean floor sends an ultrasonic wave and receives an echo back 1.2 s later. Calculate the depth of the ocean at this point. (The velocity of sound in sea water is 1400 m s^{-1}.)

***34** A student produces a note by blowing across the mouth of an open-ended piece of plastic pipe 0.2 m long. Calculate the frequency of the third harmonic.

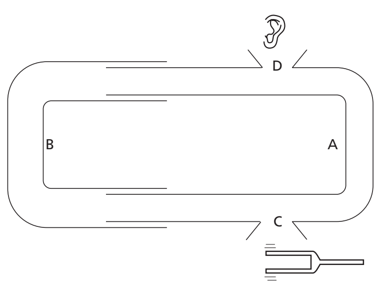
***35** A 40 cm organ pipe is open at both ends. If air is blown over one end what is the fundamental frequency emitted from this pipe?

****36** Open-ended and closed-ended pipes can produce the same fundamental frequency.
(a) Calculate the fundamental frequency produced by a closed-ended pipe of length 25 cm.
(b) Calculate the length of the open-ended pipe that would produce the same fundamental frequency.
(c) Even though they both produce the same fundamental frequency they would sound different. Explain with calculations why this occurs.

**37 A pipe is open at both ends and is 0.58 m long.
(a) Determine the wavelength of the sound that would produce the fundamental frequency in this pipe.
(b) Calculate the frequency of the second and third overtones.
(c) What harmonic is the fourth overtone?
(The speed of sound is 342 m s^{-1}.)

**38 One method of measuring the speed of sound is to stand a distance of, say, 300 m from a wall in an open area and bang two pieces of wood together while timing and listening for the echo. When the echo is heard bang the wood together again. Continue to do this, say, 10 times, and then stop the timing. Knowing the distance to the wall and the time, the velocity of sound can be calculated. If the distance to the wall is 300 m and the time measured from the first noise to the last echo was 17.5 s, calculate the velocity of sound.

Figure 16.25
For question 39.



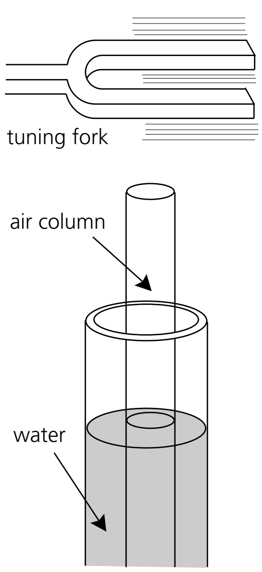
**39 The apparatus in Figure 16.25 can be used to find the speed of sound. Section A is fixed while section B is movable. A tuning fork of known frequency is sounded over one opening, C, and you listen over the other opening, D. If you make the lengths of A and B equal, a maximum of intensity is heard at D, but if B is slowly moved out a minimum of intensity, then a maximum, then another minimum will be heard.
(a) Explain why this happens.
(b) When B is moved out a distance of 21 cm the first minimum of intensity is heard. Calculate the speed of sound. (The tuning fork used had a frequency of 400 Hz.)

40 Draw diagrams to show the standing waves set up in a harp string of length *L* when the length of the string corresponds to **(a) two wavelengths; **(b)** three and a half wavelengths; **(c)** four wavelengths. **(d)** Calculate the frequency of the sound emitted from this string in each of part **(a)**, **(b)**, and **(c)**, when the length of the harp string is 0.60 m, the mass of the string is 20 g, and the string is under a tension of 120 N.

**41 Students experimenting with musical notes set up a row of test-tubes in a test-tube rack. They then fill them to various levels with water. By blowing across the top of the tubes they are able to create different notes. If the distance between the top of the water and the top of the test-tube for the first one is 8.0 cm:
(a) calculate the fundamental frequency emitted by this tube;
(b) describe what the students will need to do to create different frequencies, and calculate the distances from the top of the tubes to the water if they wish to create fundamental frequencies of **(i)** half that in part **(a)**; **(ii)** twice that in part **(a)**; **(iii)** three times that in part **(a)**.

**42 Discuss the possibility of open-ended organ pipes producing different frequency notes on hot or cold days.

Figure 16.26
For question 43.



Extension — complex, challenging and novel

***43 An open tube is placed into a container of water and a vibrating tuning fork placed over the mouth of the tube. (See Figure 16.26.) As the tube is raised so a greater length of the tube is out of the water, resonance is heard. This occurs when the distance from the top of the tube to the water level is 12 cm, and again at 50 cm. Determine the frequency of the tuning fork.

***44 To find the frequency of an unknown tuning fork (Z), two tuning forks (X and Y) of known frequency are used. X has a frequency of 245 Hz and Y has a frequency of 247 Hz. When Z is sounded with X, 30 beats are heard in 10 s. When Z is sounded with Y 10 beats are heard in 10 s. What is the frequency of Z?

*****45** A 442 Hz tuning fork is sounded at the same time as the A string of a guitar. A beat frequency of 50 beats per 10 s is heard. If a rubber band is wrapped tightly around one prong of the tuning fork and this is then sounded at the same time as the guitar string, a beat frequency of 30 beats in 10 s is heard. What is the frequency of the guitar string?

*****46** A ferry crossing the river at 10 km h^{-1} sounds its whistle as it approaches the jetty. Passengers on board the ferry hear two whistles — the whistle itself and its echo from the rock face behind the jetty. They appear to be different. What is the frequency of the reflected sound if the frequency of the whistle is 480 Hz and the speed of sound is 330 m s^{-1}?

*****47** In 1845 the Dutch meteorologist Christoph H. D. Buys Ballot first tested the Doppler effect by having two trumpet players play a musical note of 440 Hz, one on a moving flatcar of a train and the other on the station platform. While they were playing the note, a beat frequency of 3 beats per second was heard by a person at the station. How fast was the train moving?

*****48** To detect the location of cannons on a battlefield, the army used a method called triangulation. Several microphones would be placed in a straight line at 1 km intervals along the 'front line'. When the cannon noise was detected by the closest microphone, a timer would start and the delay for the second and subsequent microphones would be recorded on a paper roll. If three microphones 1 km apart were used and there was a time delay of 1.2 s to the second microphone and 3.7s to the third microphone, determine the position of the cannon (there may be more than one answer). A similar process is used to detect the epicentre of earthquakes.

Ch17-Walding 4th 25/8/04 12:01 PM Page 381

UNIT 06

Optics

CHAPTER 17

Reflection of Light

17.1 INTRODUCTION

Light has played an important part in the evolution of humans since the beginning of time. Light from the Sun has supplied the energy for plants to photosynthesise, thus producing plant growth and food for animals and humans. A by-product of this, if it can be called such, is the production of oxygen — a necessary ingredient for sustaining life on Earth.

Over the past two decades energy from sunlight has played an important part in the conservation of other forms of energy. The development of non-polluting forms of energy will add to our quality of life. Use of solar energy will play a major part in our energy needs in the future. Light energy from the Sun is used to provide energy to heat water in solar hot water systems, reducing the dependence on coal-burning electricity production. Light is used in the production of solar electricity — electricity used to provide energy for remote telephone boxes, to fuel cars that race experimentally, and for energy-conscious households of the future.

But light has a more important use — it allows us to see. It allows us to identify objects, see colours and in most cases to choose our partners.

Scientists have developed many devices that enhance our perception of the world around us, with the development of mirrors and lenses that allow us to see better — glasses; to see further — telescopes and binoculars; and to see finer detail — microscopes. This is the content of this chapter and the next three chapters: Optics — the study of light and devices that use light.

A study of mirrors and lenses will enable you to answer questions such as:

- Where is light energy being used today?
- How can we concentrate light energy to be able to use it?

17.2 THE PRODUCTION OF LIGHT

Light energy can be converted to other forms of energy and vice versa. Light energy from the Sun is converted into chemical energy stored in plants, as well as into heat energy, and electrical energy. Other forms of energy such as heat energy and electrical energy are converted into light energy, such as when a light is turned on. Objects that emit their own light energy such as light bulbs, light from a star, or even a hot flame are called **luminous** objects. But these are few, because we see most objects by the reflection of light. When a light source illuminates objects they reflect light to our eyes; these objects could not be seen in a dark room. These are **non-luminous** bodies.

Light travels at a speed of 3×10^8 m s^{-1} in a vacuum and, contrary to Newton's original proposal, at a slower speed in water, glass, or any other medium.

17.3 LIGHT RAYS

Contrary to the belief of small children and our early ancestors, we see objects because light from these objects travels to our eyes, and not the other way around. It is no use covering your eyes, the 'bogey man' will still be there and able to see you.

To determine the position of an object requires narrow beams of light, **light rays**, to reach your eye, preferably your two eyes. Your brain traces these rays back to where they appear to meet. This, your brain tells you, is where the light originates. The wider the base for triangulation the better the positioning of the object. We call this 'stereoscopic vision'.

EI Activity 17.1 DEPTH PERCEPTION

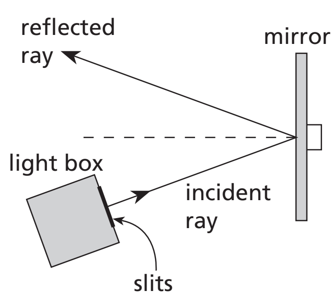
1. Ask a partner to hold up a finger within your reach.
2. Close one eye and try to touch your finger on the top of theirs.
3. How close did you get?
4. Try this again with both eyes open.
5. How close did you get this time?
6. Is one eye better than the other?

When only one eye is open, rays enter this eye at either side of the pupil, creating a narrow base and producing poor depth perception. With both eyes open, the base for triangulation is greater and so is the depth perception. Animals and fish that are hunters have eyes placed wide apart on the front of their heads so as to improve depth perception.

We use rays of light to determine where objects or images are. Rays of light travel in straight lines from objects. Importantly, rays from distant objects are close enough to be considered parallel.

Light boxes are common devices used to produce thin beams or rays of light for the investigation of optics in the laboratory (Figure 17.1).

Figure 17.1
Light boxes that produce thin beams of light (rays) are used in laboratory optical investigations.



A **laser** is another device used to produce thin beams of light. These also have the added convenience of emitting light of one wavelength. Laser stands for 'light amplification by stimulated emission of radiation'. Briefly, light is produced when the atoms of the laser medium are excited by electrical discharges or intense light flashes. When these atoms return to their unexcited state they give off energy in the form of light of a particular frequency and phase. Notice that you cannot see the light of a laser or even the light from a light box unless it strikes a wall or an object, as your eye and brain only respond to light when it strikes your eye. However, the following activity will allow you to see the laser beam without looking directly into it **as this is very dangerous and can damage your eye.**

EI Activity 17.2 LASER LIGHT

1. Place a laser at one end of the laboratory and turn it on so the beam strikes the wall at the other end of the room.
2. Hit a chalk-filled duster with a ruler around the area where the beam passes.
3. What do you notice?
4. Try to explain why this occurs.

PLANE MIRRORS 17.4

Plane mirrors normally consist of flat pieces of glass that have their backs coated with a thin layer of aluminium, and with lacquer to stop the aluminium from flaking. However, up until 1857 mirrors mainly consisted of highly polished pieces of steel. In that year Jean Foucault developed a method of silvering glass to make mirrors, thus producing a lighter and better-quality mirror than the common polished metals ones that tarnished. This brought about large advances in the development of astronomical telescopes.

Laws of reflection

What happens to a ray of light when it strikes a mirror? Everyone would say it is reflected. But in what way is it reflected?

Activity 17.3 LAWS OF REFLECTION

EI

1 Stand a mirror on a piece of white paper.
2 Using a light box with one slit, shine a ray onto the mirror at various angles.
3 In each case draw in the incident ray, the reflected ray and the position of the mirror.
4 Measure the angles between the normal to the mirror and the rays.
5 What do you notice?

We call:
- the ray that strikes the mirror the **incident ray**
- the ray that leaves the mirror the **reflected ray**
- the perpendicular to the mirror **the normal**
- the angle between the incident ray and the normal **the angle of incidence**
- the angle between the reflected ray and the normal **the angle of reflection**.

These are shown in Figure 17.2.

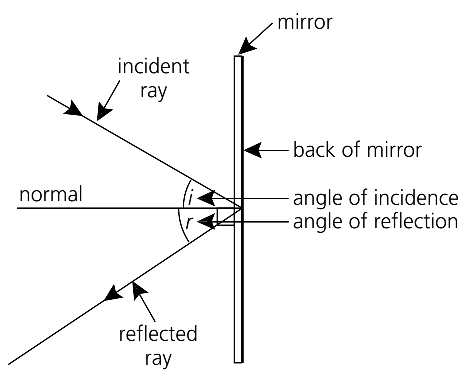


Figure 17.2
The common terms associated with rays of light and plane mirrors ($\angle i = \angle r$).

The above activity should have demonstrated the first law of reflection:

The angle of reflection is equal to the angle of incidence.

There is something else you may have noticed from the activity even though it may have been regarded as trivial. To obtain the angles in the activity, **the incident ray, the normal and the reflected ray all lie in the same plane**, that is, the plane of the paper.

If many parallel rays strike the mirror they leave the mirror parallel to each other, as shown in Figure 17.3.

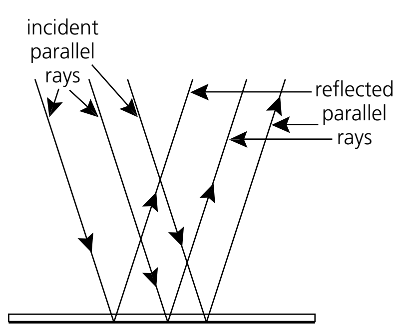


Figure 17.3
Specular reflection from a plane mirror. Parallel incident rays produce parallel reflected rays.

NOVEL CHALLENGE

You walk towards a plane mirror at 1 m s^{-1}. *How fast does your image approach you?* The mirror now approaches you at 1 m s^{-1}. *How fast does your image approach you now?*

Activity 17.4 REGULAR REFLECTION

1 Shine the light from a light box with a number of slits producing parallel light onto a plane mirror.

2 What do you notice about the rays that are reflected from the mirror?

This property is called **regular** or **specular reflection**. However, if these parallel rays are incident on an uneven surface such as a sheet of paper or a table top the reflected rays are not parallel.

Activity 17.5 DIFFUSE REFLECTION

1 Shine light from a light box with a number of slits onto a sheet of paper instead of a mirror.

2 What do you notice this time?

This is called **diffuse reflection**. It is not that the laws of reflection are being broken; it is just that the surface is uneven and the incident rays are not striking the surface, or parts of the surface, at the same angle (Figure 17.4). The angle of incidence still equals the angle of reflection for each ray at the particular point of contact.

NOVEL CHALLENGE

In the April 1984 edition of *New Scientist* magazine, a report appeared on the work of British inventor Charles deSelby. DeSelby reasoned that when you look at yourself in a mirror you are not seeing yourself at that instant in time but when you were a fraction of a second younger (the time it takes light to travel from your face to the mirror to your eyes). He set up two parallel plane mirrors facing each other and produced an enormous number of images as the light reflected back and forth. If you place your head between the mirrors you can verify this (even at home). Each successive image was further away in time than the one before. He used a telescope to peer at the receding images and he said he noticed that his image appeared successively younger until he finally noticed he looked like a young boy of particularly beautiful countenance. *Was deSelby a big liar or what? What is wrong with his theory?*

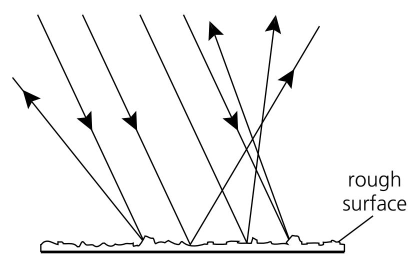


Figure 17.4
Diffuse reflection occurs from a rough surface. Parallel incident rays do not produce parallel reflected rays.

Because of this, images cannot be produced by surfaces that produce diffuse reflection.

Images

If you look into a mirror you can see the image of yourself but where is this image? The position of this can be found with a little investigation.

Activity 17.6 IMAGES

1 Place a sheet of paper on a styrofoam board.

2 Stand a mirror on this paper.

3 Place a pin in the paper a distance of about 10 cm from the mirror — this is the object.

4 Move your head to a position at an angle to the mirror and observe the image of the pin. Place two pins in line with this image. These represent a reflected ray.

5 Move your head to another position and place two more pins in line with the image — another reflected ray.

6 Mark the position of each pin and the mirror.

7 Draw up the reflected rays and extend them to where they meet. This will give you the position of the image (Figure 17.5).

8 What did you find?

You would have found that when this ray diagram is drawn and extended back (this is what your brain does), the rays appear to meet behind the mirror. Notice that they only appear to do so. This is called a **virtual image** as the rays do not pass through it and therefore the image could not be focused on a screen.

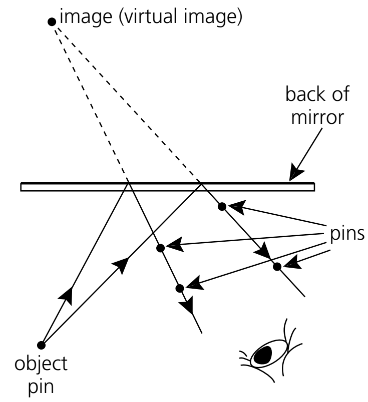
You may have also noticed that:

the image is the same distance behind the mirror as the object is in front

and

the line joining the object to the image is perpendicular to the mirror

Figure 17.5
Two rays are needed to determine the position of the image.

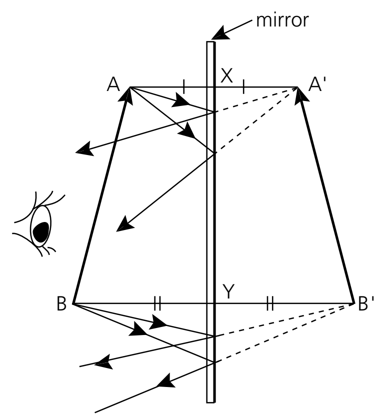


EI Activity 17.7 TRIANGULATION

This is just a short activity to demonstrate the importance of using a wide base of observation to locate images in mirrors.

1. Ask a student at the back of the classroom to speak and ask a student at the front, with eyes closed, to indicate where the speaker is.
2. Repeat the procedure but ask two students sitting close together at the front to indicate where the speaker is. (This gives a better indication of the source of the sound.)
3. Repeat the procedure but this time ask two students sitting a large distance apart at the front to indicate the source of the sound.
4. What did you find?

Figure 17.6
The image of an object can be determined using ray diagrams.



You get the idea — you need two rays, and the wider the base of observation the better and more accurate the placement of the source of the sound.

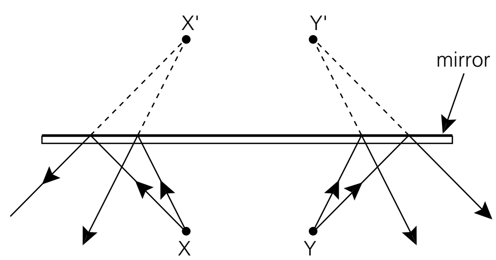
Ray diagrams, the principles of reflection, and triangulation can be used to diagrammatically show the relationship between objects and their mirror images.

Figure 17.6 shows the image of an arrow AB in a mirror. Using two rays from point A reflecting from the mirror to your eyes the image of A can be established. The same applies to B. Remember — to position an object two rays are needed and your brain, using the learnt fact that light travels in straight lines, traces the rays back to where they appear to come from.

Notice the arrow slopes the opposite way. This should have been predicted as it has already been established that the image of an object lies the same distance behind a mirror as the object does in front and is on the perpendicular to the mirror. Therefore $AX = XA'$, and $BY = YB'$. The same applies for all points between A and B. Notice also that the image $A'B'$ is the same size as AB.

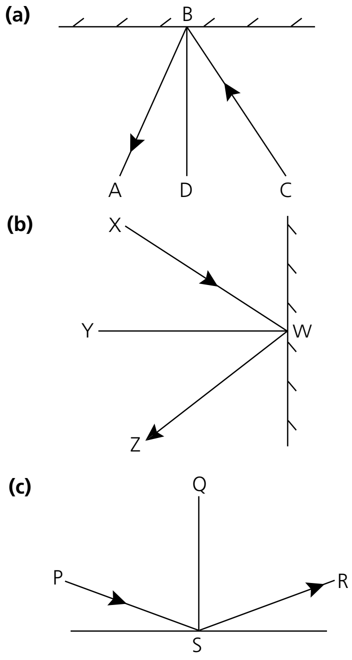
Another curious fact about images in plane mirrors is that the image is **laterally reversed**. This is seen when you observe yourself in a mirror. If you wink your right eye it is the image's left eye that does the winking — left and right are reversed. Again, this can be established by drawing ray diagrams (Figure 17.7). The object X is on your left and Y on your right. In the image, X′ is on the image's right.

Figure 17.7
A ray diagram shows lateral inversion in a plane mirror.



Plane mirrors have been used for many years in various ways; for example, as beauty aids, and in cameras. They were used by First World War soldiers in the trenches, and are used in submarine periscopes and at football matches.

Figure 17.8
For question 2.



People often think that a photograph of themselves is not very complimentary. This can be explained because they are used to seeing themselves in a mirror. The image they see of themselves has been laterally inverted, unlike the photograph.

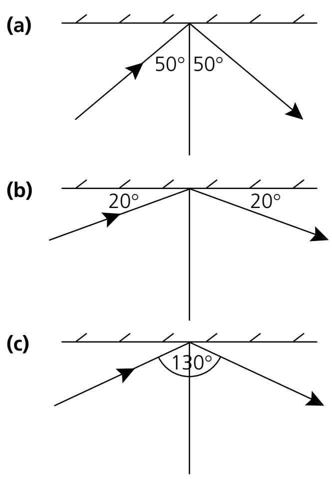
Questions

1. Give examples of luminous and non-luminous bodies.
2. For the three cases shown in Figure 17.8, state the angle of incidence, the angle of reflection, the incident ray, the reflected ray, and the normal.
3. For each of the cases shown in Figure 17.9, state the size of the angle of incidence.
4. A student walks toward the front door of the school building at night. If she approaches the doors at a speed of 2 m s^{-1}, at what speed will her image in the doors approach her?
5. If a plane mirror produces images that are laterally reversed, then explain with the aid of a diagram why periscopes do not produce images that are reversed.

Corner reflectors

Interesting reflections occur from plane mirrors when three plane mirrors are placed along the three axes (*x*, *y*, and *z*) like the internal corners of a cube (Figure 17.10). This is called a **corner reflector**. Corner reflectors reflect light directly back towards the source no matter what angle the light strikes the reflector from. 'Cats'-eyes' on many roadways are made of corner reflectors. They also have many applications in science; for example, a laser geodynamic satellite 'LAGEOS' has over 400 corner reflectors and is used to measure continental drift by bouncing laser beams from the satellite. Scientists have placed corner reflectors on the Moon to reflect laser pulses from the Earth and thus accurately measure the changes in the Earth–Moon distance.

Figure 17.9
For question 3

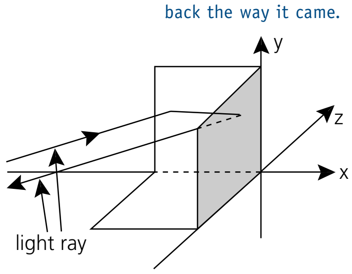


One-way mirrors

Everyone has seen police movies where a witness stands behind a 'one-way mirror'. The witness can see the criminal but the criminal cannot see the witness. How do these work? One-way mirrors rely on the lighting in the rooms. Just as you can very easily see your reflection when you try to look through a window on a dark night, the same occurs when a criminal tries to see through the one-way mirror. The room containing the 'line up' is very well lit while the observation room is dark. The criminal looking into the window sees his or her reflection and the glass appears to be a mirror. However, the witness in the dark room can observe the criminal, as light from the well-lit room passes through the glass. The 'one-way mirror' can be improved if a thin layer of metal is coated on the mirror's back surface. This improves the reflection properties but still allows enough light to be transmitted for observations to be made.

CURVED MIRRORS 17.5

Figure 17.10
A corner reflector, which consists of three perpendicular mirrors, reflects light back the way it came.



Curved mirrors are just as common today as plane mirrors but in many cases not as obvious. Can you identify some places where they are used?

A curved mirror can be one of two types, either **convex** or **concave**, depending on where the reflecting surface is. Curved mirrors are normally spherical mirrors, that is, they come from a part of a sphere. Imagine you had a hollow glass sphere. If you could take a section out of the sphere and silver either the outside or the inside you would have a spherical mirror. If the outside is coated this becomes the back of the mirror, producing a concave mirror. If the inside is silvered it becomes a convex mirror. Curved mirrors can make light rays converge (come together) or diverge (spread out), as shown in Figures 17.11 and 17.12.

Features of spherical mirrors

The centre of the sphere of which the mirror forms a part is called the **centre of curvature (C)**. The line through the centre of the mirror to the centre of curvature is the **principal axis**.

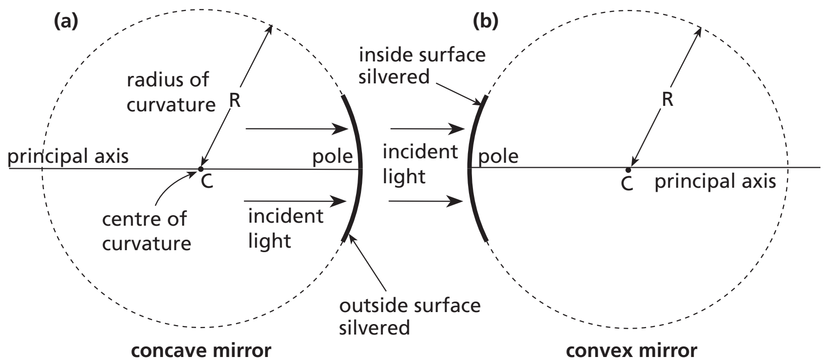


Figure 17.11 The spherical section becomes a concave **(a)** or convex **(b)** mirror, depending on which surface is silvered.

The point at which light rays parallel to the principal axis converge, in the case of a concave mirror (**converging** mirror), or appear to converge when extended back, in the case of a convex mirror (**diverging** mirror), is the **principal focus** (F). (See Figure 17.12.) This point could be found experimentally using a light box with multiple slits. In the case of the convex mirror these rays would have to be traced back to establish the focus — a virtual focus. In each case the focus could be found geometrically using the laws of reflection. (See Figure 17.13.)

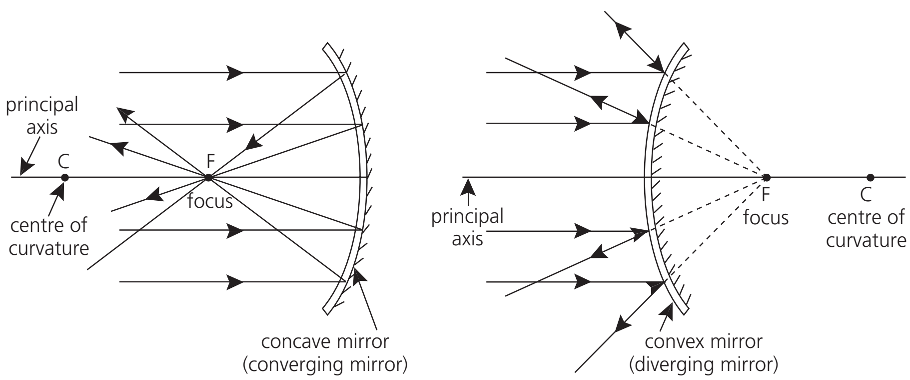


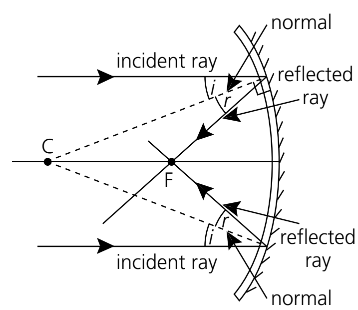
Figure 17.12 Common terms associated with spherical mirrors.

In either case the focal point is found to be half the distance from the mirror to the centre of curvature ($c = 2f$). The distance from the mirror to the focus is the **focal length** (f) of the mirror.

Measuring the focal length of a concave mirror

There are a few methods for determining the focal length of a concave mirror including finding the position of the image and then using the mirror formula. However, the simplest methods use the principle of 'parallel rays converge at the focal point'.

Figure 17.13 Using the law of reflection, $\angle i = \angle r$, with a number of rays, the focal point can be determined.



EI Activity 17.8 FOCAL LENGTH

1. Shine the light from a light box with a slide that contains three or four slits on to the concave mirror. The focal point is where these rays intersect after reflection.
2. Use the light from a distant object outside the window of the laboratory. Focus the light from this object onto a screen placed in front of the mirror and measure the distance from the mirror to the screen when a clear, distinct image falls on the screen.
3. How do the two methods compare?

Ch17-Walding 4th 25/8/04 12:01 PM Page 389

But how do you measure the focal distance of a convex mirror?

Images

The position of images seen in curved mirrors can be determined by drawing ray diagrams. To establish the position of the image requires drawing at least two rays. Any number of rays can be drawn using the laws of reflection, but these require the use of protractors to ensure the angle of incidence equals the angle of reflection. For this reason three easily drawn rays are normally used:

A A ray parallel to the principal axis reflects through the principal focus, or appears to come from this point in the case of a diverging mirror.

B The reverse of **A**. A ray through the focus reflects parallel to the principal axis.

C A ray through the centre of curvature reflects back through the centre of curvature.

(Remember — the centre of curvature is the centre of the sphere of which the mirror forms a part, therefore all rays from this point are perpendicular to the curved surface.)

Images — concave mirrors

The following examples are ray diagrams drawn to find the image of objects placed at various distances from a concave mirror.

Example 1 — object outside the centre of curvature Look at Figure 17.14. Notice that we only need to find the image of the head of the object as the foot is on the principal axis directly below the head; therefore the image of the foot will be directly below the head of the image.

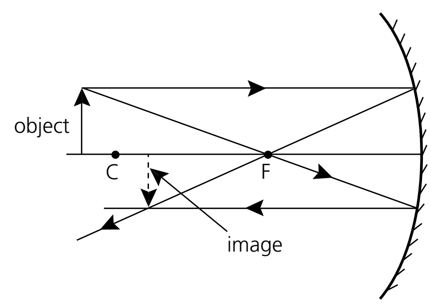


Figure 17.14
An object placed outside the centre of the curvature of a concave mirror produces an image that is smaller, real, inverted, and between F and C.

Example 2 — object at C See Figure 17.15.

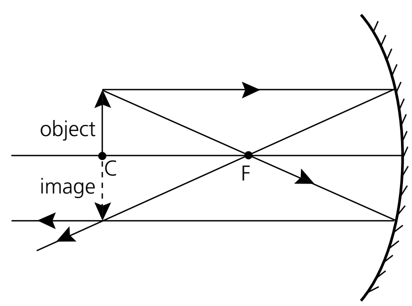


Figure 17.15
An object placed at the centre of the curvature of a concave mirror produces an image that is the same size, real, inverted and at C.

Example 3 — object between C and F See Figure 17.16.

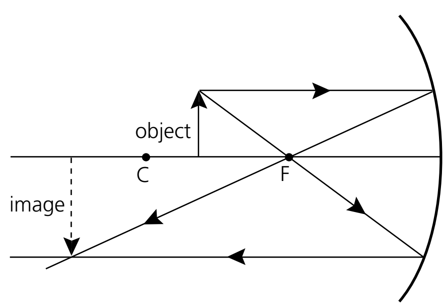


Figure 17.16
An object placed between the centre of curvature and the focal point of a concave mirror produces an image that is larger, real, inverted and outside C.

Example 4 — object between F and the mirror See Figure 17.17.

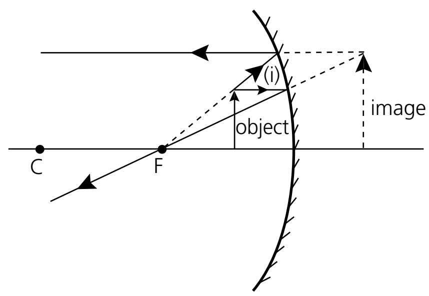


Figure 17.17
An object placed between the focal point and a concave mirror produces an image that is larger, virtual, upright and behind the mirror.

Notice that in Example 4, ray (i) does not actually pass through the focus but lines up with the focus. Also, to find the image the reflected rays have to be constructed back to where they appear to meet.

The characteristics of images are usually described using a set of common descriptors:

- size
- real or virtual
- upright or inverted
- position.

In the above examples for concave mirrors the characteristics of the images are as follows:
Example 1 — diminished (smaller); real (as the rays actually pass through the image, and because of this it can be focused on a screen); inverted; on the same side as the object between the focal point and the centre of curvature.
Example 2 — same size; real; inverted; at C on the same side.
Example 3 — magnified; real; inverted; outside C on the same side.
Example 4 — magnified; virtual; upright; behind the mirror.

Images — convex mirrors

Rays similar to those drawn for concave mirrors can be drawn to find the images in convex mirrors.

Example 1 — object a long distance from the mirror See Figure 17.18.

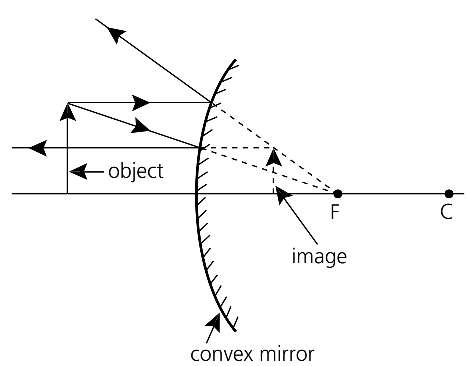


Figure 17.18
An object placed a long distance from a convex mirror produces an image that is smaller, virtual, upright, and behind the mirror.

Note 1: C and F are behind the mirror. (As C is the centre of the sphere, it has to be behind the mirror and F is half-way between C and the mirror.)
Note 2: The rays have to be traced back to where they appear to come from.
Description — diminished; virtual; upright; and behind the mirror between F and the mirror.

Example 2 — object close to the mirror See Figure 17.19.
Description — diminished; virtual; upright; and behind the mirror between F and the mirror.

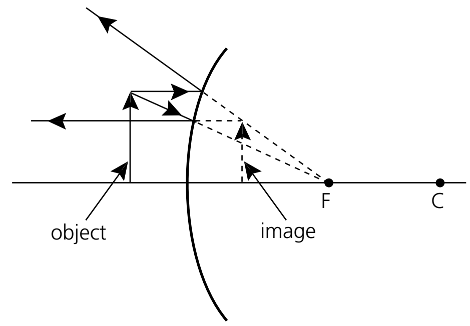


Figure 17.19
An object placed close to a convex mirror produces an image that is smaller, virtual, upright, and behind the mirror.

Notes on images

- In each of the above cases when virtual images are formed they are always upright.
- Where real images are formed they are always inverted.
- Convex mirrors always form virtual images.

Magnification

Magnification (M**)** is the size of the image compared with the size of the object.

$$M = \frac{H_i}{H_o}$$

where H_i is the height of the image; H_o is the height of the object.

Example 1

An object of height 1.0 cm is placed 6.0 cm in front of a concave mirror of 4.0 cm focal length.

(a) Draw an accurate ray diagram to locate the image.
(b) Describe the image.
(c) Find the magnification.

Solution

(a) See Figure 17.20.
(b) Magnified, real, inverted, outside C on the same side as the object.

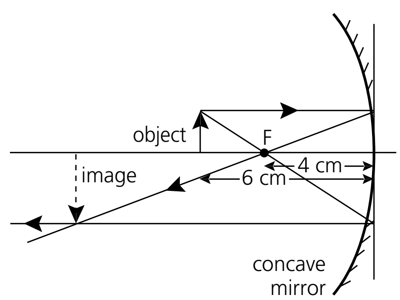


Figure 17.20
For sample problem.

(c)

$$M = \frac{H_i}{H_o} = \frac{1.8\text{ cm}}{1.0\text{ cm}} = 1.8$$

Note: to make drawing of ray diagrams easier, that is, without the use of compasses to draw the mirror, we can use a vertical line through the back of the mirror to represent the reflecting surface. This is a reasonable approximation as long as we stay within the middle part of the mirror.

Example 2

A convex mirror has a focal length of 6.0 cm. An object of height 1.5 cm is placed 2.0 cm in front of the mirror.

(a) Draw an accurate ray diagram to find the position of the image.
(b) Describe the image.
(c) Find the magnification.

Solution

(a) See Figure 17.21.
(b) Diminished, virtual, upright, behind the mirror between the mirror and F.

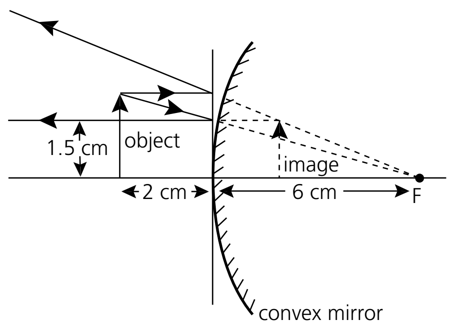


Figure 17.21
For sample problem.

(c)

$$M = \frac{H_i}{H_o}$$

$$= \frac{1.0 \text{ cm}}{1.5 \text{ cm}}$$

$$= 0.67$$

Questions

6 State as many differences as you can between concave and convex mirrors.

7 Which of the two mirrors, concave or convex, **(a)** spreads parallel light out, or diverges the light; **(b)** focuses parallel light to a point, or converges the light?

8 Use a ray diagram and an appropriate scale to find the position of the image of a 10 cm high object placed 1.5 m in front of a concave mirror of focal length 20 cm.

9 An object of 2.0 cm height is placed 5.0 cm in front of a diverging mirror of 10 cm focal length.

(a) Draw a ray diagram to find the position of the image.
(b) Describe the characteristics of the image.
(c) Use the magnification formula to find the height of the image.

17.6 THE MIRROR FORMULA

We need a more accurate method of finding the position of an image without drawing ray diagrams, which are not very accurate. We can do this by using a formula.

Consider the following derivation and refer to Figure 17.22.

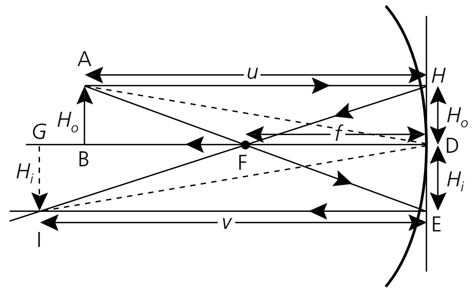


Figure 17.22
Used for the derivation of the mirror formula.

$$\triangle ABD \simeq \triangle IGD$$

$$\therefore \frac{H_o}{H_i} = \frac{u}{v}$$

$$\triangle ABF \simeq \triangle EDF$$

$$\therefore \frac{H_o}{H_i} = \frac{BF}{DF}$$

$$= \frac{u-f}{f}$$

$$\triangle HDF \equiv \triangle FGI$$

$$\therefore \frac{H_o}{H_i} = \frac{DF}{GF}$$

$$= \frac{f}{v-f}$$

$$\therefore \frac{u-f}{f} = \frac{f}{v-f}$$

$$(u-f) \times (v-f) = f^2$$

$$vu - uf - vf + f^2 = f^2$$

$$vu = vf + uf$$

$$vu = (v+u)f$$

$$\frac{1}{f} = \frac{(v+u)}{vu}$$

$$= \frac{v}{vu} + \frac{u}{vu}$$

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

This formula, $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$, relates the focal length f to the object distance u and the image distance v and is called the **mirror formula**.

The equation $M = \frac{H_i}{H_0} = \frac{v}{u}$ is used to find the magnification.

Note: since we have measurements on either side of the mirror, in front and behind, an order convention is required. We will make all measurements on the object side of the mirror positive and those behind the mirror (taken as the origin) negative. Hence concave mirrors have a positive focal length, and convex mirrors have a negative focal length.

To remember that u represents the object distance and v the image distance, recall that u comes before v in the alphabet and that light goes to the object before the image.

When using the magnification formula ($M = \frac{v}{u}$), the absolute valves of v and u should be used. That is, ignore + and − values.

Now let's see how accurate we were in drawing ray diagrams.

Example 1

An object of height 1.0 cm is placed 6.0 cm in front of a concave mirror of focal length 4.0 cm.

(a) Find the position of the image.
(b) Find the magnification and the height of the image.

Solution

(a) Using the formula, the mirror is concave, therefore $f = +4.0$ cm.

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

$$\frac{1}{4} = \frac{1}{v} + \frac{1}{6}$$

$$\frac{1}{4} - \frac{1}{6} = \frac{1}{v}$$

$$\frac{6}{24} - \frac{4}{24} = \frac{1}{v}$$

$$v = 12 \text{ cm}$$

The image is 12 cm in front of the mirror.

(b)

$$M = \frac{H_i}{H_o} = \frac{v}{u}$$

$$= \frac{12}{6}$$

$$= 2$$

$$\therefore H_i = 2H_o$$

$$= 2 \times 1$$

$$= 2 \text{ cm}$$

Example 2

A convex mirror has a focal length of 6.0 cm. An object of height 1.5 cm is placed 2.0 cm in front of the mirror.

(a) Find the position of the image.
(b) Find the magnification and the height of the image.

Solution

(a) The focal length is negative for convex mirrors as it is on the opposite side of the mirror as the object, hence $f = -6$ cm.

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

$$\frac{1}{-6} = \frac{1}{v} + \frac{1}{2}$$

$$\frac{1}{-6} - \frac{1}{2} = \frac{1}{v}$$

$$\frac{-1}{6} - \frac{3}{6} = \frac{1}{v}$$

$$\frac{-4}{6} = \frac{1}{v}$$

$$v = -1.5 \text{ cm}$$

Ch17-Walding 4th 25/8/04 12:01 PM Page 395

The image is 1.5 cm behind the mirror.

(b)

$$
\begin{aligned}
M &= \frac{H_i}{H_o} = \frac{v}{u} \\
&= \frac{1.5}{2} \\
&= 0.75 \\
\therefore\ H_i &= 0.75H_o \\
&= 0.75 \times 1.5 \\
&= 1.125\ \text{cm}
\end{aligned}
$$

Note: the image distance for real images is positive whereas the image distance for virtual images is negative.

Questions

10 A small light bulb is placed 20 cm in front of a concave mirror of focal length 15 cm.
- **(a)** Calculate the image distance.
- **(b)** Calculate the ratio of the width of the image of the bulb to the width of the actual bulb.
- **(c)** If the bulb was 1 cm across how wide would the image be?
- **(d)** What type of image is produced?

11 Students performing experiments with diverging mirrors try to locate the image of a small candle of height 2 cm in the mirror. When the candle is placed 25 cm from the mirror they see the image in the mirror to be smaller. (The focal length of the mirror is 20 cm.)
- **(a)** Draw a ray diagram to locate the image.
- **(b)** Use the mirror formula to calculate the position of the image.
- **(c)** What is the height of the image?

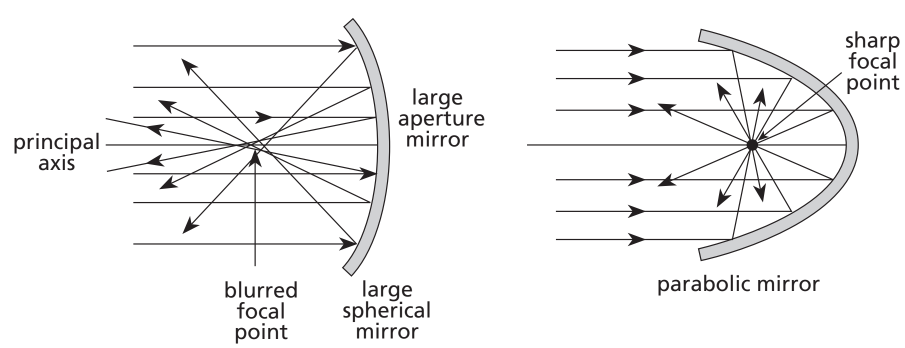
12 An object of height 2 cm is placed 4 cm in front of a diverging mirror of focal length 6 cm.
- **(a)** Draw a ray diagram to find the position and height of the image.
- **(b)** Use the mirror and magnification formulae to verify your answer to part **(a)**.

13 A concave mirror of focal length 10 cm is used to produce an image on a screen that is half the size of the object.
- **(a)** Find the position of the object and the image.
- **(b)** If you wanted to produce a real image of twice the size where would be the position of the object and the image?

SPHERICAL ABERRATION 17.7

Spherical aberration is the inability of a concave mirror to focus parallel light to a point. Parallel rays after reflecting from the mirror do not meet at a point but over a small region,

Figure 17.23
Parabolic mirrors help to eliminate spherical aberration where a blurred focal point is produced.



producing a blurred focal point rather than a sharp point. This defect of curved mirrors is called spherical aberration. It occurs more often when the mirror is large. It can be overcome by using smaller aperture mirrors or by only using the central region of larger mirrors. Special larger parabolic mirrors whose geometry results in the sharp focusing of parallel light are also manufactured to overcome this defect (Figure 17.23).

17.8 USES OF CURVED MIRRORS

Concave mirrors

- Concave mirrors are also used to produce magnified images so as to observe more detail in objects. Recall that if an object is inside the focal length then it produces an upright magnified image. Larger focal length concave mirrors are therefore used as shaving and make-up mirrors. Small concave mirrors are used by dentists.
- Concave mirrors are used as the reflectors in a number of applications where parallel or almost parallel light is required. If the light source is placed at the focal point of the mirror almost parallel light will be produced; for example, reflectors in headlights of cars, torches, and searchlights.
- Concave mirrors, because they bring together (or focus) light rays, are used to collect light energy as well as other forms of energy. Solar furnaces or ovens use large concave mirrors to concentrate light energy from the Sun onto pots and kettles placed at the focus. The biggest solar furnace in the world is located in the Pyrenees mountains in southern France. An array of computer-controlled plane mirrors (heliostats) on a nearby hill track the Sun and reflect the light onto an eight-storey-high converging mirror. The converging mirror focuses the sunlight onto a small building housing the solar furnace. Temperatures in excess of 3000°C have been reached in this experimental furnace.

  Concave mirrors are also used to concentrate other electromagnetic radiations such as radio and TV waves. Satellite dishes concentrate TV waves to be used by TV sets. Radio waves from stars can be focused onto the antennae placed at the focus of the receiving dish, which conducts the signal to an amplifier. (See Photo 17.2.)
- The same principle applies to astronomical telescopes. Larger optical reflecting telescopes concentrate visible light energy to the focus, where the eye piece or photographic equipment is placed. (Refer to Chapter 20, Optical Instruments.)
- Interestingly enough, because sound is so important to bats, bats' ears are concave in shape to collect and concentrate sound energy.

Photo 17.1
Torches use concave mirrors to produce a nearly parallel beam of light.

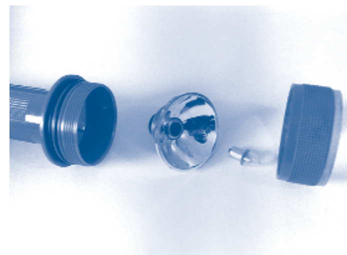


Photo 17.2
Satellite dishes enable TV sets to pick up TV waves that are reflected from satellites by concentrating TV waves on to an aerial.

Photo 17.3
Bats' ears are concave to collect reflected ultrasound waves.

Convex mirrors

Convex mirrors, because they have a wider field of view than plane mirrors and produce upright images, are used as rear-vision mirrors in cars; however, they have a disadvantage in that distances can be misjudged. They are also used on intersections of streets where vision is obscured. This allows drivers to see around 'blind' corners.

Because of their wide field of view they are also used in shops for security purposes.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***14** White paper or the desk top does not produce an image of an object. Does this mean the laws of reflection are not true for these surfaces? Explain!

***15** If you observe yourself in a plane mirror what are three significant optical facts about the image?

***16** One method of taking your own photo is to photograph your image in a large mirror. If you are standing 2.0 m in front of the mirror what distance setting should you use to obtain a photo in focus? Would an autofocus camera focus correctly?

***17** If you wish to view your whole body in a mirror what is the minimum length of such a mirror and where on the wall should it be placed?

***18** Explain the difference between a real and a virtual image.

***19** In each of the following cases state whether the light is parallel, converging, or diverging:
- **(a)** Light from a light bulb.
- **(b)** Light from a light bulb reflected from a mirror.
- **(c)** Light from a star.
- **(d)** Light from a star reflected from a concave mirror.
- **(e)** Light from a star reflected from a plane mirror.
- **(f)** Light from a star reflected from a convex mirror.
- **(g)** Light from a bulb placed at the focal point of a concave mirror.
- **(h)** Light from a bulb placed at the focal point of a convex mirror.

****20** A 1.0 cm high object is placed 10 cm in front of a concave mirror of 7.0 cm focal length.
- **(a)** Draw a ray diagram to find the position of the image.
- **(b)** Describe the nature of the image.
- **(c)** Use the mirror formula to find the exact position of the image.
- **(d)** Use the magnification formula to find the height of the image.

****21** A dentist wishes to use a concave mirror to view a patient's teeth. If he wants the image to be twice as large as the object and upright when the mirror is placed 2.0 cm from the teeth, what focal length mirror is needed?

****22** Students doing experiments with convex mirrors place a small candle 15 cm in front of the mirror. They are able to see a smaller image of the candle in the mirror but are unable to produce an image on a screen.
- **(a)** Explain why an image could not be formed on the screen. (Use a ray diagram.)
- **(b)** What sort of image is it?
- **(c)** If the focal length of the mirror was 20 cm, calculate the position of the image.

***23** Large diverging mirrors are used for security purposes in small shops. They are placed in the corners. What are the advantages of these mirrors for this purpose?

****24** A student places a 5.0 cm high candle 20 cm in front of a concave mirror. The image is formed on a screen placed 40 cm from the mirror.
- **(a)** What is the focal length of the mirror?
- **(b)** What is the size of the image?
- **(c)** Is the image virtual or real?
- **(d)** Is the image upright or inverted?

***25** You are unable to see your reflection in a dull, unpolished car; however, if the car is polished you can see your image. You have not changed the material structure of the car so why does this occur?

***26** When you go for your driver's licence you are required to read an eye chart. Issuing centres use plane mirrors to make sure the correct distance between the chart and the observer is maintained. If a distance of 6.0 m is required describe characteristics of the physical set-up of the chart and mirror as well as the chart itself to obtain this situation.

***27** Why do some police cars and ambulances have the words on their fronts as shown in Figure 17.24?

****28** Draw a reasonably accurate ray diagram to find the image of objects in the three cases shown in Figure 17.25.

Figure 17.24
For question 27.

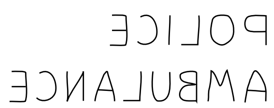
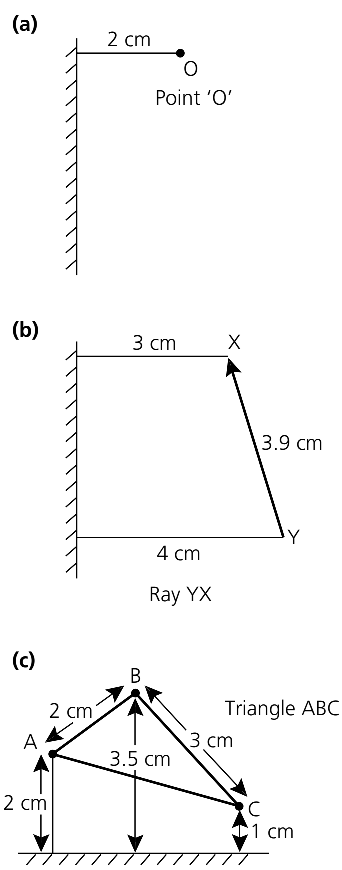


Figure 17.25
For question 28.



**29 Figure 17.26 shows the position of two mirrors at 45° to each other. A ball is placed at point X. Draw ray diagrams to find the image of the ball in the mirrors.

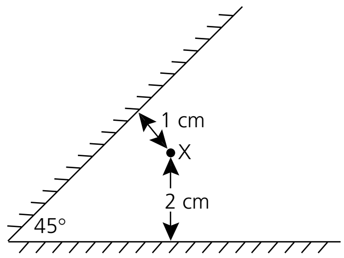


Figure 17.26
For question 29.

**30 Figure 17.27 indicates a light ray making an angle of 40° with the mirror. If the mirror is rotated so the ray now makes an angle of 10° with the mirror, through what angle does the reflected ray move?

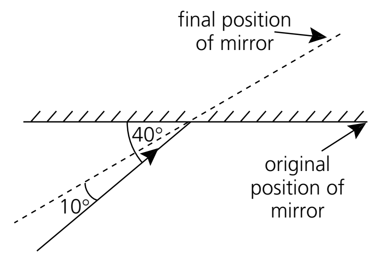


Figure 17.27
For question 30.

*31 Figure 17.28 shows a satellite dish used for receiving TV signals.
(a) What shape should the dish be?
(b) Where should the signal detector (aerial) be situated? Why?
(c) What features of the dish should be changed to improve its performance?

*32 Construct a table for concave mirrors showing the position of the object from outside the centre of curvature to inside the focus, the size of the image, the position of the image, the nature of the image, and whether it is upright or inverted.

*33 If you were sitting at the breakfast table and a piece of dust flew into your eye, explain how you could use a spoon to observe a larger image of your eye to help remove the dust particle.

*34 Some solar hot water systems use curved enclosures/supports to hold black PVC pipe in which water flows. (See Figure 17.29.) Discuss the purpose of these shiny curved enclosures and the positioning of the black PVC pipe.

*35 Figure 17.30 (opposite) shows an object with several rays reflecting from a convex mirror. Which of the rays are correctly drawn?

*36 Students going on a hike from an outdoor education centre decide to heat their food by making a solar cooker out of sheets of aluminium foil.
(a) What shape should they press the foil into?
(b) Where should the food be placed?

**37 In a laboratory experiment, students measured the position of the image of a small 4.0 cm high candle with respect to a concave mirror, and the magnification of the candle by measuring the size of the real image produced on a screen. Table 17.1 shows the results obtained.

Figure 17.28
For question 31.

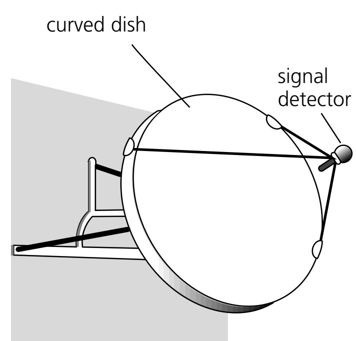


Figure 17.29
For question 34.

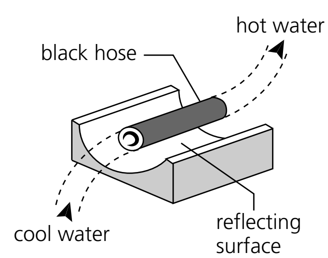


Table 17.1

Image distance (cm)	60	37.5	26.3	24	21.4	20	18.5
Image height (cm)	12	6	3	2.4	1.7	1.3	1.0
Magnification							

(a) Complete the table.
(b) Plot the graph of magnification verses image distance.
(c) Find the magnification when the image distance is 45 cm.
(d) Find the image distance when the magnification is 0.50.
(e) *From the graph*, find the focal length of the mirror.

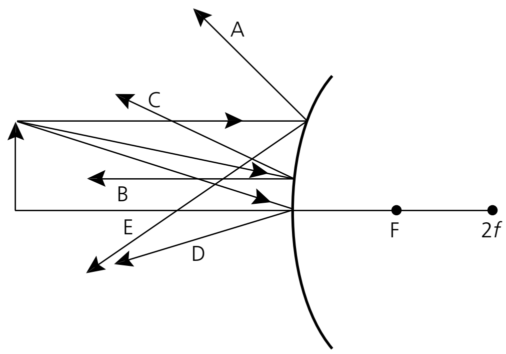


Figure 17.30
For question 35.

Extension — complex, challenging and novel

*****38** A plane mirror and a convex mirror are placed facing each other and 50 cm apart. A candle is placed on the principal axis 20 cm from the plane mirror, as shown in Figure 17.31. If the distance between the two images in the plane mirror is 40 cm, calculate the focal length of the convex mirror.

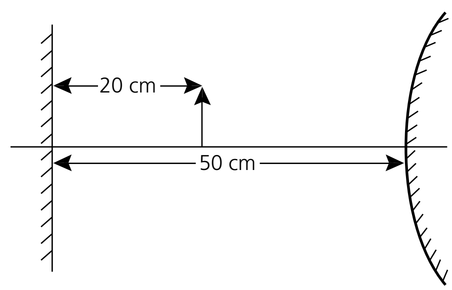


Figure 17.31
For question 38.

*****39** An object is placed 20 cm in front of a convex mirror of focal length 30 cm. A plane mirror is placed between the object and the mirror so that the image of the top half of the object in the convex mirror and the bottom half of the object in the plane mirror coincide. What distance is the plane mirror from the convex mirror?

*****40** A candle is placed in front of a concave mirror whose focal length is 20 cm. Find the position of the object and the image if **(a)** a virtual image of twice the size of the object is produced; **(b)** a real image of twice the size of the object is produced.

*****41** Students determining the focal length of a concave mirror obtained the measurements listed in Table 17.2 for the distances of the object and the image formed on the screen. Plot the graph of image distance against object distance to determine the focal length of the mirror.

Table 17.2

Object distance (cm)	60	50	40	30	25	20	15	10
Image distance (cm)	9.2	9.5	10	10.9	11.8	13.3	17.1	40

CHAPTER 18

Refraction

18.1 INTRODUCTION

Refraction is a property of waves that has been with us since time began. The refraction of light waves and radio waves through the universe has gone on since the beginning of the universe. Early hunters and spearfishermen used the refractive properties of water to accurately spear fish. However, the use of refraction, particularly in the fields of communication and medicine, has increased beyond belief over the past two decades. The development of optical fibres has revolutionised the way we receive telephone calls and TV programs (pay TV), and has reduced the time we spend in hospital with exploratory surgery.

- But what is refraction?
- How do optical fibres rely on refraction?
- Why are optical fibres in so much demand today?
- There are many uses made of optical fibres — can you name a few?

Questions related to everyday phenomena and which can be explained by refraction include these:

- Why do swimming pools or clear mountain streams seem shallower than they are?
- What causes mirages?
- Did you know a rainbow can only be observed when the rain is in front of you and the Sun is behind you?
- Did you know that the 'glass frogs' of Central America, whose bodies are so transparent you can see their insides, use refraction to vanish from sight when they slip into the water?

These questions and many more odd characteristics of the way light travels can be answered by a study of the refraction of light. By the end of this chapter many interesting phenomena will be able to be discussed with a knowledge of refraction.

18.2 REFRACTION

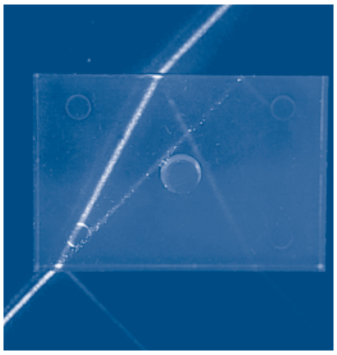
Recall the definition of refraction — it is the changing in direction of waves as they go from one medium to another. For water waves this meant that the direction of propagation of waves changed when they travelled from one depth of water to another. For light, refraction occurs when light passes from one medium to another, such as when light rays pass from air to water, from air to glass or from glass to water. This direction change can be easily observed in the case of light — the light rays themselves bend at the boundary between the media. A definition of refraction for light thus becomes:

Refraction is the bending of light rays at the boundary or interface, as they go from one medium to another.

EI Activity 18.1 REFRACTION

1. Use a light box with a narrow aperture to produce a single light ray.
2. Shine this ray at an angle other than 90° onto a block of glass.
3. What happens to the ray as it goes from air to glass?

Photo 18.1
The refraction of light as it passes through a block of glass.



You should have noticed that the ray bent as it entered the glass and bent again on exiting from the other side. (See Photo 18.1.) It has been refracted twice. If the normal had been drawn to the surface of the glass at the point the ray entered the glass, it would have also been noticed that the ray bent closer to the normal in the glass on entering and further from the normal on exiting.

We call the ray that strikes the glass **the incident ray**, and the ray that bends in the glass the **refracted ray**. The angle between the incident ray and the normal is **the angle of incidence** (**i**) and the angle between the refracted ray and the normal is **the angle of refraction** (**r**) (Figure 18.1).

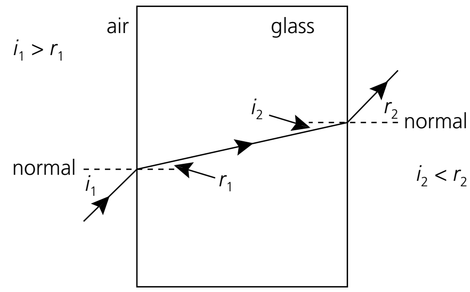


Figure 18.1
Light is refracted as it passes through glass, bending towards the normal in the more dense medium, and away in the less dense medium.

At the second surface, the surface where the ray passes from glass to air, the ray in the air is the refracted ray and the ray in the glass is the incident ray.

If Perspex is used instead of glass, similar refraction occurs except that for the same angle of incidence the angle of refraction will be different.

Similar effects are observed using any transparent material. The refracted ray bends towards the normal when the light travels from air to the material. The amount the rays bend depends on the **optical density** of the material. Optical density has nothing to do with physical density — mass, volume, etc. — but with the ability of light to pass through it.

A general rule is that **light rays bend toward the normal when they go from a less optically dense medium to a more optically dense medium**. The reverse is also true — **light rays bend away from the normal as they pass from a more optically dense medium to a less optically dense medium**. This illustrates the reversibility properties of light rays through a refractive system.

The amount of refraction that occurs results from the changing speed of light as it goes from air to the medium. (This was shown, using water waves, in Chapter 14.) Light travels faster in a vacuum or in air than in glass, water, etc. This is shown in Table 18.1.

PHYSICS FACT

Diamonds have one of the highest refractive indices of all substances. The pioneer in making fake diamonds was teenager Carroll Chatham, who in 1926 dissolved some graphite in molten iron and dropped the fiery solution into a vat of liquid nitrogen (−196°C). Instead of getting diamonds, the explosion blew out the windows in his San Francisco home. Later he had success and his son Tom now produces synthetic diamonds that have similar RI to the real ones.

Table 18.1 THE VELOCITY OF LIGHT IN VARIOUS MEDIA OF DIFFERENT REFRACTIVE INDICES

MEDIUM	VELOCITY OF LIGHT IN THE MEDIUM, v (10^8 m s^{-1})	ABSOLUTE REFRACTIVE INDEX OF THE MATERIAL, n
Air	3.00	1.00
Ice	2.31	1.30
Water	2.26	1.33
Ethyl alcohol	2.21	1.36
Fused quartz	2.05	1.46
Perspex	2.00	1.49
Benzene	2.00	1.50
Crown glass	1.97	1.52
Light flint glass	1.90	1.58
Heavy flint glass	1.82	1.65
Zircon	1.58	1.90
Diamond	1.24	2.42

The effect can be explained by use of the analogy of a car hitting a flooded section of a road. As one of the car's front wheels hits the water it slows down while the other wheel keeps going at the original speed. Therefore the direction of the car changes (if allowed). It bends into the water. (See Figure 18.2.) The car's direction changes. The new direction of the car will be closer to the normal and it will slow down.

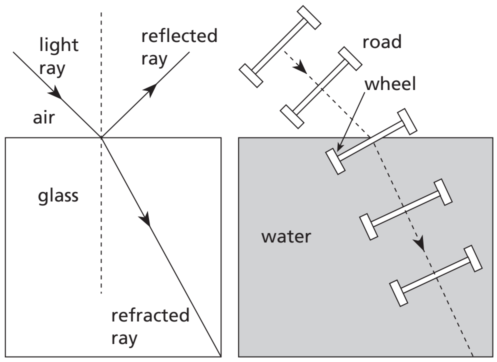


Figure 18.2
As the wheels of the car enter the water they slow down and swerve towards the normal.

The ratio of the velocity of light in air to the velocity of light in a different medium — water, glass, etc. — is constant. This constant is called the **absolute refractive index** of the material and is denoted by the symbol n. That is:

$$\frac{v_a}{v_m} = n$$

where v_a is the velocity of light in air; v_m is the velocity of light in the medium.

The refractive indices of several common materials are shown in Table 18.1. Notice that since n is a ratio it has no units.

Questions

1. Use Table 18.1 to see if you obtain the correct refractive index of the material by dividing the velocity of light in air by the velocity of light in the material.
2. Calculate the index of refraction for light going from air to a material in which its speed is **(a)** 2.6×10^8 m s^{-1}; **(b)** 1.8×10^8 m s^{-1}; **(c)** 3.4×10^8 m s^{-1}. (Is answer **(c)** possible? Explain!)
3. Calculate the speed of light in a medium whose refractive index is **(a)** 1.5; **(b)** 2.4; **(c)** 1.3.

18.3 SNELL'S LAW

In 1621 a Dutch mathematician, **Willebrod Snell** (1591–1626), discovered that the refractive index of a substance can be found using the angles of incidence and refraction. He found that if the angle of incidence was changed, the angle of refraction also changed in such a way that the ratio of the sine of the angle of incidence to the sine of the angle of refraction is always a constant for a particular material. This constant is the absolute refractive index of the material.

$$\frac{\sin i}{\sin r} = n$$

This is known as **Snell's law**.

For example, for a ray of light entering a block of glass, the ratio $\frac{\sin i}{\sin r} = 1.5$ for all values of i. Therefore the refractive index of glass $n_{a\text{-}g}$ or just $n_g = 1.5$.

The refractive indices given in Table 18.1 are the **absolute refractive indices**. They are the refractive indices obtained when a light ray travels from air to the material. Knowing the value of the absolute refractive index and the angle of incidence, the angle of refraction can be determined.

Example

Light from a light box is shone onto a block of Perspex at an angle of 30° to the normal. Determine the angle of refraction.

Solution

- $n_{\text{Perspex}} = 1.4$

$$\frac{\sin i}{\sin r} = n = 1.4$$

$$\sin r = \frac{\sin i}{1.4}$$

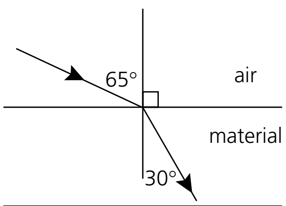
$$= \frac{\sin 30°}{1.4}$$

$$= 0.357$$

$$r = 21°$$

Questions

Figure 18.3
For question 4.



4 Use Figure 18.3 to calculate the refractive index of the material.

5 Complete Table 18.2 for light rays passing from air to a different medium.

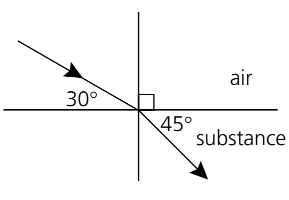
Table 18.2

	A	B	C	D
Angle of incidence, *i* (degrees)	30	60		0
Angle of refraction, *r* (degrees)	20		40	
Refractive index, *n*		1.8	1.4	

(Had a problem with D?)

6 A light ray travels from air to a substance as shown in Figure 18.4. Find the refractive index of the substance.

Figure 18.4
For question 6.



7 Students conducting experiments to find the absolute refractive index of a piece of Perspex obtained the results shown in Table 18.3 for the angles of incidence and refraction.

Table 18.3

Angle of incidence, *i* (degrees)	10	20	30	40	50	60
Angle of refraction, *r* (degrees)	6.0	13	19	25	30	35

(a) Plot a graph of sin *i* against sin *r*.
(b) From the shape of the graph what is the relationship between sin *i* and sin *r*?
(c) What is the refractive index of the Perspex?
(d) What is the angle of refraction if the angle of incidence is 54°?

The absolute refractive indices in Table 18.1 are for light going from air to the material, that is $n_{a \to g}$. But what is the refractive index of light passing from glass to air $n_{g \to a}$ as shown at the second surface in Figure 18.1?

Because of the reversible nature of light, angle r_1 = angle i_2 and angle i_1 = angle r_2. Therefore at surface 2:

$$\begin{aligned} n_{g \to a} &= \frac{\sin i_2}{\sin r_2} \\ &= \frac{1}{\frac{\sin r_2}{\sin i_2}} \\ &= \frac{1}{\frac{\sin i_1}{\sin r_1}} \\ n_{g \to a} &= \frac{1}{n_{a \to g}} \end{aligned}$$

This is called the 'reciprocal law'.

In general, the refractive index of light going from a material to air is the reciprocal of the absolute refractive index of the material:

$$n_{m \to a} = \frac{1}{n_{a \to m}} = \frac{1}{n_m}$$

Example

Find the refractive index of light going from glass to air.

Solution

$$\begin{aligned} n_{g \to a} &= \frac{1}{n_{a \to g}} \\ &= \frac{1}{1.50} \\ &= 0.67 \end{aligned}$$

If a ray of light passes from one medium to another, for example from water to glass, it is found that the ratio of the sine of the angle of incidence in water (θ_w) to the sine of the angle of refraction in glass (θ_g) is also a constant, n_{w-g}:

$$\frac{\sin \theta_w}{\sin \theta_g} = n_{w-g}$$

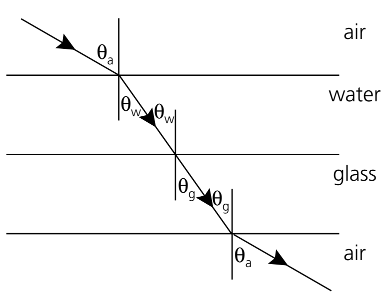


Figure 18.5
How much light rays bend depends on the optical density of the medium. They bend more in glass than in water.

From Figure 18.5:

$$\frac{\sin \theta_a}{\sin \theta_w} = n_w$$

$$\therefore \sin \theta_w = \frac{\sin \theta_a}{n_w}$$

$$\frac{\sin \theta_g}{\sin \theta_a} = \frac{1}{n_g}$$

$$\therefore \sin \theta_g = \frac{\sin \theta_a}{n_g}$$

$$\therefore \frac{\sin \theta_w}{\sin \theta_g} = \frac{\dfrac{\sin \theta_a}{n_w}}{\dfrac{\sin \theta_a}{n_g}}$$

$$\therefore \frac{\sin \theta_w}{\sin \theta_g} = \frac{n_g}{n_w} = n_{w\text{–}g}$$

In general, the **relative refractive index** for light passing from medium 1 to medium 2 is given by the formula:

$$n_{1\rightarrow 2} = \frac{n_2}{n_1}$$

where $n_{1\rightarrow 2}$ is the relative refractive index for light going from medium 1 to medium 2; n_2 is the absolute refractive index for medium 2; n_1 is the absolute refractive index for medium 1.

This results in a more general form of Snell's law, which can be used for light passing between any two media:

$$\frac{\sin \theta_1}{\sin \theta_2} = n_{1,2} = \frac{n_2}{n_1}$$

$$\therefore n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Example

Find the angle of refraction for a ray of light passing from water to glass when the angle of incidence in water is 25°.

Solution

$$n_w \sin \theta_w = n_g \sin \theta_g$$

$$1.33 \sin 25° = 1.5 \sin \theta_g$$

$$\frac{1.33 \sin 25°}{1.5} = \sin \theta_g$$

$$\theta_g = 22°$$

Questions

8 See if you obtain the same result for the relative refractive index for light passing from water to crown glass using $n_{w\text{–}g} = v_w/v_g$ and n_g/n_w.

9 A drop of soapy water ($n_{\text{soapy water}} = 1.38$) was placed onto a block of glass ($n_g = 1.5$), as shown in Figure 18.6. A ray from a laser was shone onto the water at an angle of 38°. Calculate:
(a) the angle of refraction in the soapy water;
(b) the angle of refraction in the glass;
(c) the angle at which the ray exited from the glass;
(d) the relative refractive index of light going from soapy water to glass.

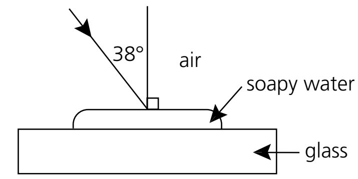


Figure 18.6
For question 9.

10 In each of the cases shown in Figure 18.7 a light ray travels from a substance, X, to air. Find the refractive index of the substance.

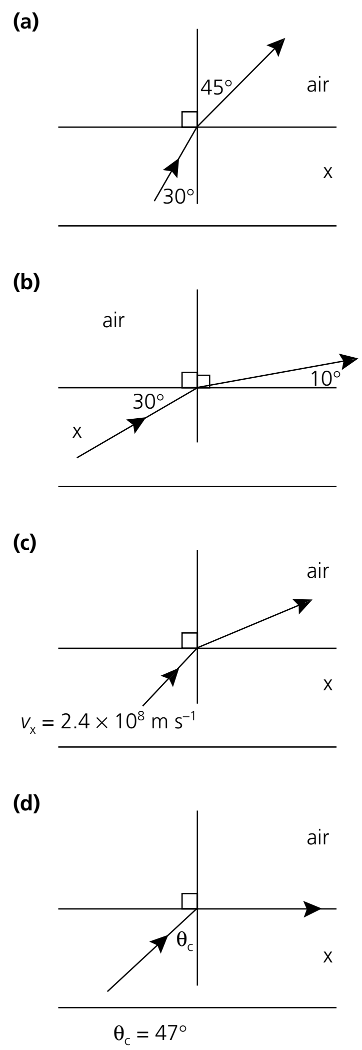


Figure 18.7
For question 10.

11 A layer of water ($n_w = 1.33$) is placed on a block of glass ($n_g = 1.52$), as shown in Figure 18.8. Calculate the angles θ_w and θ_g.

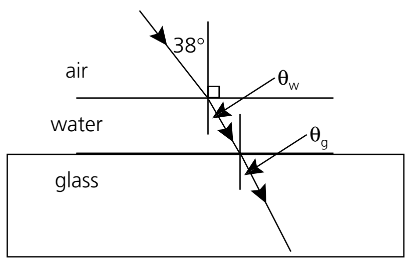


Figure 18.8
For question 11.

12 In which of the following will rays of light bend towards the normal?
(a) Glass to water.
(b) Glass to diamond.
(c) Alcohol to water.
(d) Perspex to heavy flint glass.

Colours

Refraction is due to the velocity of light changing as it goes from one medium to another:

$$n_m = \frac{v_a}{v_m}$$

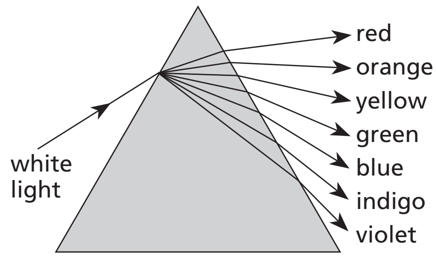
where v_a is the velocity of light in air; v_m is the velocity of light in the material.

Since $v = f\lambda$ and the frequency of waves does not change as they go from one medium to another, then:

$$n_m = \frac{v_a}{v_m} = \frac{f\lambda_a}{f\lambda_m} = \frac{\lambda_a}{\lambda_m}$$

This results in each colour of light having a slightly different absolute refractive index. This means that if white light is shone on the surface of a block of glass, for example, the angle of refraction for each colour will be slightly different. The colours will separate slightly. If this occurs at a second surface, such as the second surface of a prism as shown in Figure 18.9, the effect is increased, resulting in a very visible separation of the colours of light. This phenomenon is known as **dispersion**. (See Photo 18.2 and colour section.) The colour pattern formed is called a **spectrum**. Notice that violet light is refracted the most and red the least.

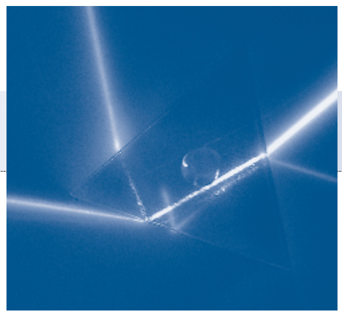
Ch18-Walding 4th 25/8/04 12:03 PM Page 407



$n_{red} = 1.515$
$n_{yellow} = 1.517$
$n_{blue} = 1.523$
$n_{violet} = 1.533$

Figure 18.9
Because the different colours of light have different refractive indices they separate when passing through a prism, creating a spectrum (see also colour section).

Photo 18.2
A continuous spectrum produced by the refraction of white light by a prism (see also colour section).



In summary:

$$\frac{\sin\theta_1}{\sin\theta_2} = n_{1,2} = \frac{n_2}{n_1} = \frac{v_1}{v_2} = \frac{\lambda_1}{\lambda_2}$$

EXAMPLES OF REFRACTION 18.4

A fish eye view To a fish or underwater diver, a tree on the shore would appear to be up in the air and objects would appear to be in different positions from where they actually are because of refraction. However, certain fish have overcome these apparent positional changes to still be able to shoot down insects by squirting a high pressure jet of water from their mouths toward their prey, which can be up to 3 m above the surface. These fish must take account of refraction with a great deal of precision to enable them to aim from under the water to make a 'hit' on an insect. Once hit, the insect falls to the water, where it becomes a meal for these incredible marksmen.

Figure 18.10
Because of refraction, fish see objects at positions that differ from their true positions.

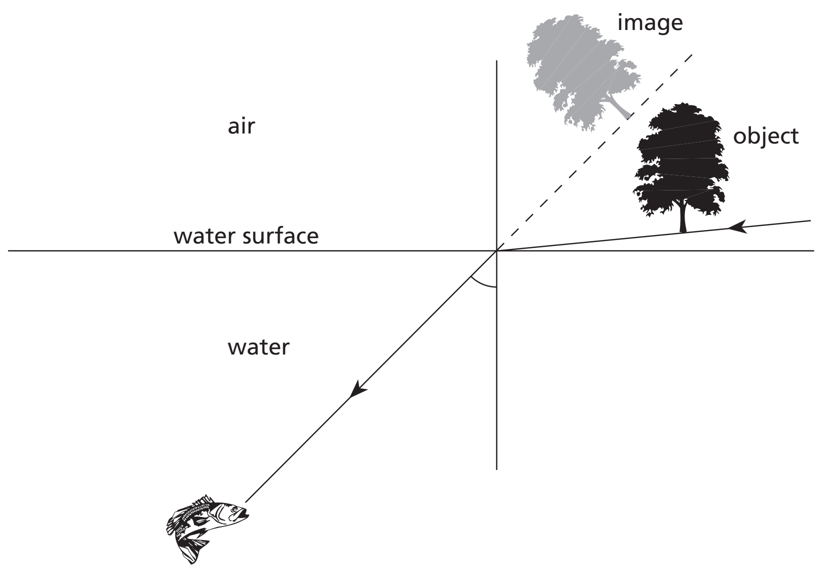
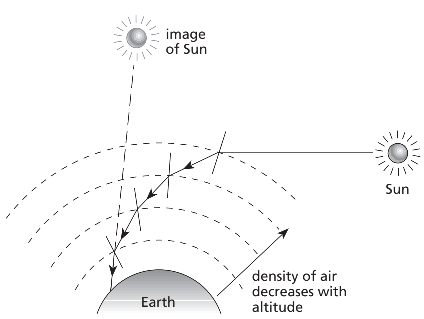


Figure 18.11
Light entering our atmosphere is refracted by layers of differing density, producing an apparent shift in the position of the stars.



Astronomers Like the water-squirting fish, astronomers have to make allowances for refraction when observing stars. Light from the stars travels in straight lines through the vacuum of space until it enters the Earth's atmosphere where it is refracted. The atmosphere of Earth is a more dense medium than the vacuum of space. Stars appear to be at different positions in the sky from where they actually are.

This occurs with light from our Sun. However, since the refractive index for light travelling from space to our atmosphere is only 1.000 29, Figure 18.11 has been exaggerated. The observed position of the Sun is only about 0.5° or one Sun's diameter higher in the sky than its real position. Since it takes the Earth about 2 minutes to rotate through 0.5° we gain approximately 4 minutes of extra sunlight a day due to refraction at sunrise and sunset.

Ch18-Walding 4th 25/8/04 12:03 PM Page 408

Apparent depth The refraction of light results in objects in different media appearing to be closer than they are. For example, a toy at the bottom of a pool will appear to be closer to the surface than it really is, to an observer standing above the pool. (See Figure 18.12.)

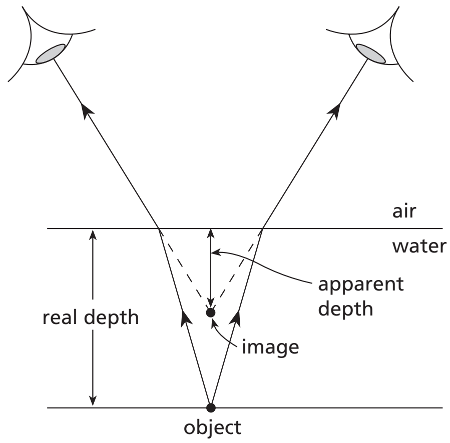


Figure 18.12
Because of refraction objects appear closer to the surface of a pool than they actually are.

Light rays travelling from the toy to your eyes are refracted away from the normal at the water's surface. To your eyes and brain, which trace these rays back to where they appear to meet, the object appears to be closer to the surface than it actually is. This depth is called the **apparent depth**. It can be derived that:

$$\frac{\text{true depth}}{\text{apparent depth}} = n$$

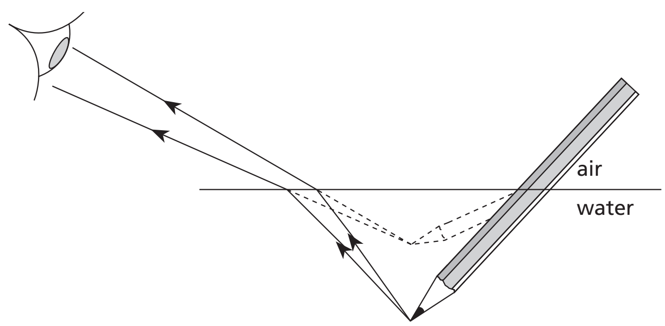


Figure 18.13
The pencil appears bent because the parts of the pencil under water appear closer to the surface.

This is why a pencil placed in water appears to be bent. (See Figure 18.13.)

Rays from the tip of the pencil are refracted away from the normal at the surface of the water. On tracing them back they appear to come from the image of the tip of the pencil, which is closer to the surface.

Example
A stone at the bottom of a pool in a creek appears to be 1.2 m from the surface. What is the true depth of the pool? (n_w = 1.33)

Solution

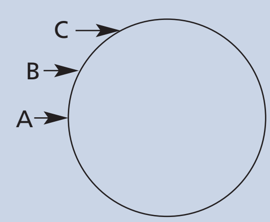
$$\frac{\text{true depth}}{\text{apparent depth}} = n$$

$$\frac{\text{true depth}}{1.2} = 1.33$$

$$\text{true depth} = 1.6 \text{ m}$$

NOVEL CHALLENGE

In 1621, French scientist Rene Descartes published a diagram showing the refraction and total internal reflection of light in a raindrop. Redraw a big circle like the one in the diagram and show the path of the parallel sunlight rays A, B and C that strike the drop at the 9, 10 and 11 o'clock positions. Assume n_{glass} = 1.5.



Questions

13 A student on a biology field trip dropped a coin in a creek. The depth of the water appeared to be 75 cm so he rolled up his sleeves to retrieve the money. What would be the consequences of such an action? Explain! (The refractive index of water is 1.33.)

TOTAL INTERNAL REFLECTION 18.5

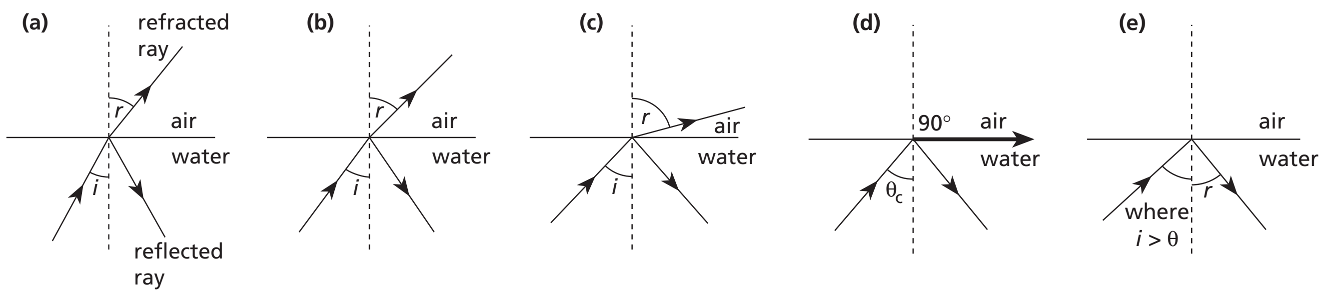


Figure 18.14
When light passes from a more dense to a less dense medium it bends away from the normal. However, after the refracted ray = 90° the incident ray is totally reflected, producing total internal reflection.

As previously discussed a ray of light bends toward the normal when going from a less dense to a more dense medium. The opposite is also true. Rays will bend away from the normal when going from a more dense to a less dense medium, as was shown in Figure 18.1. This results in an odd situation as the angle of incidence increases as shown in Figure 18.14. There comes a stage where the angle of refraction is 90° (Figure 18.14**(d)**). The angle of incidence that produces this is called the **critical angle** (θ_c). If the angle of incidence is further increased the ray of light is entirely reflected from the surface at an angle equal to the angle of incidence. This is called **total internal reflection** and occurs when light travels from a more optically dense medium to a less optically dense medium and the angle of incidence is greater that the critical angle.

For a ray of light going from water to air:

$$\frac{\sin \theta_w}{\sin \theta_a} = n_{wa} = \frac{1}{n_w}$$

When the angle of incidence θ_w is equal to the critical angle θ_c, the refracted angle $\theta_a = 90°$, and $\sin 90° = 1$, then:

$$\sin \theta_c = \frac{1}{n_w}$$

Example
Find the critical angle for a light ray passing from light flint glass to **(a)** air; **(b)** water.

Solution
(a)

$$\sin \theta_c = \frac{1}{n_g}$$
$$= \frac{1}{1.58}$$
$$\theta_c = 39°$$

(b)

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

$$1.58 \sin \theta_c = 1.33 \sin 90°$$

$$\sin \theta_c = \frac{1.33 \times \sin 90°}{1.58}$$

$$\theta_c = 57.3°$$

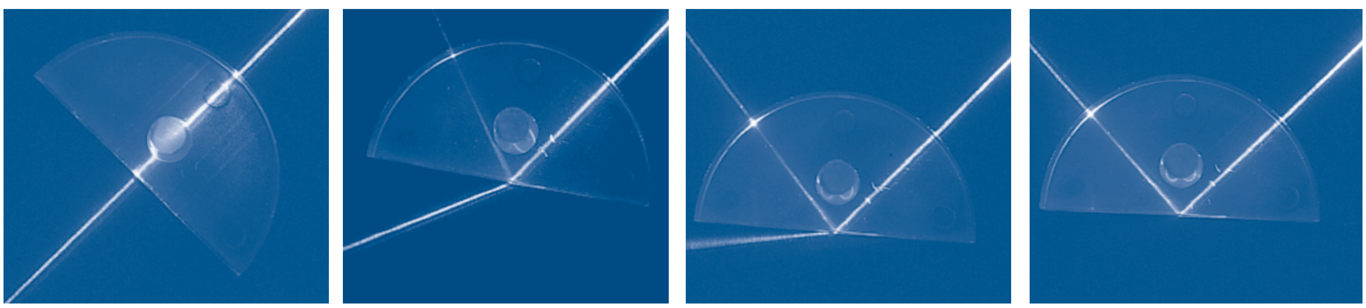
NOVEL CHALLENGE

When you squeeze the air out of a plastic soft-drink bottle while it is under water, bubbles rise to the surface. Predict what they would look like when observed from the water's surface. Think about air lenses and total internal reflection. You'd be very surprised!

Total internal reflection can be demonstrated easily by using a semicircular block of glass and a light box. (See the Photo 18.3.)

The block is placed on a sheet of white paper and a ray from a light box is directed onto the centre of the semicircular block; the ray entering the glass, as it is along a radius, is then perpendicular to the surface. Therefore no refraction occurs at the first surface. However, refraction occurs at the second surface and the ray bends away from the normal. As the angle of incidence at this surface is increased, the angle of refraction also increases. (Notice that you will start to see a faint reflected beam from this surface as light is both reflected and refracted from transparent surfaces.) When the angle of incidence is approximately 42°, the refracted beam will be along the straight surface of the block of glass; that is, the angle of refraction is 90°. If the angle of incidence is made slightly greater, the refracted beam disappears as the light beam is reflected back inside the block of glass at an angle equal to the angle of incidence. (See the Photo 18.3.) This is total internal reflection.

Photo 18.3
Total internal reflection produced by light passing from glass to air. This occurs when the angle of incidence is greater than the critical angle. The refraction of light through a semicircular block of glass showing total internal reflection.



Questions

14 In each of the following situations where a light ray passes from one medium to another state whether it is possible for total internal reflection to take place. Explain.
(a) Air to glass.
(b) Diamond to air.
(c) Water to glass.
(d) Flint glass to air.
(e) Ice to a vacuum.
(f) Crown glass to Perspex.

15 A block of ice is placed on top of a semicircular block of crown glass (Figure 18.15). At what minimum angle would all light incident on the boundary between the two surfaces be reflected?

16 A ray of light travels from one medium to another. It is found that total internal reflection occurs when the incident angle is greater than 54°. If the refractive index of the first medium is 1.49, calculate the refractive index of the second medium.

17 Students investigating total internal reflection using a semicircular block of glass notice that before total internal reflection occurs there is a faint reflected beam. Comment on the intensity of beams **(i)**, **(ii)** and **(iii)** in Figure 18.16 as the angle of incidence increases.

Figure 18.15
For question 15.

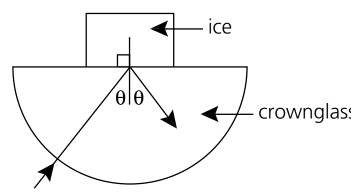


Figure 18.16
For question 17.

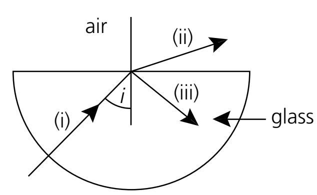


Figure 18.17
Because of total internal reflection prisms are capable of bending light rays through 90°.

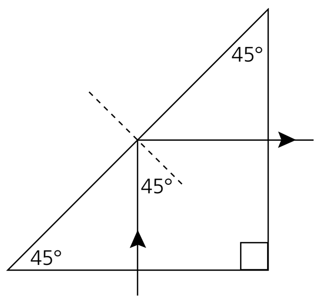


Figure 18.18
Periscopes use total internal reflection in prisms.

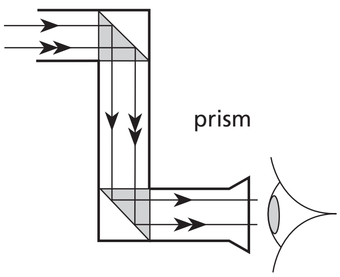


Figure 18.19
Quality periscopes use prisms rather than mirrors because mirrors produce multiple images if thick glass is used.

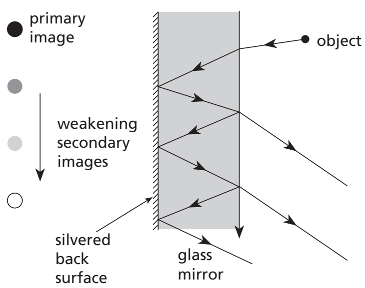
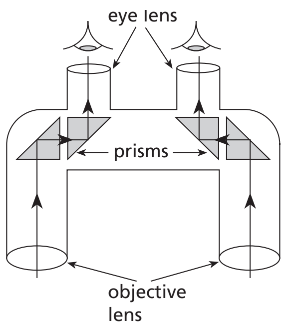


Figure 18.20
A schematic diagram of binoculars using glass prisms.



Activity 18.2 UNDERWATER BUBBLES

1 Hold an empty plastic soft drink bottle upside-down in a bucket of water, an aquarium, or a swimming pool.
2 Squeeze the bottle and watch the air bubbles rise. They look shiny.
3 Why is this?

USES OF TOTAL INTERNAL REFLECTION 18.6

Prisms

A glass 45°/45° prism as shown in Figure 18.17 can be used for total internal reflection and has many uses.

Light striking one surface of the prism at right angles makes an angle of 45° with the second surface. The angle of incidence is greater than the critical angle of 42° and the light is therefore reflected from this surface. It then strikes the third surface at right angles. The rays have thus made right-angled turns.

This makes them useful in quality periscopes (Figure 18.18). They have an advantage over mirrors because mirrors produce multiple images as light is reflected from the back and front surfaces a number of times (Figure 18.19). The coating on mirrors can also flake, reducing the reflected light intensity.

Prisms are also used in prism binoculars, making these pieces of equipment much more compact than older telescopes (Figure 18.20). In single lens reflex cameras a pentaprism is used to reflect the incoming light back to the viewfinder as well as invert it so that the photographer is actually seeing the light which is entering the camera lens used to form the image on the film.

If light is incident at right angles onto the hypotenuse of the prism it is reflected back the way it came (Figure 18.21). This property of prisms makes them useful for reflectors on bicycles and 'cats'-eyes' on roads. Notice, however, that the rays of light are inverted.

Diamonds are cut in similar ways to reflect light incident on them to produce many internal reflections and thus to sparkle.

Optical fibres

One of the major developing uses of total internal reflection is in optical fibres. Fibre optics is a branch of optics dealing with the transmission of light through fibres or thin rods of glass or some other transparent material of high refractive index. If light is admitted at one end of a fibre, it can travel through the fibre with very low loss, even if the fibre is curved.

Optical fibres have been around for decades. You might remember, or have seen, those stringy plastic lights (Fantasy lights) that were the rage back in the 1970s. These consisted of basic optical fibres. (See Photo 18.4.)

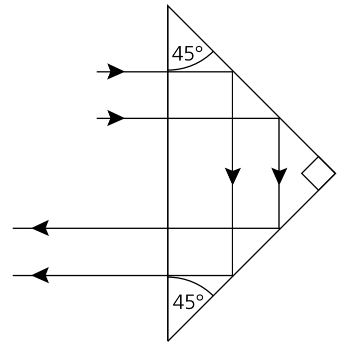
However, the number of medical and communication uses of optical fibres has exploded over the past decade.

An optical fibre consists of a very pure glass fibre as thin as a hair 0.125 mm with a layer of cladding around the outside (Figure 18.22), to protect it from damage and moisture.

The outside layer has a lower refractive index than the inside material, thus creating a situation where light propagating in the central layer is travelling in a more dense material than in the outside layer. This means the light is totally internally reflected if it strikes the boundary between the two media at an angle greater then the critical angle. Thus light is reflected and reflected and reflected along the length of the fibre, which can be bent into any shape as long as it is not kinked. Once kinked, surface cracks allow light to refract out. In underwater cables the glow at the kink attracts fish, which can eat and sever the cable.

Optical fibres can be made as either step index type where the refractive index changes rapidly at the boundary between the core and the cladding, or graded index where there is a more gradual change of refractive index from the centre to outside.

Figure 18.21
45°/45° prisms can also cause light rays to bend through 180° when the light is incident on the hypotenuse.

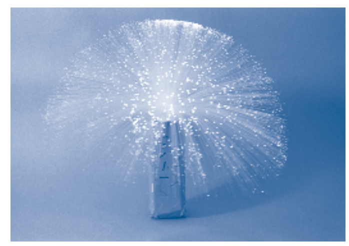


Communication

One of the first large commercial users of optical fibres was the telecommunications industry. Optical fibres were first used in the USA in the 1960s and are now replacing copper conductors in telephone and cable TV and data systems worldwide. Digital electrical signals are converted into light pulses and transmitted over optical fibre cable by switching light-emitting or laser diodes on and off. At the other end, these optical digital pulses are converted back into electrical signals by photo-transistors. Optical fibres are thinner, cheaper and lighter than equivalent copper conductors and can carry much more information. One particular fibre optic cable laid from New Jersey to Britain and France can carry 50 000 simultaneous voice conversations as well as other information, such as ten channels for cable TV and Internet data. There is no cross-talk between voice conversations and they are almost impossible to ‘bug’.

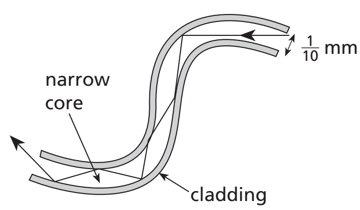
It is suggested that Australia has an international fibre optic cable capacity of over one terabit per second (1.0 Tb s^{-1}). Bandwidth capacity is forecast to increase to over 4 Tb s^{-1} by 2004. The new Australia–Japan cable connects the east coast of Australia with Japan and North America. It has doubled existing broadband capacity to the west coast of the United States and increased the capacity to North Asia 15-fold. Australian domestic networks consist of fibre-optic, wireless, satellite and microwave systems; but fibre optic is now the predominant technology. Major fibre optic networks provided by companies such as Telstra, SingTel-Optus, PowerTel, Uecomm and NextGen connect Sydney and the major east-coast cities. Total bandwidth capacity of satellites covering Australia is estimated to be 4 Gb s^{-1}.

Photo 18.4
An optical fibre lamp.



The installation of asymmetric ADSL (asymmetric digital subscriber line) technology into local telephone exchanges is connecting businesses and residential homes into a high-speed digital broadband network. SingTel-Optus uses a fibre optic cable network (CABLE) which is supported on the local powerline distribution grid. With data speeds from 256 kb s^{-1} to over 2 Mb s^{-1}, everyone will have access to Internet speeds 30–50 times faster than the standard dial-up service. New South Wales carries the majority of Australia’s Internet traffic. Over 500 Internet service providers in Australia use Internet Protocol (IP) technology, which mixes voice, data and video transfer over the same networks.

Figure 18.22
An optical fibre consists of a thin glass fibre of higher refractive index than the outside cladding layer. Thus total internal reflection is used to reflect light pulses along the length of the fibre.



NEI Activity 18.3 INTERNET CONNECTIONS

Compare and contrast ADSL, cable and the standard 56K V90 modem. In your discussion discuss the upload and download speeds of ADSL, cable and dial-up, as well as the pricing plan for the major Internet service providers in your local area. Are there any other factors that you need to consider when thinking about installing an Internet connection into your home and computer system?

Medical

The medical profession was the first to make use of optical fibres. Surgeons use bundles of fibres to look inside a person’s stomach and lungs without surgery. A bundle of fibres is introduced into the stomach via the throat. Light is shone down some of the fibres and reflected light from the stomach is transmitted back via other fibres. If, for example, an ulcer is discovered, a laser beam is transmitted down the fibres to burn and seal the ulcer. These devices are commonly called ‘endoscopes’. Most recently fibre optic endoscopes have been used coupled with colour video cameras and external video monitors to increase ease of viewing. (The word endoscope comes from the Greek *endo skopion* meaning ‘within’ and ‘to see’.)

Other effects of total internal reflection

Mirages

Mirages result from refraction and total internal reflection. On a hot day imaginary pools of water appear on the road or the desert. As rays of light from the sky reach the ground they undergo gradual refraction in the layers of hot air above the road. The rays end up hitting the hot layer just above the ground at an angle greater than the critical angle and thus are reflected from this layer to the observer who sees the layer as a pool of water. In fact, it is the reflection of the sky (Figure 18.23).

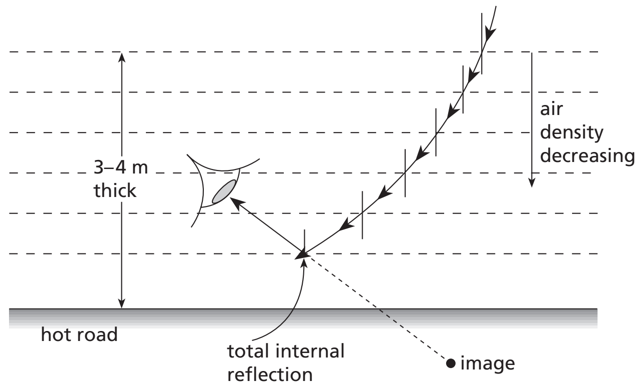


Figure 18.23
A mirage is formed when light from the sky is refracted as it passes through the different density layers of air above a hot surface. It is reflected from the bottom hot layer.

Rainbows

A rainbow forms when water droplets in the rain refract the sunlight. The process actually involves two refractions and a total internal reflection (Figure 18.24).

When sunlight from behind the observer strikes rain droplets in front of the observer the light is refracted on entering the droplet, totally internally reflected, and refracted on leaving. Red light is refracted the least and violet the most, therefore a person on the ground sees red light from high in the sky and the other colours from raindrops closer to the ground.

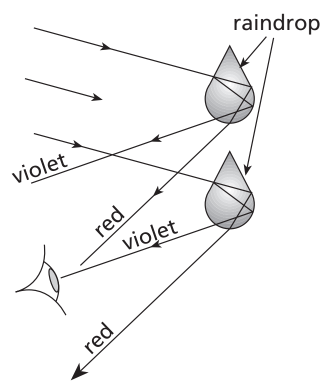


Figure 18.24
A rainbow is formed when sunlight undergoes refraction and reflection inside raindrops.

INVESTIGATING

A 'blue sky' is due to the scattering of sunlight. But so is a 'red sky'.
How can you resolve this anomaly?

Questions

18 Optical fibres have a less optically dense layer surrounding the fibre so as to produce total internal reflection. If the fibre has a refractive index of 1.70 and the outside cladding layer has a refractive index of 1.48:

(a) calculate the minimum angle at which light incident on the junction is totally internally reflected;

(b) find the speed of light in the fibre.

19 A scuba diver working in the ocean looks up to notice the Sun setting on the horizon. At what angle to the normal to the surface will he need to look? ($n_{\text{salt water}}$ = 1.38.)

Activity 18.4 RESEARCH QUESTIONS

Write a short report on one of the following topics:

Diamonds Why do real diamonds sparkle more than counterfeit ones? In your answer describe the 'brilliant cut' — one of the most common styles of faceting a diamond. How many separate facets are there in the 'brilliant cut' and who invented it?

Endoscopy Fibre optics endoscopy is used for examining the oesophagus, pancreas and bladder, among other organs, to detect the presence of cancer. Name two other medical conditions endoscopy is used for, and say how the endoscope is inserted, and how the doctor views the image.

Phone cables Find out the following about fibre optic phone cables:

1. What is the distance between boosters?
2. How is the cable joined?
3. What is the diameter of a single fibre?
4. How far underground is the cable buried?
5. What is multiplexing?

Glass fibre Early trials of optic fibres in medicine were unsuccessful but in 1970 Corning Glass overcame the problem. What was the problem and how was it solved?

UFO sightings UFO hoaxes have often been achieved by reflection–refraction phenomena. Research some of the more famous hoaxes, or try to videotape your own hoax. Late in the afternoon with the lights out in your house or garage, stand at a window and hold a torch in one hand. Direct the strong torch beam out of the window. You will be able to see the transmitted light from the backyard and sky as well as a superimposed reflection of the torch light in the window. You can make the reflected spot of light hover and shimmer as if it were out in the yard. A simple video recording of this type of reflection has been the basis of some famous UFO hoaxes in the past. Try it yourself and have some fun.

PHYSICS FACT

Have you noticed the green iridescent colours that sometimes appear on bacon and corned meat? This is not rotting meat but is caused by microscopic droplets of oil and water of differing refractive indices on the surface causing the interference of light. If it goes really green then that's the bacteria breaking down the oxygen transport protein (myoglobin) to produce green compounds. Heat will show the difference. Heat will make the oil droplets go away but not the green rot.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

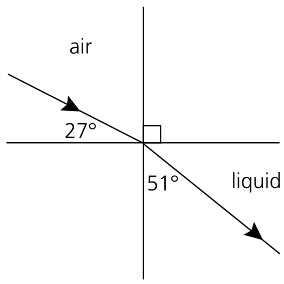
Review — applying principles and problem solving

***20** For each of the following situations where a light ray passes from one medium to another, state whether the light ray will bend away from or toward the normal:
- **(a)** Air to water.
- **(b)** Glass to air.
- **(c)** Water to glass.
- **(d)** Diamond to glass.
- **(e)** Flint glass to crown glass.
- **(f)** Perspex to a vacuum.

***21** A light ray strikes the surface of a block of fused quartz at an angle of 54° to the normal.
- **(a)** Calculate the angle of refraction.
- **(b)** Find the velocity of light in the quartz.

***22** Light from a light box is directed at an angle of 30° onto a block of glass whose refractive index is 1.5, and then onto the surface of salty water whose refractive index is 1.4. In which case will the light bend the most?

Figure 18.25
For question 23.



NOVEL CHALLENGE

Have you ever taken a photo of reflections in a still pool of water? Imagine you photographed some wallabies resting next to a still dam. How could you tell which way to hold the photo if the reflection was a perfect copy?

***23** A ray of light is directed onto the surface of a liquid as shown in Figure 18.25. Calculate the speed of the light in the liquid.

***24** Calculate the absolute refractive index for light going from air to a medium when the velocity of light in the medium is **(a)** 2.4×10^8 m s^{-1}; **(b)** 1.8×10^8 m s^{-1}; **(c)** 2.7×10^8 m s^{-1}; **(d)** 3.6×10^8 m s^{-1}.

***25** What is the relationship between the relative refractive index for light passing from glass to diamond ($n_{g\rightarrow d}$) and the absolute refractive indices of glass (n_g) and diamond (n_d)?

***26** The absolute refractive indices of certain media are **(i)** 1.78; **(ii)** 1.2; **(iii)** 2.1; **(iv)** 1.42.
(a) Calculate the speed of light in each of the media.
(b) Which substance is the most optically dense?

***27** The refractive indices of flint glass and turpentine are 1.65 and 1.5 respectively.
(a) Calculate the refractive index for light passing from turpentine to flint glass.
(b) If the angle of incidence in the turpentine is 49° calculate the angle of refraction in the flint glass.

***28** Red light from a laser of wavelength 633 nm is used in refraction experiments. What is the wavelength of this light in glass? ($n_g = 1.5$.)

****29** Biology students on a rocky shore excursion notice that a shell in a rock pool appears to be 50 cm from the surface of the water, but a physics student informs them this is not correct.
(a) What depth is the rock pool really? (The refractive index of water is 1.33.)
(b) How is the real depth influenced by the fact that it is salt water rather than fresh water?

****30** A mirage is formed when light from the sky passing through cool layers of air into hotter layers just above the road is totally internally reflected. Calculate the critical angle for light travelling from cool air to hot air. ($n_{\text{cool air}} = 1.0004$ and $n_{\text{hot air}} = 1.0002$.)

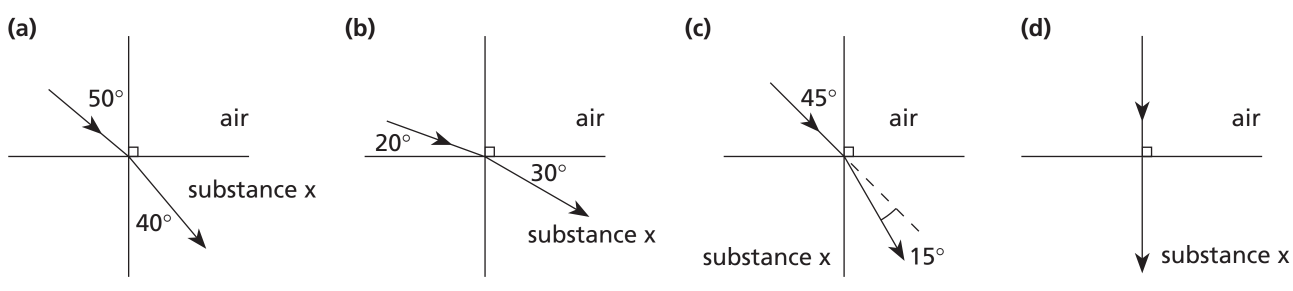
****31** Through what surface area can a fish see if it is 5.0 m deep in sea water? ($n_{sw} = 1.38$.)

****32** A pulse of light enters an optical fibre of refractive index 1.53. What is the refractive index of the cladding material if the critical angle required is 82°?

****33** The index of refraction for glass is different for different colours of light. The refractive index for blue light ($\lambda = 430$ nm) passing from air into flint glass is 1.650 and for red light ($\lambda = 680$ nm) is 1.615. If a beam containing blue and red light is shone onto a block of flint glass at an angle of 52°, find the angle between the blue and the red rays in the glass.

***34** In each of the four cases shown in Figure 18.26, a light ray travels from air to the substance. Use the diagrams to find the refractive index of the substances.

Figure 18.26
For question 34.



****35** Students experimenting with an unknown transparent substance used a light box to produce a ray of light. They shone the ray onto the unknown substance at various angles and measured the angles of refraction. Refer to Table 18.4.

Table 18.4

Angle of incidence, i (degrees)	20	30	40	50	60	70	80
Angle of refraction, r (degrees)	10	15	20	24	27	30	31

(a) Redraw the table including values of sin i and sin r.
(b) Plot a graph of sin i against sin r.
(c) What shape graph was obtained?
(d) What is the relationship that exists between sin i and sin r?
(e) From this graph find the refractive index of the substance.
(f) What is the substance?

***36** Complete the diagrams shown in Figure 18.27 where a ray of light passes from air to a block of glass.

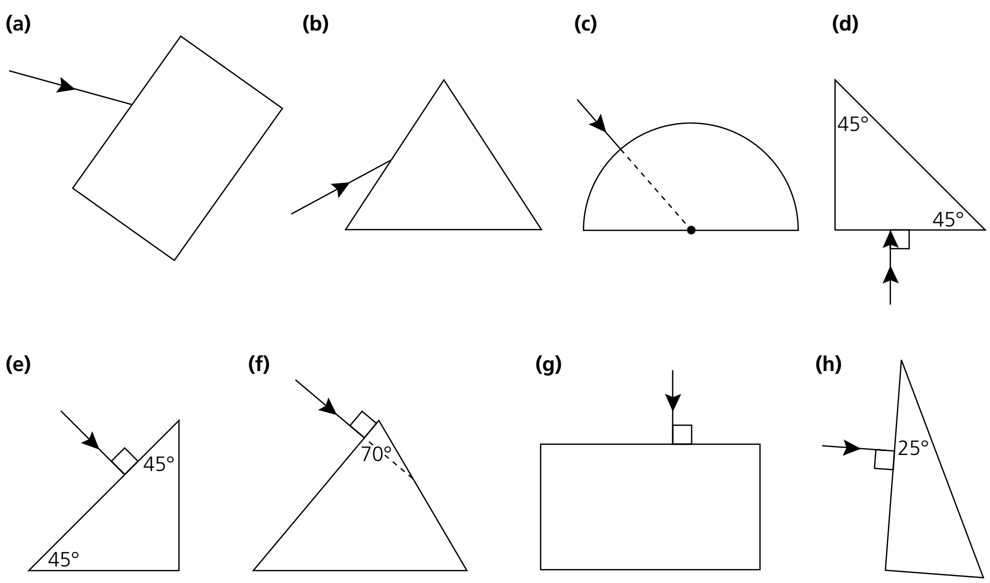


Figure 18.27
For question 36.

37 A ray of light strikes the surface of a beaker of water at the centre of the surface. If the ray hits the bottom of the beaker 4.0 cm from the centre (refer to Figure 18.28), calculate the angle of incidence. (n_w = 1.33.)

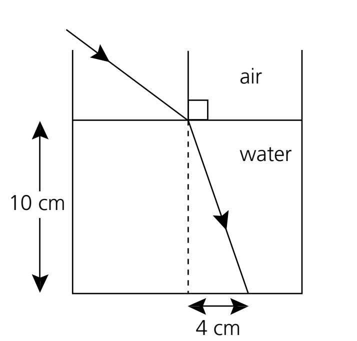


Figure 18.28
For question 37.

Ch18-Walding 4th 25/8/04 12:03 PM Page 417

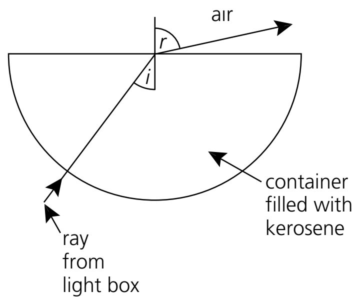


Figure 18.29
For question 38.

****38** Students were given the task of finding the critical angle for light passing from kerosene to air. They filled a semicircular container with kerosene and used a light box to produce a single ray. They intended to shine the ray through the curved side of the container onto the centre of the straight edge and gradually increase the angle of incidence until the refracted ray was along the straight edge. (Refer to Figure 18.29.) However, before they finished the experiment the power went out, but they did take some measurements before this happened. Refer to Table 18.5. Graphically find the critical angle.

Table 18.5

Angle of incidence, i (degrees)	10	15	20	25	30	35	40
Angle of refraction, r (degrees)	13	22	30	38	47	57	70

Extension — complex, challenging and novel

*****39** A beam from a laser strikes the surface of a 5.0 cm thick block of Perspex at an angle of 41° to the normal. Find the perpendicular distance between the original direction of the beam and the direction of the beam as it leaves the Perspex. (See Figure 18.30.)

Figure 18.30
For question 39.

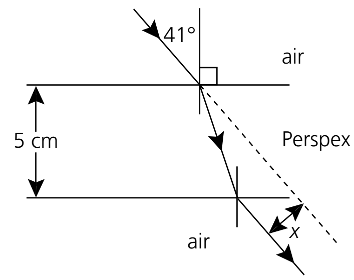
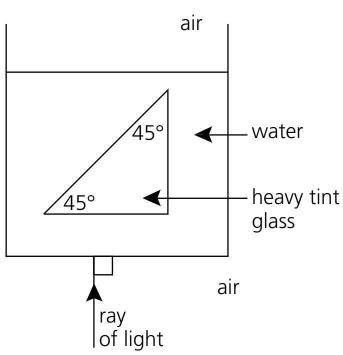


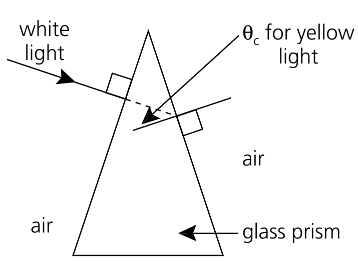
Figure 18.31
For question 40.



*****40** A 45°/45° heavy flint glass prism is placed in a beaker of water. A ray of light from a light box is incident on one side of the prism. (Refer to Figure 18.31.) Analyse the passage of this ray through the water and the glass prism. (n_g = 1.62, n_w = 1.33.) Use a labelled diagram to show the passage of the light ray.

****41** A beam of white light is shone onto a glass prism such that the angle with the second surface is the critical angle for yellow light. (Refer to Figure 18.32.) Analyse what occurs to the beam of white light.

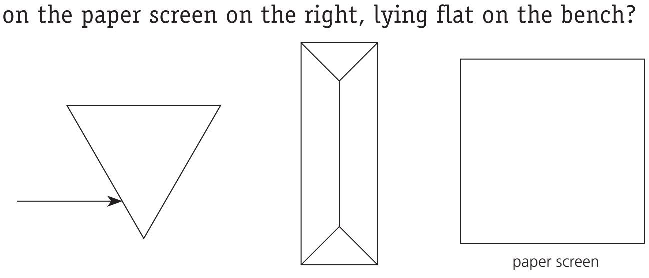
Figure 18.32
For question 41.



****42** In one of his first optics experiments, Newton laid two prisms and a sheet of paper on a bench; from overhead they appeared as in Figure 18.33.

He allowed a beam of white light to enter from the left (shown by the arrow), where it split into its colours and fell on to the second prism. What will you see on the paper screen on the right, lying flat on the bench?

Figure 18.33
For question 42.



****43** **(a)** The plastic rings holding a six-pack of beer stubbies together is made from polythene. It has the same refractive index as sea water, as a result of which animals often eat them or get stuck in them. Propose as many ways as you can of solving this problem. We can think of at least four.

(b) The velocity of light is less in saltwater than in fresh. What would a rainbow look like in a saltwater spray?

CHAPTER 19

Lenses

19.1 INTRODUCTION

Did you know that the earliest eyeglasses were thick convex lenses, which reminded their makers of lentils, hence the term 'lens', from the Latin for 'lentil beans'?

Lenses play a major part in everyday life and in the scientific world. They are present in spectacles, contact lenses, microscopes, telescopes, overhead projectors, and photocopiers, to name a few. Simple lenses are used in science classes for mineral identification and for map reading in geography.

Lenses are transparent devices that refract light. Differently shaped lenses refract light differently. Once they were made of glass but now, to reduce the weight, many lenses are made of plastic. Magnetic lenses are used to focus charged particles in particle physics. Eyes of animals contain lenses made of organic material that can change shape to produce a range of depth of vision.

At the end of this chapter you may be able to answer questions such as these:

- What devices around me rely on lenses of various types?
- If lenses refract light why don't we see colours when using spectacles due to the fact that each colour of light is refracted differently?

19.2 SHAPE OF LENSES

Lenses are curved pieces of glass or plastic in many shapes, although they all serve a similar purpose — they refract or bend light in one way or another. Several shapes are shown in Figure 19.1. However, the ones most commonly used are **biconcave** (concave) or **biconvex** (convex) lenses.

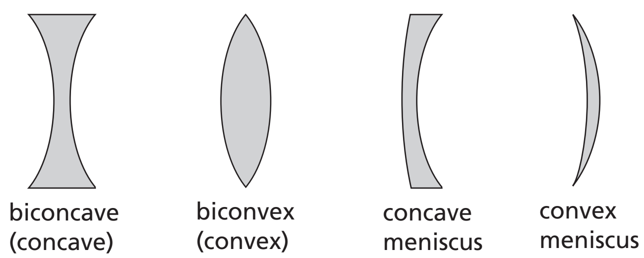


Figure 19.1 Some of the different shapes of lenses.

One common feature of all lenses is that light incident on the lens always bends toward the thickest part. Recall the refraction of light through a prism (Figure 19.2). Light is refracted toward the normal at the first surface and away from the normal at the second surface. The overall effect is that light always bends toward the thickest part of the prism.

Ch19-Walding 4th 25/8/04 12:07 PM Page 419

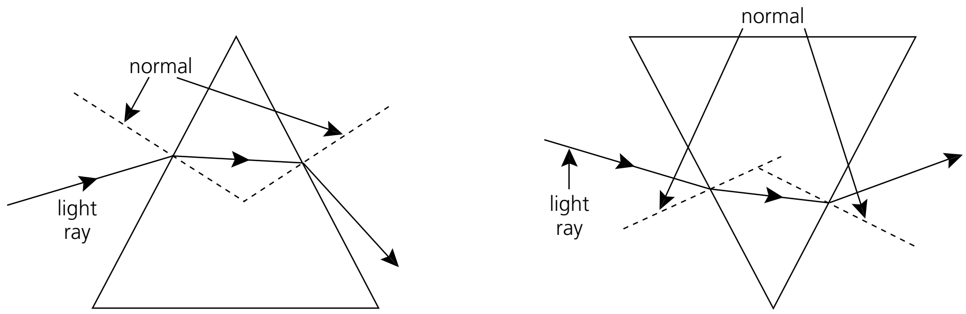


Figure 19.2
Light through a prism always refracts toward the thickest part of the prism.

We can regard a lens as being a system of tiny prisms of slightly different shapes, so that light rays are refracted at both sides of the lens, which results in the bending of light rays toward the thickest part. This makes light rays incident on a convex lens bend toward the centre, thus focusing or converging the light. This type of lens is therefore also called a **converging lens**. Light incident on a concave lens spreads or diverges the light. This lens is also called a **diverging lens** (Figure 19.3).

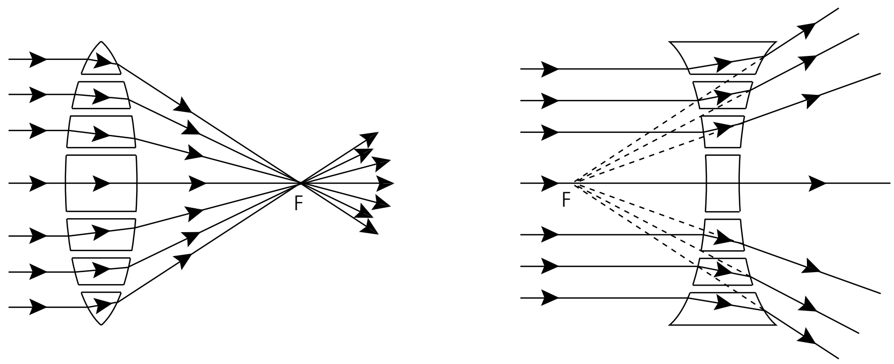


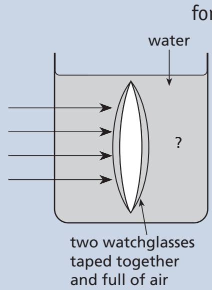
Figure 19.3
Lenses can be considered to be made up of tiny prisms, thus light will always refract to the thickest part of the lens.

Other shaped lenses produce effects that depend on their shape. *Note:* refraction occurs at both sides of the lens but to simplify the drawing of ray diagrams we will draw a line through the centre of the lens and have all refraction occur at this line.

FEATURES OF LENSES 19.3

NOVEL CHALLENGE

You have made an air lens by taping two watchglasses together. This does nothing in air but what will it do underwater? 'Myopic people see better underwater' — account for this.



Similar terms to those associated with mirrors are used in describing lenses. Each face of the lens can be drawn using a compass. The centre of the circle that produces the curved surface is called the **centre of curvature**. Each face can have a different curvature. However, for our discussions we will use lenses whose sides have the same **radius of curvature**. The line joining the centre of curvature through the optical centre of the lens is called the **principal axis**. If rays of light parallel to the principal axis are incident on a convex lens the rays converge to a point on the other side of the lens — the **principal focus** (F). Since lenses can be used either way around they have two focal points, one on either side of the lens. If rays of light parallel to the principal axis are incident on a concave lens they diverge. However, if these rays are traced back they appear to come from a point on the same side of the lens as the light originates. This is a **virtual** focus. The distance from the focal point to the optical centre of the lens is the **focal length** (f). More powerful lenses have shorter focal lengths and are much thicker (Figures 19.4 and 19.5).

Lens makers use the unit of the **dioptre** (D) to define the **optical power** of a lens. This is equivalent to the reciprocal of the focal length (in metres). For example, the optical power of a 20 cm focal length convex lens is $1/0.20$, which equals $+5$ D. For a concave lens this becomes $1/-0.20 = -5$ D. The term power in this case refers to the ability of the lens to refract light. High power lenses will refract light to a greater degree than low power lenses.

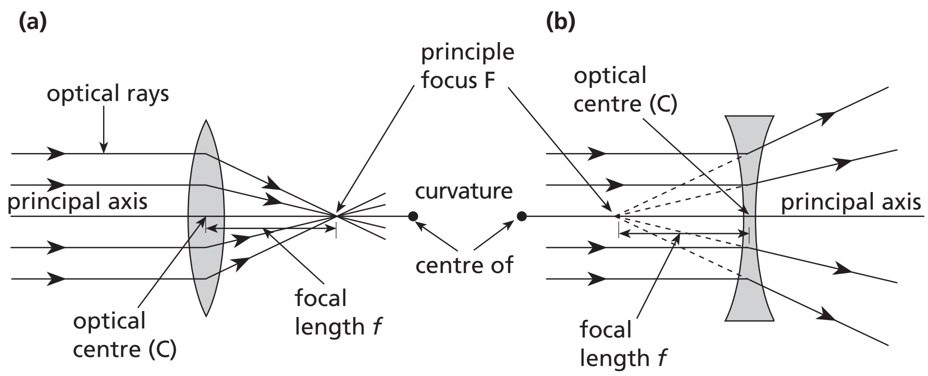
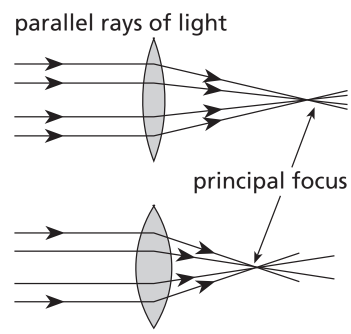


Figure 19.4
Features of convex **(a)** and concave **(b)** lenses.

19.4 RAY DIAGRAMS AND IMAGES

If you look through a magnifying glass (a convex lens) at an object you can see the object. In actual fact you are only seeing the image of the object. This can be easily verified if you look at an object a long distance away while holding the lens at arm's length. The image is upside-down.

Figure 19.5
Thick lenses have smaller focal lengths than thin lenses and are more powerful.



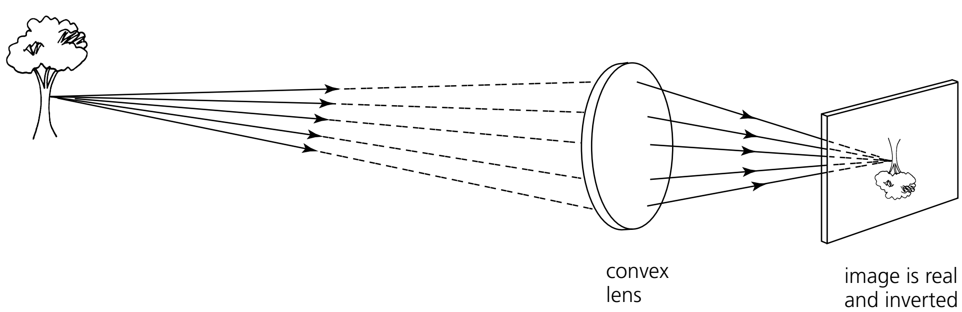


Figure 19.6
Images of distant objects are inverted, indicating that when you look through a lens you are observing the image and not the object itself.

Ray diagrams

Ray diagrams can be drawn to find the position and nature of an image. Again, any number of rays can be drawn to determine the position and nature of the image but many would require the use of protractors, many calculations, and measurements of the angles of refraction at both surfaces to produce an accurate ray diagram. However, three easily drawn rays are most commonly used to simplify the drawing of ray diagrams (Figure 19.7).

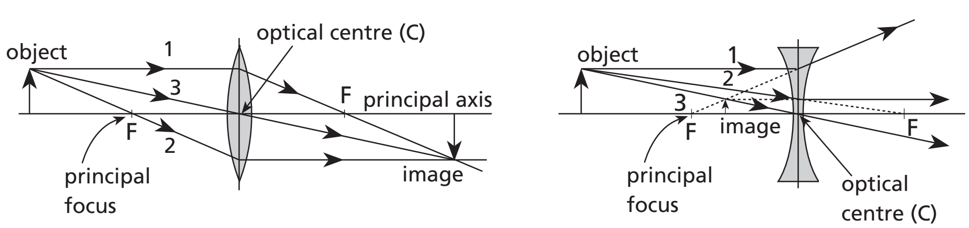


Figure 19.7
The three rays that are used to find the image produced by a lens: one parallel to the principle axis (1), one through the focal point (2), and one through the optical centre of the lens (3).

Ch19-Walding 4th 25/8/04 12:07 PM Page 421

NOVEL CHALLENGE

A man wakes up and his digital clock reads 5:20. He then realises that there is a glass of water in front of the display acting like a convex lens, and he has arisen too early. What is the real time? Use a diagram to show this.

1 The ray parallel to the principal axis refracts through the principal focus for a convex lens, and appears to come from the focus for the concave lens.
2 A ray through the focus refracts parallel to the principal axis for the convex lens. For a concave lens the ray that is lined up with the focus on the opposite side refracts parallel to the principal axis.
3 The ray through the centre of the lens continues unchanged in direction.
Remember, to find the position of the image requires the use of two rays. The third can be used to double-check.

Images in convex lenses

Figure 19.8 shows a few examples of ray diagrams. The object has been placed at various positions with respect to the convex lens — greater than 2*f*, at 2*f*, between 2*f* and F, and between *F* and the lens (Figure 19.8).

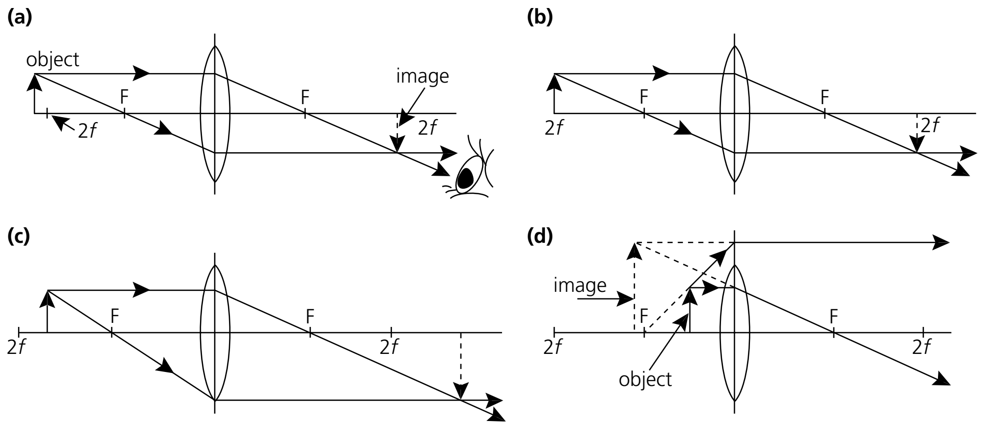


Figure 19.8
(a) When the object is outside 2*f* the image produced is: real, inverted, smaller, and between F and 2*f* on the opposite side of the lens. **(b)** When the object is at 2*f* the image produced is: real, inverted, the same size, and at 2*f* on the opposite side of the lens. **(c)** When the object is between 2*f* and F, the image produced is: real, inverted, larger, and outside 2*f* on the opposite side of the lens. **(d)** When the object is between F and the lens, the image produced is: virtual, upright, larger, and between F and the lens on the same side of the lens.

Notes:
- As an object moves toward F the image moves out and increases in size.
- When the object is between F and the lens a virtual image is formed.

Characteristics of the image

The terms used to describe the image are identical to those used for mirrors: real or virtual, upright or inverted, and smaller or larger.
In the previous examples:
In Figure 19.8**(a)**: the image is real (as the rays pass through the image), inverted, diminished, and between F and 2*f* on the opposite side of the lens.
In Figure 19.8**(b)**: the image is real, inverted, the same size, and at 2*f* on the opposite side of the lens.
In Figure 19.8**(c)**: the image is real, inverted, larger, and outside 2*f* on the opposite side of the lens.
In Figure 19.8**(d)**: the image is virtual (as the rays do not actually pass through the image), upright, larger, and between F and the lens on the same side of the lens.

NOVEL CHALLENGE

A concave mirror and a convex lens are placed in water. *Does their focal length change and if so, why?*

Images in concave lenses

The same three rays can be used to find the position and nature of an image in a concave lens. However, two important rules must be remembered:
- The rays always bend toward the thickest part of the lens.
- The focal point, the point where light rays parallel to the principal axis converge (or appear to come from), is on the same side of the lens as the object.

Figure 19.9 shows the use of rays to find the position and characteristics of the image in a concave lens.

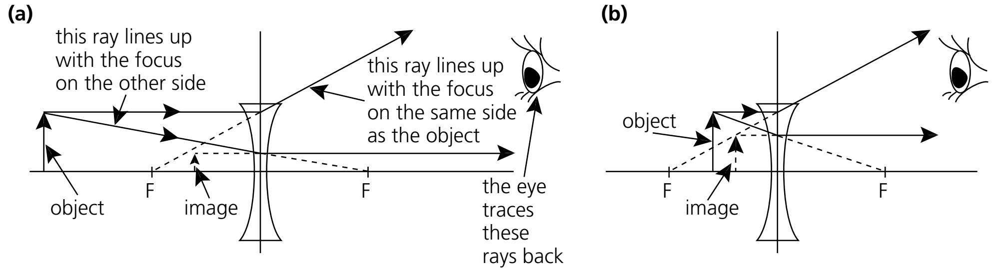


Figure 19.9
(a) When the object is at a distance from the concave lens, the image produced is virtual, upright, smaller, and between F and the lens on the same side of the lens.
(b) When the object is close to a concave lens, the image produced is virtual, upright, smaller, and between F and the lens, on the same side of the lens.

It will be noticed that whatever the distance of the object, even close to the lens, the image is virtual, upright, smaller and between the lens and F on the same side as the object. Concave lenses always produce these types of images no matter where the object is placed.

Questions

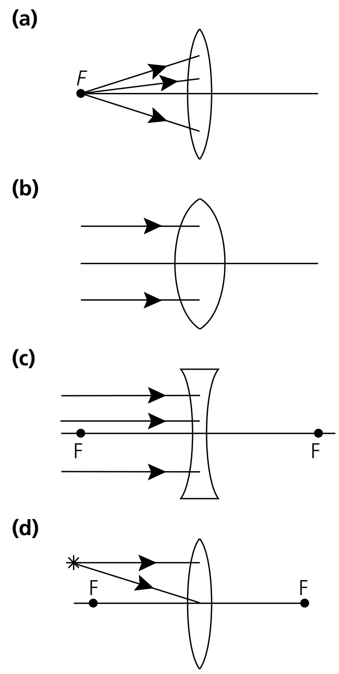
1 Complete Table 19.1 indicating the nature and position of the image as the object moves in from a long distance to a point where it is close to the convex lens.

Table 19.1 NATURE OF THE IMAGE

OBJECT POSITION	REAL/VIRTUAL	INVERTED/UPRIGHT	SMALLER/LARGER	POSITION OF IMAGE
Outside 2*f*				
At 2*f*				
Between 2*f* and F				
At *f*				
Between F and the lens				

2 The power of a lens depends on its .... and therefore its ....
3 Complete the ray diagrams shown in Figure 19.10.
4 An object 5.0 cm high is placed 20 cm from a convex lens of focal length 15 cm. Draw a ray diagram to scale to find the position and characteristics of the images.
5 Use a ray diagram to find the position of the image of an object 25 cm in front of a concave lens of 10 cm focal length. (Use a scale of 1 cm = 5 cm.)

Figure 19.10
For question 3.



19.5 THE LENS FORMULA AND MAGNIFICATION

As for mirrors, the position and size of the image can be found mathematically. (See Figure 19.11.)

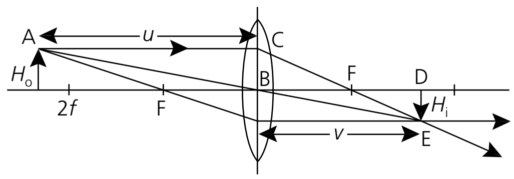


Figure 19.11
Using similar triangles the lens and magnification formulae can be derived.

NOVEL CHALLENGE

Plot a graph of v (y-axis) against M (= v/u) on the x-axis using suitable data. *Prove that the intercept on the y-axis equals the focal length. But what does the slope of the line equal? You'll be surprised and delighted.*

By using triangles as in Figure 19.11 the formula for the magnification can be derived as:

$$M = \frac{H_i}{H_o} = \frac{v}{u}$$

The **lens formula** $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ can also be derived using this figure,

where H_o is the object height; H_i is the image height; u is the object distance from the centre of the lens; v is the image distance from the centre of the lens; f is the focal length.
Note: When using the magnification formula, use the absolute values of v and u (don't worry about v being + or −).

Activity 19.1 A SIMPLE PROOF

Use similar triangles as in Figure 19.11 to derive the magnification and lens formulae.

Even though the formulae are derived using a convex lens they also apply to concave lenses but care needs to be taken in their use in solving problems. Again, since we have measurements on both sides of a lens, an order convention is required:

- All object distances (u) are positive.
- Since light passes through lenses, all image distances (v) on the opposite side of the lens (real images) are positive.
- Images formed on the same side of the lens as the object are virtual images, and their distances are negative.
- The focal length (f) of a convex lens is positive, and of a concave lens is negative.
- All measurements are made from the optical centre of the lens.

Example 1

A small 2.0 cm high candle is placed 50 cm in front of a 20 cm focal length convex lens. Find:

(a) the position of the image;
(b) the magnification;
(c) the height of the image;
(d) the position of the image of the candle if it is placed inside the focal point at a point 10 cm from the lens.

Solution

(a)

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

$$\frac{1}{v} + \frac{1}{50} = \frac{1}{20}$$

$$\frac{1}{v} = \frac{1}{20} - \frac{1}{50}$$

$$= \frac{5}{100} - \frac{2}{100}$$

$$= \frac{3}{100}$$

$$\therefore\ v = 33.3 \text{ cm}$$

The image is at 33.3 cm distance on the other side of the lens.

(b)

$$M = \frac{H_i}{H_o} = \frac{v}{u}$$
$$= \frac{33.3}{50}$$
$$= 0.67$$

(c)

$$H_i = 0.67 \times H_o$$
$$= 0.67 \times 2$$
$$= 1.33 \text{ cm}$$

(d)

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$
$$\frac{1}{v} + \frac{1}{10} = \frac{1}{20}$$
$$\frac{1}{v} = \frac{1}{20} - \frac{1}{10}$$
$$= \frac{1}{20} - \frac{2}{20}$$
$$= -\frac{1}{10}$$
$$\therefore v = -10 \text{ cm}$$

The image is 10 cm away from the lens and is on the same side of the lens as the object.

Example 2

A small light 4.0 cm high is placed 25 cm in front of a concave lens of 10 cm focal length. Find **(a)** the position of the image; **(b)** the height of the image.

Solution

(a) (*Note:* since the focal length of a concave lens is negative, $f = -10$ cm.)

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$
$$\frac{1}{v} + \frac{1}{25} = \frac{1}{-10}$$
$$\frac{1}{v} = \frac{1}{-10} - \frac{1}{25}$$
$$= -\frac{5}{50} - \frac{2}{50}$$
$$= -\frac{7}{50}$$
$$\therefore v = -7.1 \text{ cm}$$

The image is at 7.1 cm distance on the same side of the lens as the object.

(b)

$$\frac{H_i}{H_o} = \frac{v}{u}$$

$$= \frac{7.1}{25} \times 4$$

$$= 1.1 \text{ cm}$$

Questions

6 A small candle (3.0 cm high) is placed 5.0 cm in front of a convex lens of 20 cm focal length.
(a) Draw a ray diagram to find the image. Describe its nature.
(b) Use the lens formula to find the location of the image.
(c) Use the magnification formula to find the height of the image.

7 A 5.0 cm high object is placed 30 cm in front of a concave lens of 10 cm focal length.
(a) Find the position of the image.
(b) Find the height of the image.
(c) Describe the image.

8 In each of the following cases draw a ray diagram to find the position and characteristics of the image.
(a) A 2.0 cm high object is placed 10 cm in front of a 20 cm focal length convex lens.
(b) A 5.0 cm object is placed 18 cm in front of a 6.0 cm focal length concave lens.
(c) A 10 cm high object is placed 50 cm in front of a 25 cm focal length convex lens.

9 A 2.0 cm high object is placed in front of a convex lens. A real image 4.0 cm high is produced on a screen 30 cm from the lens.
(a) Calculate the magnification. **(b)** Find the position of the object.

10 Students use a 8.0 cm high candle to produce images using a convex lens. At one particular object position a 2.0 cm high image is produced on a screen placed 10 cm from the lens.
(a) Calculate the magnification.
(b) Calculate the position of the object.
(c) What is the focal length of the lens?
(d) Describe the image produced.

11 An object 5.0 cm high is placed 1.8 m in front of a convex lens. An image 6.8 cm high is produced on a screen. Find the focal length of the lens.

DEFECTS IN LENSES 19.6

Spherical aberration

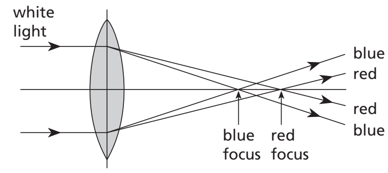
Spherical aberration is the inability of a convex lens to refract rays of light to a precise point. The word aberration comes from the Latin *aberrans*, meaning 'to wander'. This is caused by light rays that strike the outer edges of the lens being refracted more than those near the middle. This can be overcome by using a 'stop', which reduces the size of the aperture through which light can pass so that only the middle portion of the lens is used. It can also be overcome by using special parabolic lenses, or a combination of lenses as in camera lens systems.

Chromatic aberration

As discussed in Chapter 18, each colour of light is refracted by different amounts, therefore when white light passes through a prism violet light is refracted more then red light. In the case of convex lenses each colour of light is focused to a slightly different point (Figure 19.12). This produces a coloured haze around images and is called **chromatic aberration**, from the Greek word *chromo*, meaning 'colour'.

This can be overcome by using a concave–convex lens combination, as shown in Figure 19.13. Since the concave lens diverges light away from the principal axis it tends to cancel the effect of chromatic aberration produced by the convex lens. However, the two lenses need to be made of different materials, such as flint glass and crown glass, or this effect will still occur. The focal lengths of the two lenses are chosen to produce the desired focal length of the combination. The new lens is called an '**achromatic doublet**'. Today, especially in camera lens design, techniques such as multilayer coating onto the lens elements and low dispersion glasses in the lens elements all contribute to lens systems with very low aberration.

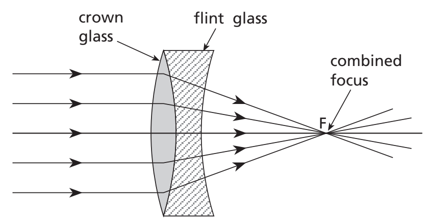
Figure 19.12
Chromatic aberration occurs when each colour of light is refracted by different amounts through a convex lens.



19.7 THE FOCAL LENGTH OF CONVEX LENSES

There are a number of practical methods used to find the focal length of convex lenses:
- You could locate the image of an object and then use the lens formula. The most accurate way is to move a small candle in until the image distance is the same as the object distance. The object and the image are then at $2f$.
- A second method is to use a light box with slits producing a number of parallel rays. The point where these rays converge after passing through the lens is the focal point.
- The third method is similar to the second method but without using a light box. This involves finding the image of a distant object — a building, a tree, or the Sun. Since rays from these distant objects are just about parallel they will converge to the focus, producing an image at the focus.

Figure 19.13
Chromatic aberration can be overcome by using an achromatic lens.



19.8 USES OF LENSES

- Magnifying glasses consist of a single convex lens. To produce a large upright image the object needs to be placed inside the focal point.
- As many optical instruments use lenses, the use of these in such things as cameras, the eye, spectacles, telescopes, microscopes, etc. will be discussed in Chapter 20, Optical Instruments.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***12** Find the power of the following lenses.
(a) A concave lens of focal length 25 cm. **(b)** A concave lens of focal length 2.0 m. **(c)** A convex lens of focal length 20 cm. **(d)** A convex lens of focal length 100 cm.

****13** What is the focal length and the type of lens that has an optical power of **(a)** –5 D; **(b)** 10 D; **(c)** 25 D; **(d)** –50 D?

****14** What type of lens **(a)** converges light; **(b)** can form upright or inverted images; **(c)** always forms a virtual image; **(d)** always forms smaller images; **(e)** is used in a magnifying glass?

****15** An object 15 mm high is placed in front of a convex lens of 12 cm focal length. The image is found 30 cm from the lens. **(a)** Find the location of the object **(i)** using a ray diagram; **(ii)** using the lens formula. **(b)** Find the height of the image **(i)** using a ray diagram; **(ii)** using the magnification formula.

**16 A magnifying glass is used to read the fine print on a legal document.
(a) Where should the document be placed in respect to the magnifying glass?
(b) If the lens has a focal length of 20 cm, and the print is 1.0 mm high, find the size of the image when the lens is placed 15.0 cm above the document.

*17 A 4.0 cm high object is placed 52 cm in front of a convex lens of 25 cm focal length. Find **(a)** the position of the image; **(b)** the height of the image; **(c)** the nature of the image; **(d)** the position of the image if this object is moved to 10 cm from the lens.

*18 A 10 cm high bulb is placed 20 cm from a concave lens of 20 cm focal length. Find **(a)** the position of the image; **(b)** the magnification; **(c)** the height of the image.

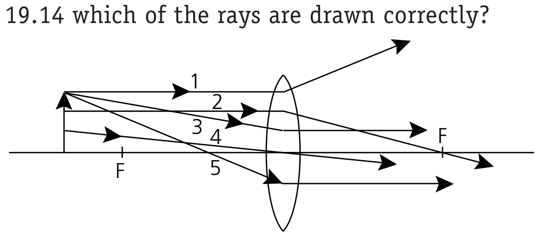
19 A thin lens of optical power +2 D was used to read the small print (2.0 mm high) in a newspaper advertisement. If the paper was placed 20 cm from the lens **(a) state what type of lens was being used; **(b)** describe the image produced; **(c)** find the position of the image; **(d)** find the height of the image.

**20 In order to take a sharp photograph of a distant object with a camera the position of the convex lens had to be adjusted until the lens was a distance of 50 mm from the film.
(a) What was the focal length of the camera lens?
(b) This camera was then used to take a photograph of a flower 50 cm from the camera; what was the distance between the lens and the film then?

*21 Describe experimentally how you would find the focal length of **(a)** a convex lens; **(b)** a concave lens.

*22 In Figure 19.14 which of the rays are drawn correctly?

Figure 19.14
For question 22.



*23 Explain how the focal length of a convex lens can be found by focusing the light from a distant building onto a screen.

**24 Students experimenting with convex lenses find the image distances corresponding to several object positions. These are shown in Table 19.2.

Table 19.2

Object distance (cm)	30	35	40	45	60	80	100	120
Image distance (cm)	150	88	67	56	43	36	33	32

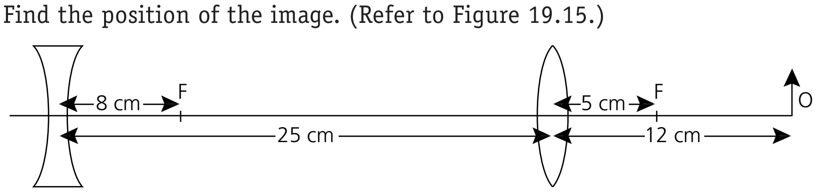
(a) Plot the graph of image distance against object distance.
(b) From the graph determine the focal length of the lens.
(c) Find the position of the image when the object is 70 cm from the lens.

*25 A philatelist (stamp collector) wishes to view stamps to identify detail in them.
(a) What type of lens would he require?
(b) Where would he need to place the stamps with respect to the lens?
(c) How would the optical power of the lens affect the image produced?

Extension — complex, challenging and novel

***26 A convex lens of focal length 5.0 cm and a concave lens of focal length 8.0 cm are placed 25 cm apart. An object is placed 12 cm in front of the convex lens. Find the position of the image. (Refer to Figure 19.15.)

Figure 19.15
For question 26.



***27 A plane mirror and a convex lens of 18 cm focal length are set up as shown in Figure 19.16. A small light source is placed 25 cm from the mirror.
(a) Describe what will be observed.
(b) Find the position of the image(s).

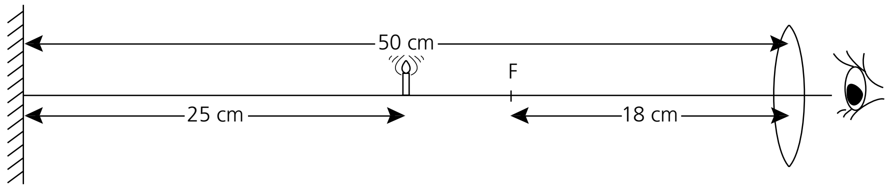


Figure 19.16
For question 27.

***28 Students performing an experiment using a convex lens placed a small candle at various distances from the lens. The images were located and the image distance for each object distance was recorded. However, when the image distances were recorded they were recorded in the wrong places in the table (Table 19.3). Rewrite the table with the image distances in the correct columns and plot the graph of object distance against image distance. From the graph find the focal length of the lens.

Table 19.3

Object distance (cm)	25	30	45	50	60	80	100
Image distance (cm)	60	27	100	36	33	25	30

***29 A slide projector is used to produce images on a screen 3.6 m from the slide. If the image is required to be 8 times as large as the object, what focal length lens is needed?

***30 A candle and a screen are placed 4.0 m apart. What focal length lens could be used to produce an image of the candle on the screen 7 times as large as the candle itself?

***31 A convex lens has a focal length of 20 cm. Where must an object be placed to produce a magnification of 5 if the image is **(a)** upright; **(b)** inverted?

***32 Students wishing to find the focal length of a concave lens found that they had a dilemma. They could not focus parallel light from a distant object onto a screen. When they used a candle they could see the image but again could not obtain an image on a screen. However, they conducted an experiment where they moved a 10 cm high candle in from 100 cm to a point 5.0 cm from the lens. At certain object distances the height of the image was judged by a number of students. The results are given in Table 19.4. Use as many of these data as necessary, generate more data from this table, plot a graph, or do calculations, to find an accurate value of the focal length of this lens.

Table 19.4

Object distance (cm)	100	80	60	40	20	10	5
Estimated height of the image (cm):							
• Student 1	1.5	2	2.5	3	5	7	9
• Student 2	1.75	2	2.5	3.5	4.5	7	8
• Student 3	1.75	2	2.5	3.25	5.5	6	8

***33 A lens forms an image on a screen with a magnification of 3.0. The screen is moved 20 cm closer and the object is then moved until the image is again in focus. The magnification is now found to be 2.5. Determine the focal length of the lens.

CHAPTER 20

Optical Instruments

INTRODUCTION 20.1

The application of artificial lenses and mirrors is widespread; however, one of the most important natural optical instruments is the human eye. But it, too, like the human body itself, has its defects and limitations. The eye not only can have inherent defects, but its performance can change with age. It is also unable to see fine detail or make observations of distant events. For these reasons scientists have developed many complex instruments that assist the eye in overcoming these limitations.

Other animals also have inherent eye defects. For example:

- chickens' eyes contain only cones thus making them unable to adapt to the dark
- owls' eyes contain only rods, which enables them to see well at night but they blink all day long. Can owls recognise colours?

NOVEL CHALLENGE

The word **pupil** comes from the Latin *pupilla* = 'a doll'. When you look at a reflection of yourself in someone's eye you see a small doll-like image of yourself. Now someone was really creative with language. *Quick now, is the image upright or inverted?*

THE HUMAN EYE 20.2

It was not until the Middle Ages that the human eye was studied in any scientific sense. Around the year 1000, an Arabian scholar, Abu Alhazen, investigated the complementary colour of after-images. (Stare at an object then close your eyes to see an image of the complementary colour.) In the early 1600s Johannes Kepler described the eye in terms of a pinhole camera but this left the world puzzled about how we saw things the right way up. Rene Descartes in 1640 proved that the eye produces inverted images (which the brain re-inverts), by experimenting with the eyes of dead oxen. From then on, theories of human vision developed rapidly.

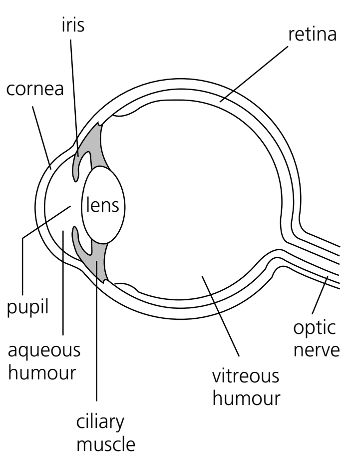


Figure 20.1
A schematic diagram of the eye.

The main function of the human eye is to form images on the **retina**. Light enters the eye through the transparent **cornea** and passes through the **lens system** to be focused onto the light-sensitive retina, which responds by sending signals to the brain via the **optic nerve**.

The amount of light entering the eye through the pupil is controlled by a diaphragm, the **iris**, which reacts to the amount of light available. In very bright daylight the iris allows little light though the pupil, which may be as small as 2.0 mm. At night the iris opens up, enlarging the pupil, which may become as big as 6.0 mm. The iris gives a person's characteristic eye colour.

EI Activity 20.1 PUPILS LOOK INSIDE THE EYE

The best way to study the optics of an eye is to cut it open. For this activity you'll need a bull's eye, scalpel, forceps, gloves and scissors.

1. Examine the eye (Photo 20.1**(a)**) and look for the optic nerve tube among the white fat and tissue at the back. Remove as much of this white stuff as you can.
2. Use a scalpel to puncture the eyeball and watch the vitreous humour ooze out.

Ch20-Walding 4th 25/8/04 12:10 PM Page 430

3 Cut the eye open completely and note the bright blue iridiscent retina. Locate the optic nerve spot. (Photo 20.1**(b)**.)

4 Cut out the lens and cornea (Photo 20.1**(c)**) and place it over some newspaper to reveal the magnification. Estimate the magnification.

5 Remove the lens and note how hard it is. Can you cut it?

(a) **(b)** **(c)**

Photo 20.1
(a) A bull's eye — as supplied by the abattoir.
(b) The retina turned inside out. The optic nerve is visible as a dark spot in the middle.
(c) A bull's eye lens and cornea with the black rim of the iris visible. The letters are magnified about 1.7 times.

The lens system

Light from objects is focused onto the retina by refraction, which takes place at the cornea, the lens, and to a lesser extent in the liquid between the pupil and the cornea — the **aqueous humour** — and the liquid between the lens and the retina — **the vitreous humour**. In fact, the curved cornea, which has an optical power of 40 D, does most of the refracting. However, it is the lens, with a power of 20 D, that allows us to focus on distant objects as well as near objects. The shape and focal length of the lens is controlled by the **ciliary muscles**. When an object is at a large distance the muscles contract, making the lens long and thin and of long focal length, which brings the object into focus on the retina. The furthest distance that objects can be seen in focus is the **far point**. This point varies from person to person and changes with age. It can be from several metres to hundreds of metres. As an object moves closer the muscles relax making the lens shorter and thicker and decreasing the focal length. The closest point at which a sharp image is formed on the retina is the **near point**. If the object is brought closer than this a blurred image is seen. For most people the near point lies between 10 cm and 25 cm. This ability of the lens to adjust, thus enabling humans to see far and near, is called **accommodation**.

NOVEL CHALLENGE

Two eyes are better than one for depth perception.
Would three be better?
Why or why not?

EI Activity 20.2 THE NEAR POINT

1 Hold this book at arm's length.

2 With one eye closed, focus on one word and slowly bring the book towards you stopping when the word starts to become blurred. This is your near point.

3 Have a partner measure this distance.

4 Repeat this with your other eye. Which eye is better?

5 What did you find about the near point for students in your class?

NOVEL CHALLENGE

You blink on average once every 5 seconds while you are awake. *How many megablinks per annum is this? When you go into a shopping centre your blink rate falls to one every 12 seconds. Propose a reason for this.*

NOVEL CHALLENGE

There are three forms of colour blindness: *protanopia, deuteranopia* and *total*. Find out what they mean then propose a reason for the names based on the meaning of the prefixes: pro = one, deut = 2.

Light rays focus on the retina, which contains light-sensitive cells called rods and cones. The **rods** react to the level of brightness whereas the **cones** react to colour as well as brightness. The cone cells are concentrated near the centre of the retina and are thought to contain three separate receptors, one each for red, green, and blue light. The rods are more concentrated on the outside of the retina, which makes colour identification in peripheral vision poor.

Cones are more active in daylight and cannot identify colour in dim light. Rods are more active in dim light. In dim light the rods secrete a light-sensitive chemical, **visual purple**, which surrounds the tips of the rods and makes them extremely sensitive to dim light. However, it is destroyed by all but red light and therefore can only accumulate in the dark. This allows people over a period of half-an-hour or so to adapt to poor light conditions.

Ch20-Walding 4th 25/8/04 12:10 PM Page 431

Activity 20.3 RODS AND CONES

1 Research:
   (a) whether other animals' eyes contain rods and/or cones
   (b) whether it is true that dogs are colour-blind. Do their eyes contain rods only?

2 Check yourself. A colour blindness test card is reproduced on the inside back cover. Cards such as these are used by optometrists to assess colour blindness.

The optic nerve carries signals from the retinal cells to the brain, which has to interpret these signals as the image formed on the retina is inverted and turned from left to right. However, there are no cells at the point where the optic nerve leaves the eyeball, which results in a blind spot.

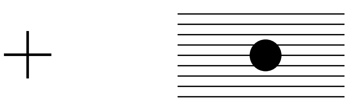
Activity 20.4 THE BLIND SPOT

Figure 20.2
Find the blind spot of the eye.

1 Close your left eye, keeping your right eye open.
2 Look at the '+' in Figure 20.2 with your right eye while holding the book at arm's length.
3 Gradually bring the book toward you while you keep looking at the '+' .
4 You will arrive at a distance at which the '•' disappears out of the corner of your eye because light rays from this point hit the retina at the blind spot — the point where the optic nerve leaves the retina.
5 Try this again looking at the '+' in Figure 20.3 this time.
6 What do you notice about the '•' and the lines when the dot is focused on your blind spot?

Activity 20.5 FLOATERS

Figure 20.3
Demonstrating how the brain tries to overcome the blind spot.



Floaters are bits of debris that come away from the inside surface of your eye.

1 Stare at something light coloured and notice the circular blobs that drift around. They tend to move upward. Why is this?
2 Stare at the ceiling and you may notice the floaters congregating at the centre. This happens as they drift into your fovea. New floaters can indicate a retinal tear.

Eye defects

Long-sightedness (Hypermetropia)

Long-sightedness is being able to see distant objects but unable to focus on near objects. It comes from Greek *hyper* = 'beyond', *metros* = 'to measure'. It is due either to the eyeball being too short, or to the inability of the lens to relax enough, so the focal length is too long and the image is formed behind the retina. (See Figure 20.4.)

This can be corrected by using a converging lens, which brings the rays closer together before passing through the eye's lens.

INVESTIGATING

Close your eyes and press an eyelid at the edge close to your nose.
How do you explain the big black circle with a yellow outline on the other corner of your eye? Hmmm!

Short-sightedness (Myopia)

This occurs when people can focus on close objects but distant objects are blurred. This is due either to the eyeball being too long, or to the inability of the muscles to contract sufficiently, making the focal length too short. The image is formed in front of the retina. (See Figure 20.5.)

It can be corrected using a diverging lens, which spreads the light rays out.

Ch20-Walding 4th 25/8/04 12:10 PM Page 432

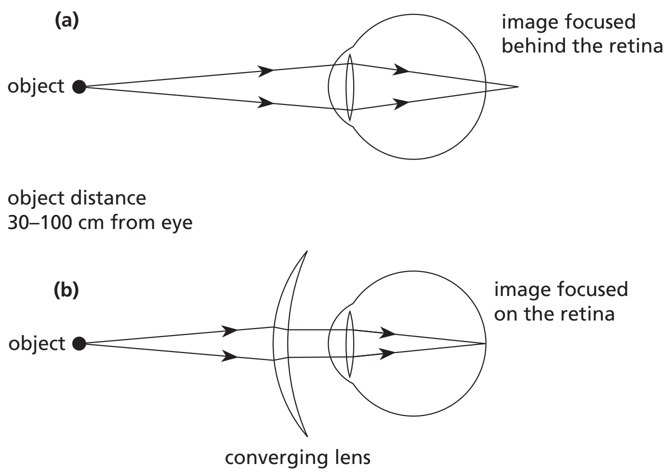


Figure 20.4
In hypermetropia the image is formed behind the retina **(a)**. This can be corrected by the use of a convex lens **(b)**.

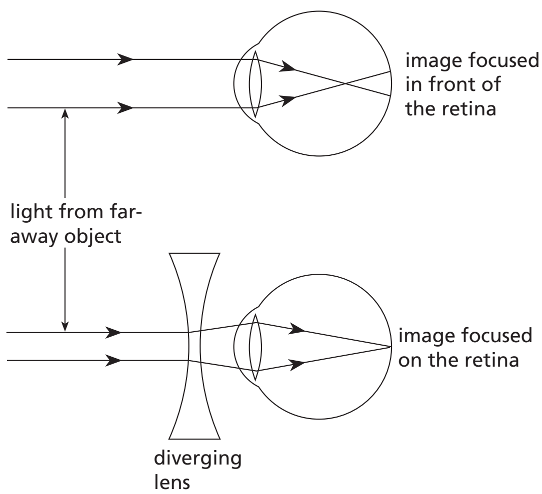


Figure 20.5
In myopia the image is formed in front of the retina **(a)**. This can be corrected by the use of a concave lens **(b)**.

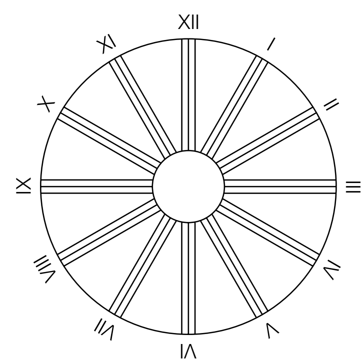
Presbyopia

Presbyopia is the inability to focus on distant or close objects — a mixture of hypermetropia and myopia. This commonly occurs with the deterioration of eyesight with age. *Presby* is Latin for 'old'. The lens is unable to change shape as the ciliary muscles weaken. Bifocals, where the top half of the spectacles are diverging lenses and the bottom half are converging lenses, are needed to correct this.

Astigmatism

This occurs if the cornea is not spherical. *Stigma* is Greek for 'point'. If the cornea is more curved in one direction than another, a person may find objects or parts of an object blurred in a particular direction whereas other parts are in focus. This can be corrected by specially shaped lenses that correct the variable curvature of the cornea.

Figure 20.6
If you have astigmatism some lines are blurred while others are sharp.



EI Activity 20.6 ASTIGMATISM

You can check to see if you have astigmatism by looking at Figure 20.6. If one line appears sharp and the others are blurred then you may have this defect.

Did you know that a Central American fish, the Anableps, has two retinas per eyeball and egg-shaped lenses? This fish swims just below the surface of the water with its large eyeballs protruding half above and half under the water. Because of the difference in refraction that occurs when light passes from air to the eye, and from water to the eye, images occur at different distances from the lens, requiring retinas at different positions.

NOVEL CHALLENGE

In 1970 Dr Fyodorov of the Soviet Union treated a near-sighted man who had glass slivers in his eyes. After the operation his near-sightedness had been cured. Propose what the 'radial keratomy' operation did.

Questions

1 **(a)** Explain why the human eye has a blind spot.
(b) What does the brain do to overcome the non-formation of an image at the blind spot?
(c) Do you think other animals would also have blind spots in their eye? (Would insects, for example flies, have many blind spots?)

2 How could you decide whether a person is long-sighted or short-sighted by examining their spectacles?

3 In the following examples state, with reasons, the likely defect of the eye, and indicate how it can be overcome.
(a) A person needs to move a book to 50 cm to be able to read it but can read street signs, etc. clearly.
(b) A student looking at a blackboard sees vertical lines in focus but horizontal lines are blurred.
(c) An older person needs to move the morning paper a long distance from his eyes to read it, and needs to sit close to the picture screen to see a movie.

4 What is the difference between the medical professionals known as an optometrist, an ophthalmologist and an orthoptist. To which of these people would you go for **(a)** spectacles, **(b)** detached retina symptoms, **(c)** lazy eye, **(d)** cataracts, **(e)** glaucoma, **(f)** conjunctivitis ?

NEI Activity 20.7 VISION EFFECTS

'See' if you can answer these questions.

1 Why does your vision persist on your retina for longer than a blink (0.1 s)?

2 Press the inside corner of your eye when closed and you will see a circular image in the other corner. You have just cut off blood flow to the retina, but why the circular image?

3 If you are long-sighted and hence need reading glasses, you may find that squinting makes close-up things clearer. Does this work by narrowing the effective pupil size and giving you a 'pinhole effect' or does the pressure on your eyeball change its shape somewhat so that it temporarily corrects the defect?

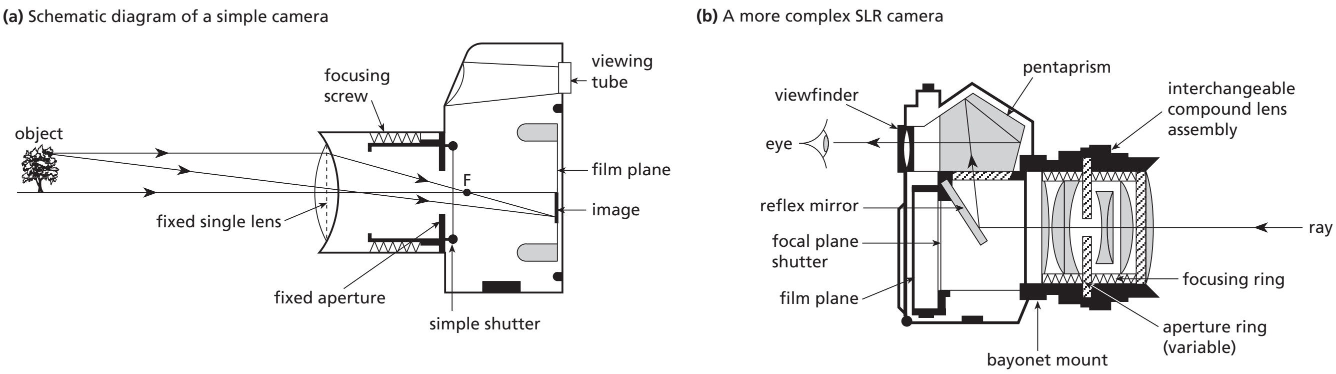
4 The fovea has an enormous number of cells (about 250 000). Why is this?

5 Contact lenses are used mainly for short-sightedness. For long-sightedness you would need a concave contact lens. Can you design one that fits the curve of the eyeball?

6 The eye disease 'prosopagnosia' has a strange effect. If you look at a hollow face mask on a rotating table it appears to be rotating the wrong way when its inside surface is viewed. Prosopagnosia prevents this. Can you find an explanation for this effect?

CAMERAS 20.3

A camera works in a similar fashion to the eye. Light from the object enters through the lens system and is focused onto the film containing light-sensitive chemicals (silver halides), which are changed by the image-forming light rays. This is called making an **exposure**. A correct exposure is made when sufficient light is allowed to enter the camera and cause the correct amount of chemical change in the film. This is controlled by both shutter speed and aperture diameter.



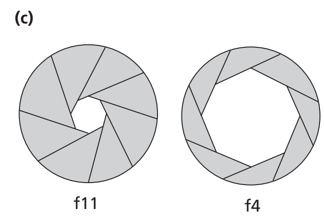


Figure 20.7
A schematic diagram of a simple camera **(a)**; and a more complex SLR camera **(b)**; **(c)** the variable aperture ring.

Unlike the **lens** of the eye, the lens of the camera cannot change its focal length. Therefore, to form a sharp image of objects at various distances the lens is movable. When objects are far away, the distance between the lens and the film is the focal length. As a distant object moves nearer, the lens has to be moved further from the film to keep the object in focus. This is achieved by use of the **focusing screw**. However, there is a limit to the closeness an object can reach while still in focus, for each particular type of camera and lens. Single lens reflex cameras have the ability to easily change lens systems. These lens systems often have a range of focal lengths over which they can operate. They are called **zoom** lenses. Camera lenses are made from glass in good-quality cameras, but in cheaper ones plastic (usually Perspex) is used.

The **shutter** controls the amount of light entering the camera. If objects are moving quickly then it is necessary to catch the object in one position otherwise the image will be blurred. Therefore, the time interval that the shutter is open needs to be very short, maybe 1/500 second. Because less light is entering the camera this will necessitate the use of a 'fast' film. These films are more sensitive to light than 'slow' films, which can be used when the shutter speed can be longer, allowing more light to enter the camera. However, 'fast' film may not give the picture quality of 'slow' film. It is usually more grainy. The speed of any film is stated as its ISO rating. An ISO400 film is four times as fast as an ISO100 film. In modern films the ISO rating is coded onto the film canister to be 'read' by the camera.

All cameras are fitted with an **aperture**, which restricts the area through which light can pass. Simple cameras have fixed apertures, but other cameras have a variable iris diaphragm (corresponding to the iris of the eye) in the lens system. As the aperture is nearly circular, its area is proportional to the square of its diameter. So when you double the diameter, you let through four times as much light; when you halve it, you let through one-quarter the amount of light.

Apertures are denoted by f-stops or f/numbers, calculated by dividing the focal length of the lens by the 'effective aperture'. The effective aperture is the width of the light beam that must enter the lens to fill the real aperture (or 'stop') completely. It is generally not the same as the diameter of the aperture itself, as most lens front elements are positive and therefore converge a beam of light to fill the aperture. The f/number series is as follows: f1, f1.4, f2, f2.8, f4, f5.6, f8, f11, f22, f32, and f45. Each number admits twice as much light as the next higher one, and half as much as the next lower. The bigger the number, the smaller the aperture, and vice versa. Most lenses allow you to set half-stops, although the f/number is not shown, unless the camera is an electronic type with an LCD panel or the f/stop is the maximum aperture.

Under the same lighting conditions, all lenses, regardless of focal length, produce the same image brightness at the same f/number.

Photo 20.2

The aperture setting and the shutter speed affect the photo produced.

Lenses are described by two parameters: their focal length and their maximum aperture. So you might have a 35 mm f2.8 lens, or a 28–70 mm f3.5–4.5 zoom. Note that the maximum aperture of the zoom has changed with focal length, and the numbers quoted here are not full stops.

It's common to talk about 'the speed of the lens', which is pretty confusing, as it has nothing to do with shutter speed. The speed of the lens is its maximum aperture; a fast lens is one with a big maximum aperture for its focal length (maximum apertures tend to become smaller as the focal length increases). So the so-called 'sports lenses' are 300 mm f2.8 telephotos and are usually very expensive.

The size of the aperture also controls the **depth of focus** or **depth of field**. If a small aperture is used objects from far and near will be in focus at the same time, giving a large depth of field. If a large aperture is used then a small depth of field is obtained, with near objects being in focus while background objects are blurred. (Compare the two photos.) When you use the lens's maximum aperture you are said to be shooting 'wide open', and when using its minimum you are shooting 'stopped down'.

For correct exposure when taking photographs, the shutter time and aperture settings will depend on:

- the brightness of the object
- the sensitivity of the film
- the effect or depth of field desired.

The correct exposure will be determined by the camera's light meter.

NEI Activity 20.8 CAMERAS AND FILM TYPES

1 Using an encyclopaedia or other reference, find out how a pentaprism in a camera works and make a drawing to illustrate.

2 Use the library or Internet to research **(a)** types of films; **(b)** types of camera systems: compact, SLRs, TLRs, medium format, view cameras, digital cameras.

Questions

5 For a simple camera to produce a photo of a distant object:
(a) state whether the lens will be moved closer to or further from the film;
(b) give an indication of the aperture size required for a good depth of field.

6 Aperture settings of cameras are in terms of f/numbers or f-stops.
(a) What do these numbers marked on the lens indicate?
(b) What is the size of the effective aperture of a 10 cm focal length camera lens if the setting is f4?
(c) An exposure of $\frac{1}{125}$ s at f8 would be equivalent to an exposure of $\frac{1}{250}$ s at what aperture, assuming that film of the same speed is used for both?

7 A lens of a simple camera has a focal length of 12 cm. What is the distance needed between the lens and the film to produce a sharp image when the object is at **(a)** infinity; **(b)** 40 cm; **(c)** 6.0 cm?

TELESCOPES 20.4

Controversy exists over the actual inventor of the telescope although it is usually attributed to a Dutch spectacle maker, Hans Lippershey, in 1608. However, it was Galileo who quickly made his own version in Venice in 1610 and turned it toward the heavens. The modern science of astronomy was born. As Venice was a glassmaking city, the developments of lenses became rapid. The microscope was invented soon after.

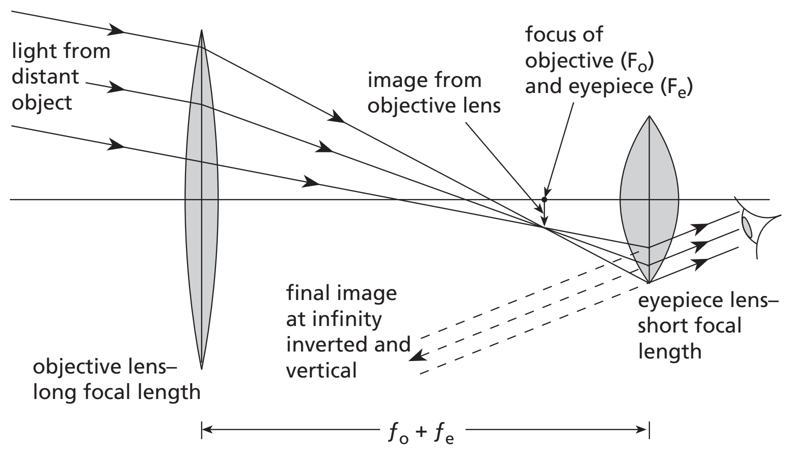


Figure 20.8
A refracting telescope contains a long focal length objective lens and a short focal length eyepiece lens.

INVESTIGATING

Binoculars solve one problem inherent in the design of telescopes. What is it?

This type of telescope consists of two convex lenses. The objective lens has a long focal length and the eyepiece lens has a short focal length. These lenses are set up so their focal points coincide (Figure 20.8). Parallel light from a distant object, after being refracted by the objective lens, forms an inverted image at the focal point of the eyepiece lens. This produces a final image at infinity. The eye uses its own lens to bring the parallel rays to focus on the retina. Notice that the image is inverted, but astronomers are accustomed to this. The magnification of a telescope is given by the formula:

$$M = \frac{f_o}{f_e}$$

where f_o is the focal length of the objective lens, and f_e is the focal length of the eyepiece lens.

The length of the telescope is also controlled by the focal lengths. The telescope has a minimum length of $f_o + f_e$.

Terrestrial telescopes (in Latin *terra* = 'Earth') are a lot smaller than astronomical telescopes and have an extra lens to produce an upright image, which is important when observing earthly things such as a cricket match or while bird watching (Figure 20.9).

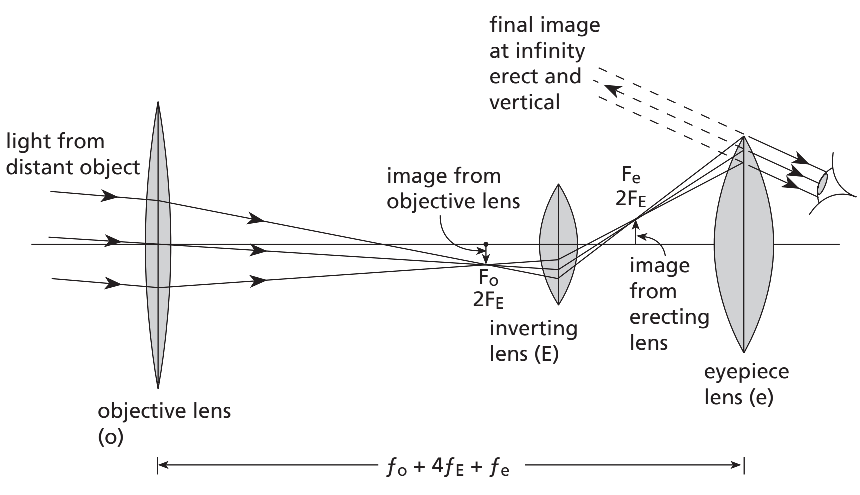


Figure 20.9
Terrestrial telescopes have an extra lens so images are upright.

Galilean telescopes use a diverging lens as the eyepiece, which in effect makes the length smaller and produces an upright image.

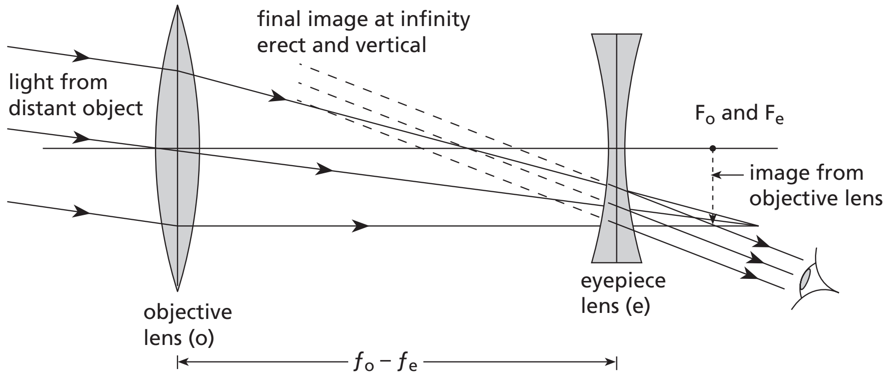


Figure 20.10
A Galilean telescope's eyepiece is a concave lens. This produces upright images but gives a narrower field of view.

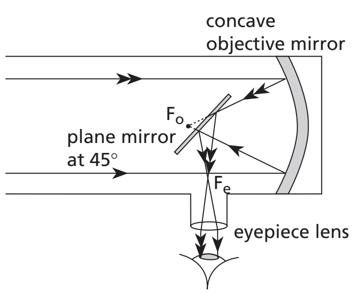
However, astronomical telescopes are normally not refracting telescopes as the large lenses required tend to sag and it is very expensive to manufacture good quality, large focal length lenses. Astronomical telescopes are normally reflecting telescopes.

Reflecting telescopes

One of the problems associated with large refracting telescopes is chromatic aberration. Isaac Newton realised that it would be difficult to avoid, so in 1665 he developed a reflecting telescope. It used a large concave mirror to collect light from distant objects. This light converged onto a plane mirror tilted at 45°; this reflected the light to the focal point of the eyepiece, which acted in the same way as that of an astronomical telescope. These telescopes have an advantage in that mirrors are less costly than lenses and can be supported at their backs, therefore they can be very large.

The world's largest optical mirror, polished to the highest perfection, has a diameter of 8.2 m, a surface of more than 50 m^2, and weighs 23.5 t. It was built in 1995 and put in place on 17 April 1998 at the first Very Large Telescope Unit (VLT-UT1) at the Paranal Observatory located on Cerro Paranal in the Atacama Desert, Northern Chile. The VLT now consists of four 8 m telescopes which can work independently or in combined mode, providing the total light-collecting power of a 16 m single telescope, currently making it the largest optical telescope in the world. The VLT is strong enough to detect a glow-worm at 10 000 km. Compared with this, the largest objective lens used in a refracting telescope is quite small — 1.34 m in diameter. Because of the very large diameter mirrors used in reflecting telescopes, they are able to collect large amounts of light, allowing astronomers to see fainter objects in the heavens.

Figure 20.11
Reflecting telescopes use large concave mirrors as the objectives. This allows them to collect much more light than refracting telescopes.



Questions

8 Copy and complete Table 20.1, which indicates the focal lengths of the objective and eyepiece lenses, and the magnification obtained by a refracting astronomical telescope.

Table 20.1

	FOCAL LENGTH OF OBJECTIVE LENS	FOCAL LENGTH OF EYEPIECE LENS	MAGNIFICATION	LENGTH OF THE TELESCOPE
(a)	4.0 m	10 cm		
(b)		5.0 cm	10	
(c)	3.0 m		60	
(d)		2.0 cm		2.0 m

9 Give reasons for most astronomical telescopes being reflecting rather than refracting telescopes.

20.5 MICROSCOPES

A **simple microscope** consists of a single convex lens or magnifying glass. When the object is placed inside the focal length a large upright image is produced. The lens is moved until the largest sharpest image is seen. This occurs when the image is at the 'near point' of the eye.

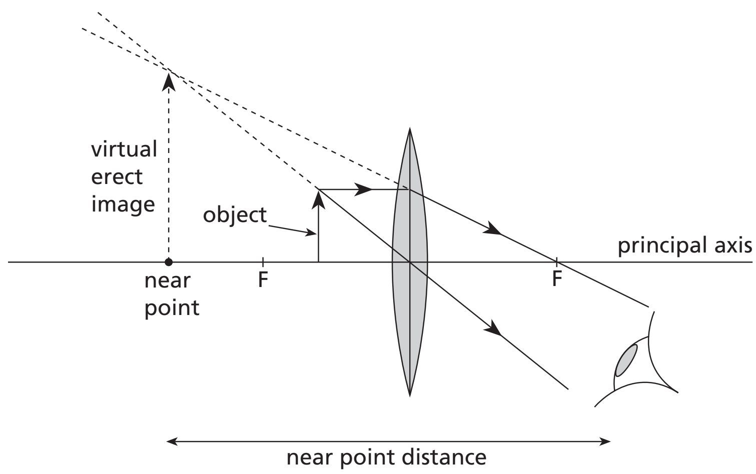


Figure 20.12
A simple microscope produces magnified, upright images at the near point of the eye.

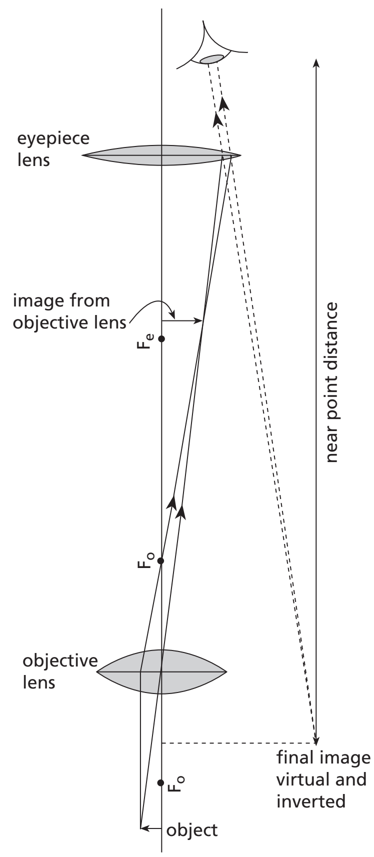
The magnification of the magnifying glass is expressed as the ratio of the near point distance to the focal length of the lens. In general, the near point is taken as 25 cm.

$$M = \frac{25}{f}$$

For example the magnification of a convex lens of 10 cm focal length, when used as a simple microscope, is $\frac{25}{10}$ = 2.5.

A **compound microscope** consists of a converging lens of short focal length (5.0–10 cm) and an eyepiece lens of longer focal length. The light from the object, which is illuminated by reflected light from a mirror or by direct light, passes through the objective lens, which is positioned so that the object is just outside its focal point. This produces a large inverted real image just inside the focus of the eyepiece lens, which acts like a magnifying glass. A large inverted, virtual image is thus produced at the near point of the eye. Since both lenses magnify, magnification of 400 or more can be obtained. Microscopes usually have a variety of objective lenses to produce a range of magnifications and good-quality microscopes contain achromatic lenses to eliminate chromatic aberration.

Figure 20.13
A compound microscope uses a short focal length objective lens and a long focal length eyepiece lens to produce upright, magnified images at the near point of the eye.



Questions

10 Find the magnification of a convex lens when it is used as a simple microscope if the focal length is **(a)** 20 cm; **(b)** 2.0 cm; **(c)** 100 cm; **(d)** 5.0 cm.

11 Cheap microscopes produce images with coloured fringes. Explain why this occurs and how it is overcome in more expensive microscopes.

PROJECTORS AND ENLARGERS 20.6

A **slide projector** (Figure 20.14) consisting of a system of mirrors and lenses produces a real, magnified, inverted image of a piece of photographic slide on a screen. As light leaves the lamp in all directions, the concave mirror reflects the light backward through the slide, intensifying the light through the slide. The **condenser** lens spreads the light evenly over the surface of the slide. The movable **projector lens** focuses a sharp, magnified, inverted image onto a screen. Because the image is inverted, the slide has to be placed in the projector upside-down.

Figure 20.14
A schematic diagram of a slide projector.

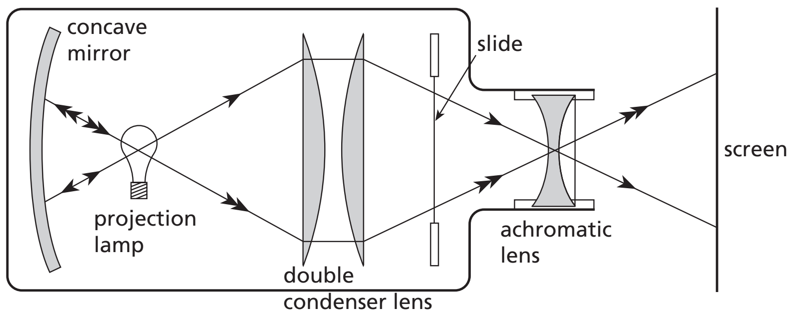
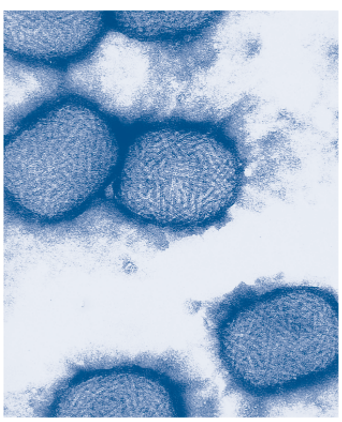


Photo 20.3
Really high magnifications can only be obtained using a scanning electron microscope. This photo shows a Pigeon Pox Virus at 160 000× magnification. Each viron is 200 nm across so you'd fit 1/4 million across the width of this photo. Taken by Mr Howard Prior at the Animal Research Institute, Yeerongpilly, Brisbane.



A **photographic enlarger** is very similar to the slide projector with the slide replaced by a film negative and the screen replaced by a piece of photographic paper containing light-sensitive chemicals, which change when light falls on them. The amount of light emerging from the enlarger can be controlled by adjusting the aperture (the f/number).

The latest versions of these instruments are referred to as **data projectors**. You might like to look back at Chapter 16 (section 16.10) to find out about the electronics and optics in these devices using LEDs and DLPs.

NEI Activity 20.9 OVERHEAD PROJECTOR

1 If you have access to an OHP, examine the optics and draw a diagram to show the way an image is produced.

2 Locate the Fresnel lens and describe its construction.

3 State one safety feature incorporated into the OHP.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

*12 Why do you think doctors shine a light in your eyes to see if you are conscious?

*13 With age, as the muscles weaken, what do you think happens to the far point?

*14 Draw a well labelled diagram of the human eye.

*15 (a) What is hypermetropia?
(b) Explain how this defect of the eye might be caused.
(c) Indicate how it can be overcome.

***16** For the human eye what controls:
(a) the focusing of an image on the retina;
(b) the focal length of the lens;
(c) the amount of light entering the eye;
(d) whether you can see in dim light?

***17** **(a)** Explain the operation of the shutter and the aperture in allowing light into a camera in order to control exposure.
(b) Explain the difference between film speed and lens speed in photography.

***18** Compare the similarities and differences between an eye and a camera by copying and completing Table 20.2.

Table 20.2

FEATURE	EYE	CAMERA
Detects an image		
Focuses light to produce an image		
Controls the amount of light entering		
Controls mechanism to produce a field of view		
Sensing apparatus/material that produces the image		

***19** What would happen to the length and magnification of a refracting telescope if the objective lens was replaced by one of shorter length?

***20** The focal length of the eyepiece lens of an astronomical refracting telescope is 4.0 cm. What focal length objective lens is required to give the telescope a magnification of 10?

***21** What is the focal length of a magnifying glass whose magnification is **(a)** 25; **(b)** 50; **(c)** 10?

***22** Compare the similarities and differences between a refracting astronomical telescope and a microscope.

***23** Specify whether the following use a concave or a convex lens.
(a) The eye.
(b) Spectacles to correct myopia.
(c) The projector.
(d) The terrestrial telescope.
(e) The microscope.

***24** Why must a slide be placed upside-down in a slide projector?

***25** In an enlarger what is the purpose of **(a)** the concave mirror; **(b)** the condenser; **(c)** the lens?

****26** A slide projector contains a lens of 5.0 cm focal length. The distance between the slide and the lens can be adjusted from 5.1 cm to 6.0 cm.
(a) Calculate the minimum distance between the lens and the screen that produces a sharp image on the screen.
(b) Calculate the maximum distance between the lens and the screen that produces a sharp image.
(c) If the size of the slide is 7.8 cm^2, calculate the maximum size of the sharp image produced.

NOVEL CHALLENGE

A searchlight lights up the sky when used on Earth. What would you see if you did the same thing on the moon?

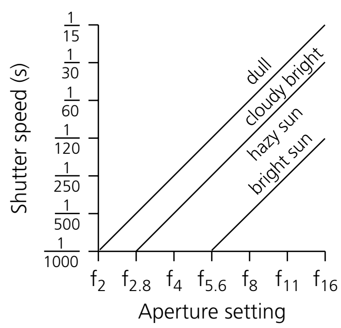
**27 Which of the following form virtual images when in normal use?
(a) The eye.
(b) The camera.
(c) The refracting telescope.
(d) A Galilean telescope.
(e) An enlarger.
(f) A microscope.

**28 Several eye defects are now being treated by the use of lasers. Research the use of lasers in modifying the cornea to enable people to see without spectacles.

29 If you wanted to start a fire without matches you could use your friend's spectacles to concentrate rays of light from the sun. Is this possible if your friend has **(a) hypermetropia; **(b)** myopia; **(c)** presbyopia? Explain each answer!

**30 Figure 20.15 can be used in determining the correct exposure when taking photos using a fast film. Use this figure to determine (with explanation) what an appropriate setting would be when the camera is used to take a photo of:
(a) a fast moving car in bright sunlight;
(b) a bird sitting in the branches of a tree;
(c) a flower in the field on an overcast day;
(d) a wedding couple on a bright sunny day.

Figure 20.15
For question 30.



**31 In William Golding's *Lord of the Flies* the character Piggy uses his glasses to start a fire. Later the boys break Piggy's glasses but he cannot identify them at close range because he is short-sighted. Find the flaw in this narrative.

Extension — complex, challenging and novel

***32 A coin collector wants to examine some old coins using a magnifying glass. If the magnifying glass has a focal length of 10 cm, where must the eye and coin be positioned to produce the largest sharpest image?

***33 A student who is far-sighted needed to hold a book at a distance of 50 cm to be able to read. What focal length corrective lens is required to enable this student to read the book at the normal near point of 25 cm? Assume the distance between the lens and the eye is negligible.

*****34** Figure 20.16 shows the schematic diagram of an 'episcope' used to produce an image of a page of a book on a screen. Analyse the figure and explain how it operates.

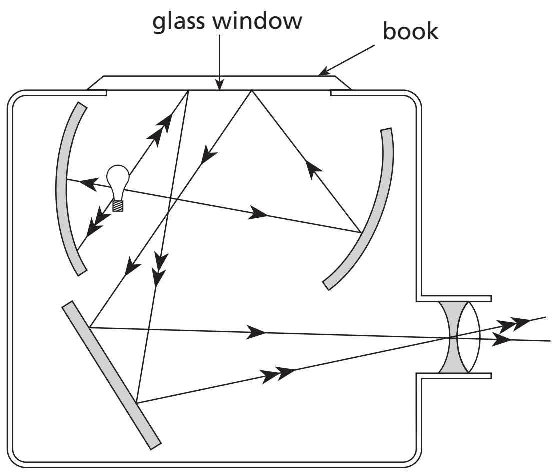


Figure 20.16
For question 34.

*****35** When you hold a pinhead close to your eye and let it be illuminated by light from a pinhole in some paper (see Figure 20.17) you see a strange image. The image is dark, magnified and inverted. Try it and offer some explanations.

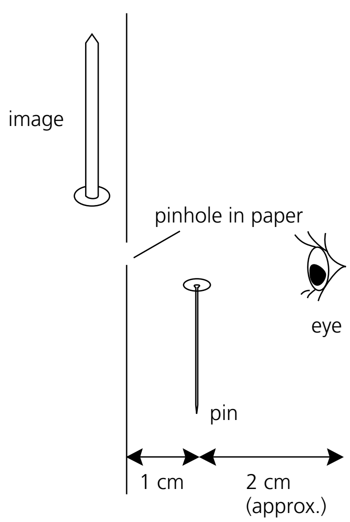
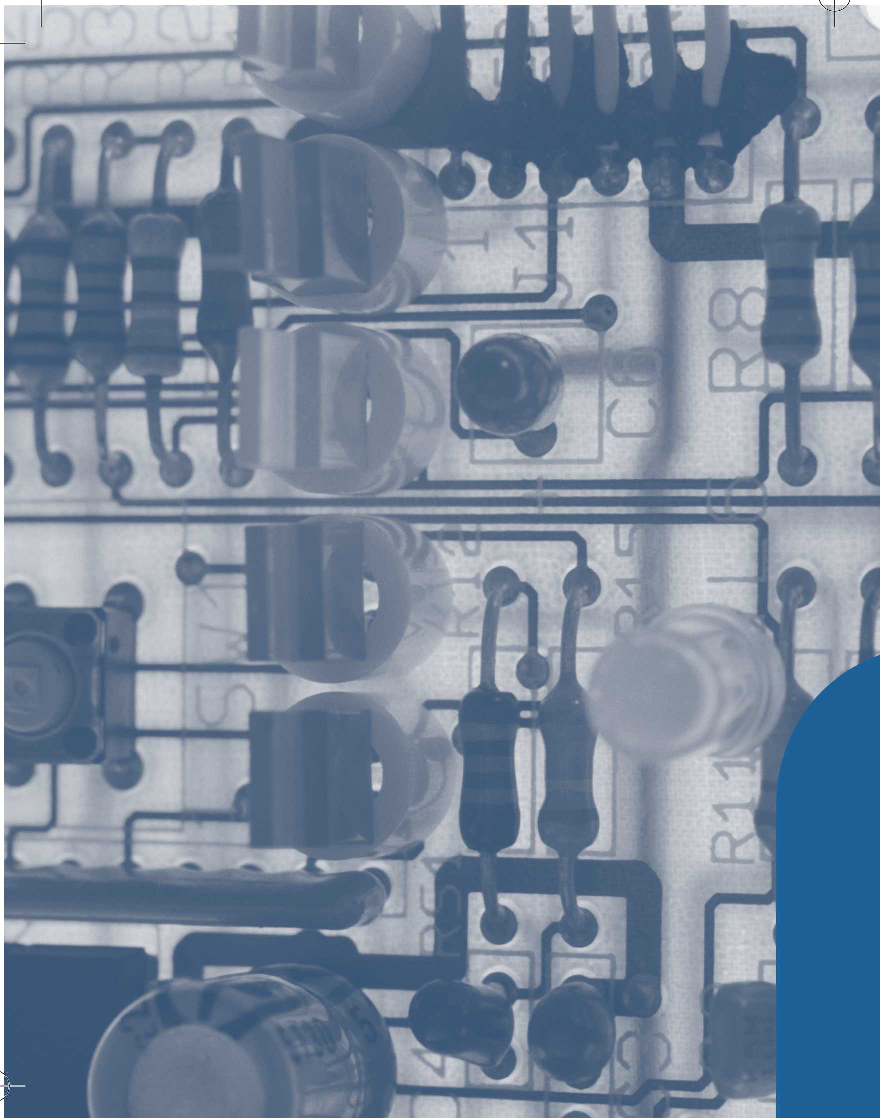


Figure 20.17
For question 35.



UNIT 07

Electricity & Electronics

CHAPTER 21

Electrostatics

21.1 INTRODUCTION

Electrostatic effects are very common in our everyday experience. Have you ever wondered about peculiar electricity effects? For example, what do the following situations have in common?

- You've greeted your best friend in the library after wandering around looking for books and you both get an electric shock as you touch.
- Removing a synthetic shirt or blouse at night in a darkened room leads to a display of tiny little sparks.
- After travelling in the car you receive a rather nasty little electric shock from the door handle when you get out.

Did you realise that these effects all occur after one type of material has been rubbed against another? Other examples of similar effects:

- Dust always seems to stick to the screen of the television set or computer monitor and gets worse as you try to take it off with a cloth.
- Some cars travel around with funny belts hanging down onto the road.
- Lightning always strikes the oldest and tallest trees in the forest.
- Some groups of balloons never seem to want to stay together properly.
- People sitting on plastic chairs in the office pose a possible danger to sensitive computer or electronic equipment.

These are all due to the presence of electric charge built up on objects around us as a result of frictional processes. This electric charge is very important in nature. Electric forces and charges control many natural effects and are seen in dramatic circumstances such as lightning strikes. Much of our modern technology relies on controlling electric charges, either trying to eliminate their effects or making use of their attracting or repelling properties. In this chapter the aim is to understand the nature of electric charge and the ways in which charge behaves. This will help us to understand the operation of application devices such as spark dischargers, electrostatic generators, photocopiers and fax machines, lightning arrestors and even the various forms of biological electrostatic defences such as those possessed by animals like the electric eels and rays. Physicists regard the force of electricity as a fundamental force of nature that is ultimately responsible for other forces such as **friction**, contact pushes, **adhesion** and **cohesion**.

21.2 ELECTRIC CHARGE

The first step in understanding electrostatic effects is a knowledge of the structure of atoms and matter. Recall that all matter is composed of atoms, which are the building blocks consisting of a very small and dense central nucleus, containing **protons** and **neutrons**, and layer-like regions called clouds surrounding this nucleus, which contain **electrons**. Figure 21.1 shows the relative diameters of a typical atomic nucleus as well as the outer electron cloud

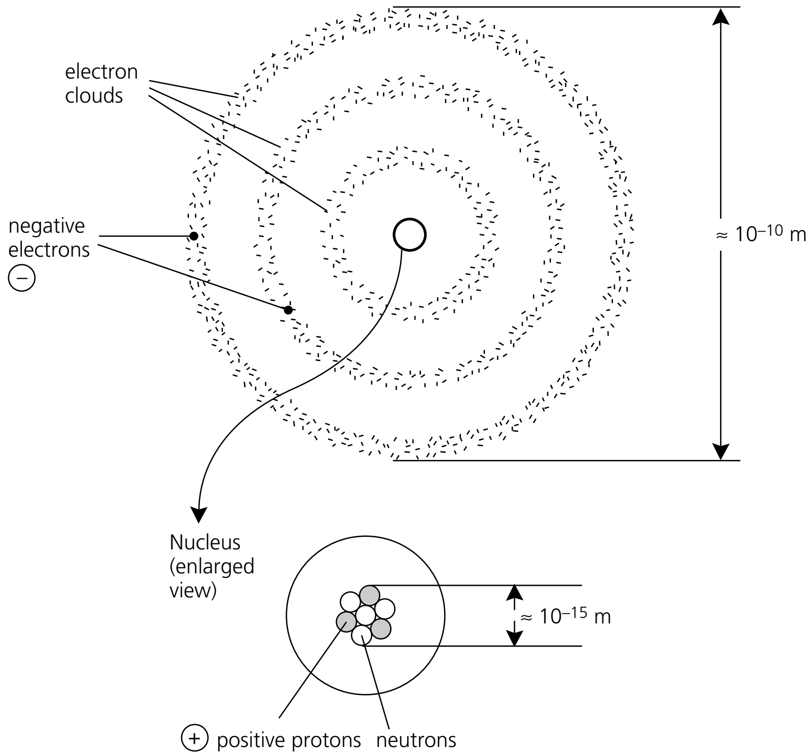


Figure 21.1
Simple model of the atom.

for a general atom. From this type of model it is possible to conclude, as did Lord Rutherford in about 1913 using alpha particle scattering experiments, that the atom is mostly empty space. It is also possible to conclude that the particles within the nucleus are very tightly bound together, with the protons being positively charged and the neutrons being **neutral**, that is, with no net charge. The electrons, especially the outermost ones, are very loosely bound to the nucleus in most atoms and are negatively charged.

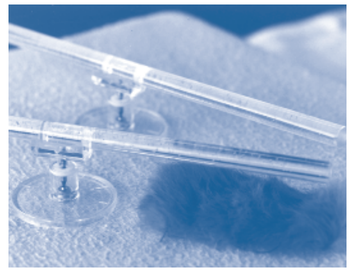
Table 21.1 ATOMIC PARTICLE PROPERTIES

	ELECTRON	PROTON	NEUTRON
Relative charge	−1	+1	0
Coulomb charge	-1.6×10^{-19}	$+1.6 \times 10^{-19}$	0
Mass	9.11×10^{-31} kg	1.673×10^{-27} kg	1.675×10^{-27}
Atomic location	orbital cloud	central nucleus	central nucleus
Discovered	1897 J. J. Thompson	1913 E. Rutherford	1932 J. Chadwick

Table 21.1 compares the properties of the three fundamental atomic particles. Any atom that has equal numbers of protons and electrons is said to be neutral as it will have no net electric charge. Remember that the neutrons act simply as a nuclear glue and do not alter the positive–negative balance of the atom. The simplest atom in nature, the hydrogen atom, only has one electron and one proton, whereas lawrencium, a very complex atom, has 103 electrons, 103 protons and 157 neutrons. If an individual atom of any element is made to gain or lose some of its electron particles then it is said to have become an **ion**. Positive ions have lost electrons and negative ions have gained electrons over and above the normal atomic number. Ions are quite important in various types of chemical reactions.

In electrostatics it is normal to consider blocks of materials that have become **electrically charged**; this means that the material itself, such as a glass rod, has either had extra electrons placed onto it, or had some of its atomic electrons removed from it. It has thus become net electrically charged or electrified. The ancient Greeks had found that the material they called 'Elektros' attracted bits of hair and small pieces of straw dust when it was rubbed with cloth or animal fur. Today we call this material amber. It is actually fossilised tree resins. Similarly, in laboratory experiments it is easy to show that Perspex rods rubbed with rabbit fur can attract small torn pieces of paper, or that polythene strips if rubbed with the same rabbit fur will subsequently repel each other when freely suspended on a set of free pivots. This type of experiment is often best carried out under a set of heat lamps so as to provide a very dry atmosphere in order to prevent moist ionised air quickly dissipating the electric charges.

Photo 21.1
Free pivot apparatus with charged rods.



This technique of using a cloth or piece of fur to rub a solid such as glass, Perspex, wax or polythene will electrify the object due to a process called 'friction charging'. In this process the energy supplied to the outermost atomic electrons allows them to move from the material with the least affinity or attraction for electrons to that material with the most affinity for electrons. The process is also referred to as **triboelectric** separation of charge. The word is derived from the Greek *tribein*, meaning 'to rub'. Electrons are therefore transferred from one object to another, one object becoming positive as it loses electrons, say the rabbit fur, and the other object becoming negative as it gains electrons, say the polythene strip. Note that in this example of the separation of charge process, the fur will most likely lose its charge quite quickly either by direct contact with the experimenter's hand or by loss to the atmosphere. It will thus regain neutrality. This quite often makes it difficult to show that the fur has in fact become electrically charged.

Table 21.2 TRIBOELECTRIC SERIES

+	Atoms have a low affinity for electrons
	Rabbit or cat fur
	Acetate
	Perspex
	Glass
	Wool
	Lead
	Silk
	Paraffin wax
	Polythene
	Ebonite
	Copper
	Rubber
	Amber
	Sulfur
	Gold
−	Atoms have a high affinity for electrons

Every material's atoms have their own specific tendency to gain or lose electrons easily. Table 21.2 lists the triboelectric series showing several materials in order, from those that have a low affinity for electrons and will tend to become positively charged to those that have a high affinity and become negatively charged in frictional experiments. This series is easy to read because any material will become positive by losing electrons if rubbed with any other material lower in the series list. For example, glass can become positively charged when rubbed with a silk handkerchief but negatively charged if rubbed with rabbit fur. Acetate or Perspex rods can become positively charged if rubbed with a woollen cloth while polythene or ebonite rods will become negatively charged when rubbed with the same woollen cloth.

In the triboelectric separation process the frictional charging simply involves a transfer of negative charge or electrons from one object to another. It is important to realise that

NOVEL CHALLENGE

In the mid-1700s, French experimentalist François du Fay observed that a charged gold leaf was attracted by some electrified substances and repelled by others. He called the two types 'vitreous' and 'resinous'. Use Table 21.2 and your knowledge of what substances are classified as vitreous and resinous to decide if the resinous rod would have a positive or negative charge.

PHYSICS FACT

In the eighteenth century, British sailing ships had their gunpowder store (the 'magazine') lined with copper to make it waterproof. Sailors had to put on thick felt slippers to avoid generating a spark by electrification. They learnt this the hard way.

NOVEL CHALLENGE

In 500 BC, Greek philosopher Thales of Miletus noticed that cork dust was attracted to a charged amber rod. It was not until 1500 (almost 2000 years later) that people noticed that after a while the cork dust was repelled.

Why was the dust repelled and why do you suspect that people did not notice it earlier?

objects can never gain or lose positive protons in this process. The net amount of charge lost by one object of the pair is gained by the other. We can state a more general law of **conservation of charge** which is:

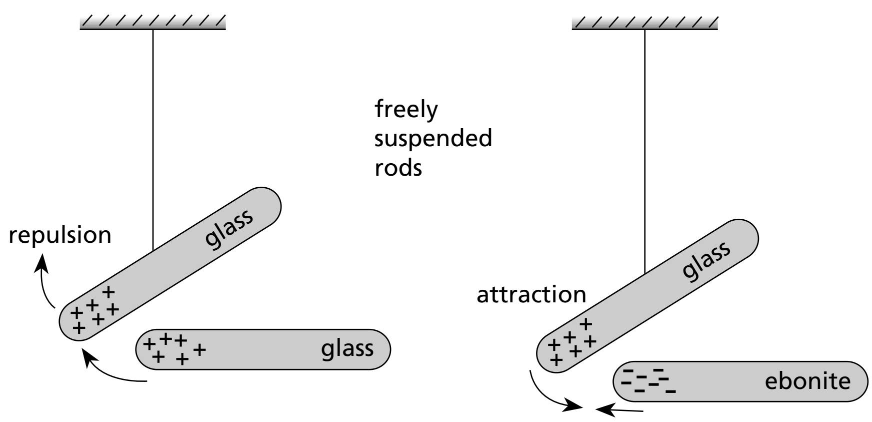
The net amount of charge produced in any transfer process is zero.

Electric charge is considered to be one of the fundamental properties of matter. Recall that in earlier chapters on force and space physics the idea of an inverse square law of attraction between two masses was established. This force of gravity depended on the value of the respective masses and was inversely proportional to the square of the separation of the masses. Similarly, a force of attraction or repulsion can be detected between electrically charged objects. This force due to electrostatic charge also obeys an inverse square law and depends on the respective amounts of charge on the objects just like the gravitational force between masses. The mathematical form of the **law of electrostatics** is further discussed in Section 21.5 but at this point the most important general features of the law are illustrated in Figure 21.2 and stated as:

Like charges repel and unlike charges attract.

The effects of this law can be observed with charged rods freely suspended on pivot apparatus.

Figure 21.2
Like charges repel and unlike charges attract.



TEST YOUR UNDERSTANDING

Plastic food wrap (e.g. *Gladwrap*) clings because it is electrostatically charged during manufacture.

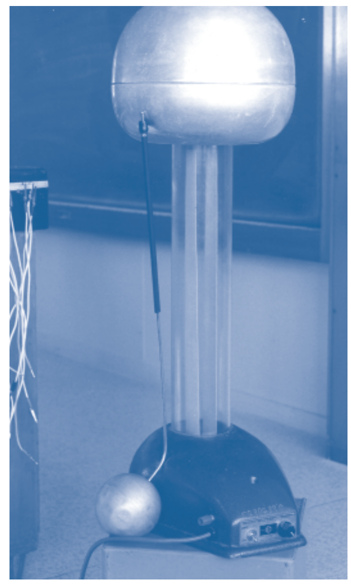
If this is the case, why doesn't it discharge when you touch it with your hands?

The process of frictional charging can be used in the development of machines that generate and store large quantities of electric charge. Research and experimentation within fields such as materials science, sub-atomic structure of matter, nuclear medicine, atmospheric physics and meteorology require electrostatic generators that provide very powerful force fields and electrostatic potential differences. These topics are further discussed in Section 21.6. A school laboratory machine that usually is a smaller model of the larger more complex research machines is the **Van de Graaff** generator. This device consists of a polished metallic dome that is supported on an insulating column. Charge is transported to or from this dome by a rubberised belt running over a pair of rollers and a system of point combs or knife-edge strips placed very close to the moving belt at the top and bottom. A high-voltage electrical system within the base of the machine continuously provides electrons and a motor rotates the rubberised belt so that charge is moved. As the Van de Graaff generator is operating, charge builds up on the metal dome, eventually producing a very high potential difference between the dome and the Earth. This may reach thousands or even millions of volts. As more charge is moved to the dome the motor has to work harder and reduces in speed. An equilibrium is established when charge movement onto the dome is balanced by charge loss to the surrounding moist air. If any neutral or earthed object is bought close to the dome of the machine it is likely that a single electric spark will jump to the object and the machine will lose all its built-up charge very rapidly. This is called discharge. (See Photo 21.2.) This effect is similar to the type of process that causes lightning to occur in nature.

The Earth itself is a particularly interesting object electrically. The Earth can be considered as either a source of electric charge or as a sink for electric charge. The process of transferring electric charge to or from the Earth is called **earthing** or grounding and requires an electrical conductor, which is discussed in the next section. Charges are free to move in a conductor either to or from the Earth so that neutrality is maintained. It is common to refer to the process of discharging any electrified object so that it becomes electrically neutral as 'earthing' that particular object.

21.3 CONDUCTORS AND INSULATORS

Photo 21.2
Van de Graaff generator.



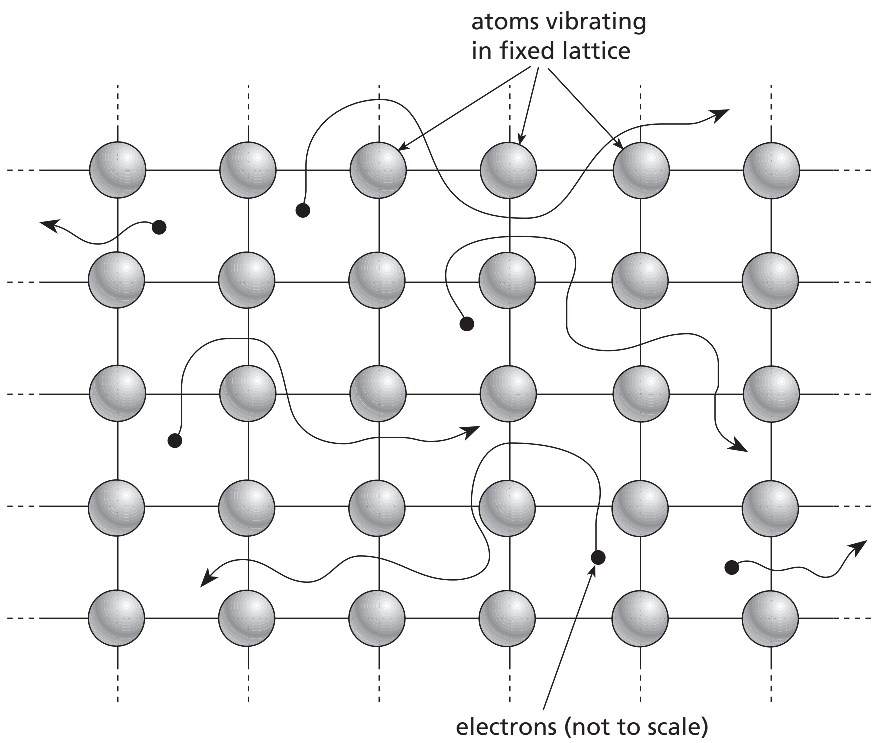


Figure 21.3
Electrons can move freely within the regular lattice arrays of good conductors.

Within solids the atoms are generally arranged in a simple geometric pattern or lattice, whereas in liquids and gases the atoms are much more free to move. A metallic solid has a very regular lattice array of atomic nuclei and the outermost electrons of each atom are quite easily able to move throughout this lattice array. This type of structure is called metallic bonding. (See Figure 21.3.) Metals are called good electrical **conductors** because any charge placed on them is free to move through the lattice. Other substances that do not allow free charge movement, or whose atomic electrons are not free to move, are called **insulators**. Charges that are placed onto any conductor will be free to evenly distribute themselves over the surface of the conductor, whereas charge placed onto the surface of any insulator will stay in the same place. (See Figure 21.4.) The human body acts as a conductor and this is why touching any charged object will lead to the removal of charge from the object.

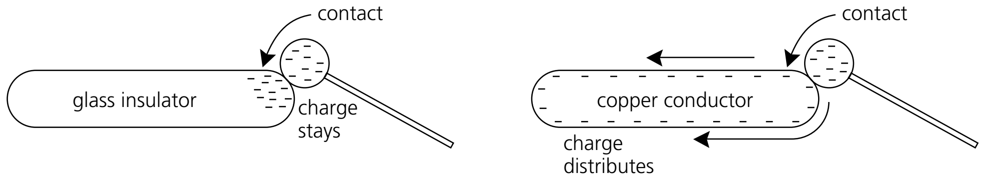


Figure 21.4
Charge movement within conductors and insulators.

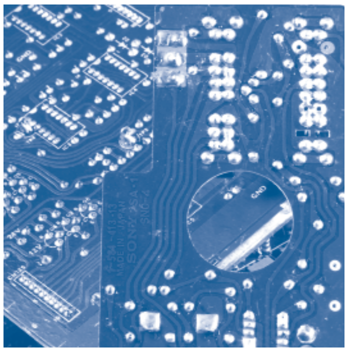
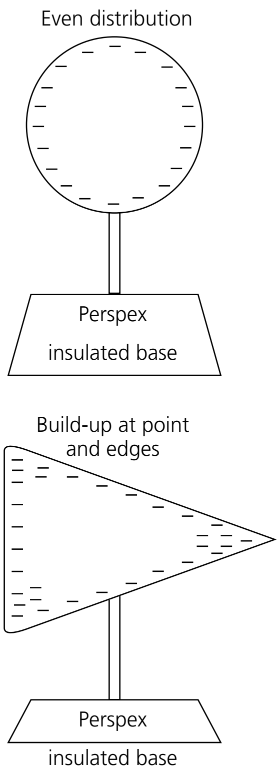


Photo 21.3
Circuit board (copper tracks).

Most metals are very good conductors of electricity, as are liquids that contain charged ions, such as salty water or acid solutions. The most common insulators used to prevent the flow of electric charge are rubber, plastics, paper, glass or ceramics. Insulators are seen in electrical wiring as either a direct coating over the copper conductor or as a stand-off insulating support on power poles in the street. Most household electronic devices have internal circuit boards that contain conducting copper tracks on an insulating fibreglass base. **Semiconductors** are a modern class of materials including silicon, germanium, gallium arsenide and various metal oxides that, in their natural state, are relatively poor conductors compared with metals. However, the conductivity of these crystalline materials can be artificially improved by the addition of selected impurity elements into their crystal structure. This class of semiconductors has become the basis of modern electronic chips and will be discussed more fully in Chapter 23. Their conductivity is variable and can even be switched on and off. Consider what happens when negative charge is placed onto a perfectly symmetrical conductor such as a metal sphere. Remember that the charges are free to move and because they all equally repel each other, they will reside equally distributed over the sphere's surface. If, however, the excess negative charges are placed onto a non-symmetrical (asymmetrical) conductor surface such as a metallic cone, then the charges will repel each other in such a way that they will tend to pile up at the edges and at the point of the cone. As a result, charge leakage to the surrounding air will occur more rapidly at edges and points on a conductor. An important application of this effect occurs with lightning rods or arresters on tall buildings, which allow charges built up during high winds and electrical storms to dissipate very quickly and prevent possible lightning strikes. (See Figure 21.5.)

Figure 21.5
Charge distribution in cross-section over regular and irregular conductors.



DETECTING AND DEFINING CHARGE 21.4

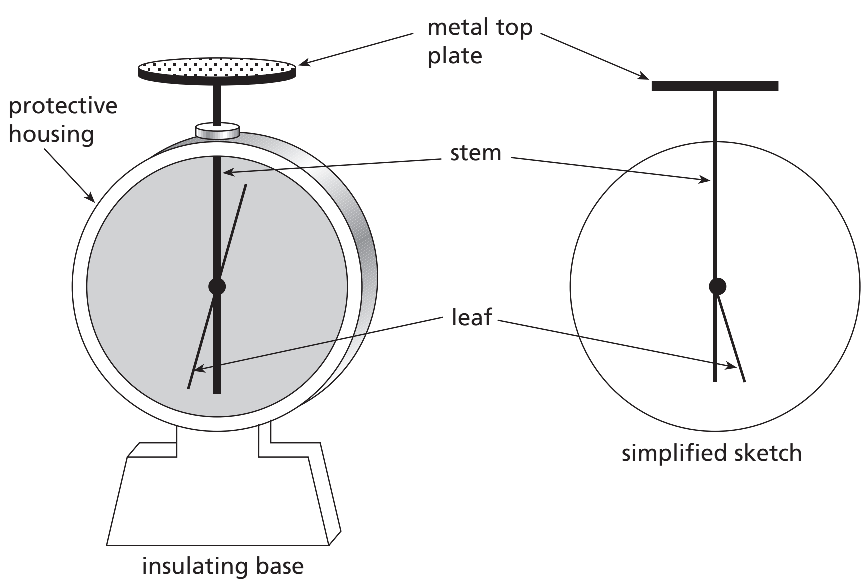


Figure 21.6
An electroscope detector.

Ch21-Walding 4th 25/8/04 12:22 PM Page 450

The property of mutual repulsion of like electric charge can be put to use in the design of a simple instrument that will detect the presence of charge. The **electroscope** consists of a metal top plate and stem cased within an insulating protective housing. Attached to the centre stem is either a very fine metal foil strip (leaf) that hangs vertically or a freely pivoted counterbalanced arm. (Figure 21.6.) The electroscope can either detect the presence of charge on any object brought close to the metal plate, or it can be charged itself by contact. If a positively charged rod is brought close to the metal plate, the leaf strip rises as it is repelled away from the stem, as free electrons in the stem and leaf strip are being attracted to the metal plate. If a negatively charged rod is brought close to the metal plate, the leaf strip rises as it is repelled away from the stem again, but this time due to free electrons being forced downward to the leaf strip and the stem. In either case, once the charged rod's influence is removed the leaf strip will collapse back to the stem because no net charge separation within the electroscope occurs (Figure 21.7**(a)**).

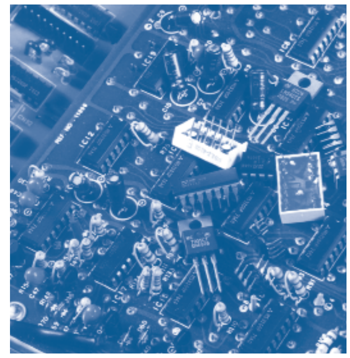


Photo 21.4
IC chips.

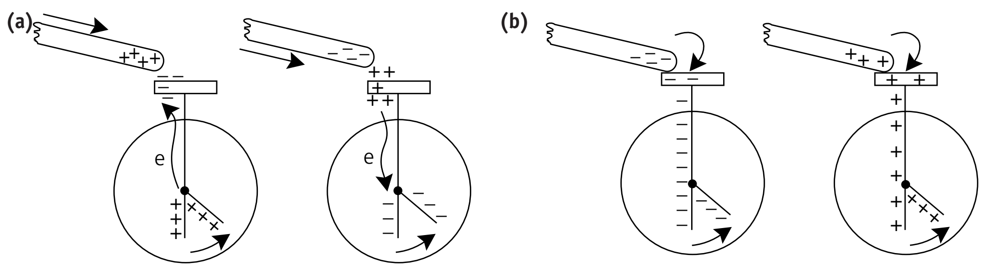


Figure 21.7
(a) Charging an electroscope.
(b) Contacting an electroscope; leaf repulsion will remain.

The electroscope can be charged by contact with the metal plate. If, for example, a positively charged rod touches the top plate and is rubbed across it, the charge is shared with the electroscope. The leaf will be repelled away from the stem, due to equal positive charge and will remain repelled even if the contacted charged rod is removed. Similarly if a negatively charged rod is touched to the plate and rubbed across it, negative charge is shared with the electroscope, and the leaf and stem mutually repel. Again the electroscope will remain charged even if the negative rod is removed (Figure 21.7**(b)**).

If the electroscope is charged with a known polarity, it can then be used to test for the presence of a charged rod. The leaf of an electroscope of known polarity will fall if a rod of opposite charge is brought into close proximity to the metal plate or will rise further if a rod of similar charge is brought into close proximity. Its leaf will also fall if an uncharged rod is brought near, so the only sure test for a charge is repulsion.

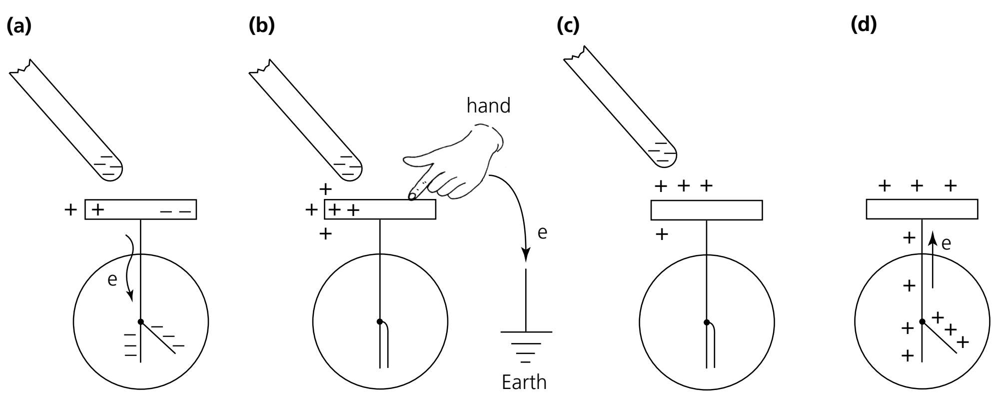
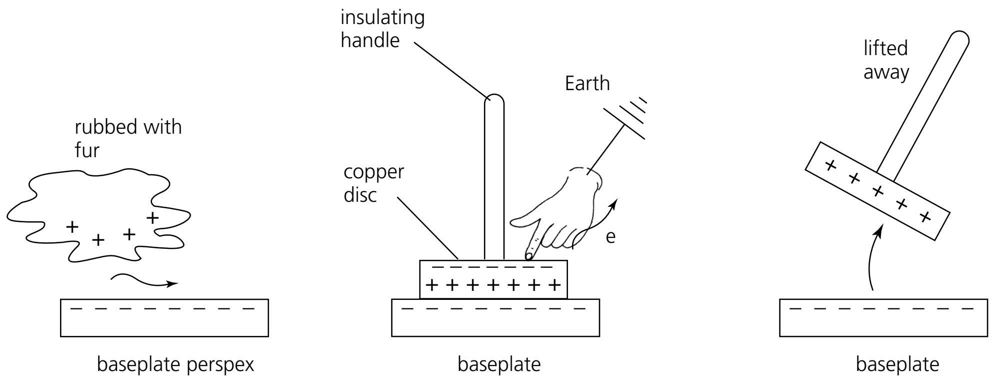


Figure 21.8
Charging an electroscope by induction.

The electroscope can also be charged by a process known as **electrostatic induction**. This involves a stage of earthing the metal plate during the charging sequence. The important fact to remember is that charging an electroscope by induction will always produce an opposite charge polarity between the electroscope and the charged rod used for induction charging. Figure 21.8 shows the steps necessary. Consider a negatively charged rod brought into close proximity to the metal plate. This repels electrons away from the plate to the leaf and stem and the leaf rises (Figure 21.8**(a)**). If the metal plate is now earthed, the leaf falls as excess electrons are forced to earth through the hand, leaving the electroscope positively charged (Figure 21.8**(b)** and **(c)**). As the hand and then the rod are removed the electroscope leaf and stem again repel due to redistribution of positive charge equally over the electroscope (Figure 21.8**(d)**).

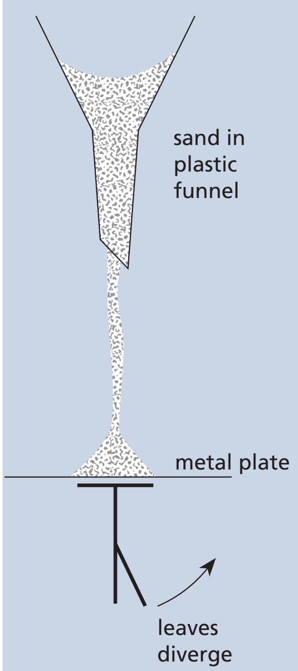
An interesting device that can be used to provide a source of electric charge is the **electrophorus**. This makes use of both friction charging and earthing techniques. The electrophorus consists of a large plastic or wax baseplate and a metallic disc on an insulating handle, which can sit on the baseplate. To charge the device, the baseplate is rubbed vigorously with woollen cloth or rabbit fur. This creates charge on the baseplate. The metallic disc is grasped by the insulating handle and placed flat down on top of the charged baseplate. Charge separation across the width of the metallic disc occurs due to attraction of the charges on the baseplate. If the top surface of the metallic disc is now earthed momentarily by touch and then the disc is lifted up from the baseplate, the disc will become electrically charged. This device is surprisingly effective in laboratory situations for demonstrating electrostatic effects. (See Figure 21.9.)

Figure 21.9
Charging a simple electrophorus.



NOVEL CHALLENGE

When sand falls through a plastic funnel onto a metal plate on an electroscope, the leaves diverge. Explain this, if you can.



Electric charge is measured in units called coulombs. The unit is named after Charles-Augustin **Coulomb**, a French physicist (1736–1806). One coulomb of electric charge is defined as the charge on 6.25×10^{18} electrons. If an electrified object has an excess of 6.25×10^{18} electrons or has lost this number of electrons, it will have a charge of one coulomb (1 C) negative or positive respectively. This means that the charge on a single electron particle can be calculated as $\frac{1}{6.25 \times 10^{18}} = 1.6 \times 10^{-19}$ C negative. The symbol Q is often used for the quantity of charge on any object.

In practice, with electrically charged objects in the laboratory, very large numbers of electrons are moved. Common units of electric charge become microcoulombs (μC). When charge moves through a metallic conductor at a given rate an **electric current** is produced. By definition, if one coulomb of electrons pass any given point in a conductor per second then a current of one ampere is flowing:

$$1 \text{ ampere (A)} = 6.25 \times 10^{18} \text{ electrons per second}$$

Example

A Perspex rod is rubbed with a piece of silk. It acquires a charge of 0.02 μC. **(a)** Is this charge positive or negative and how many electrons are transferred? **(b)** If all this charge is now transferred to an electroscope with charge +0.01 μC, what will be the net state of the electroscope?

Solution

(a) By using the triboelectric series, the charge on the Perspex rod is + 0.02 μC positive.
- 0.02×10^{-6} C = $0.02 \times 10^{-6} \times 6.25 \times 10^{18}$ electrons, or
- 1.25×10^{11} electrons transferred from Perspex to silk.

(b) Total charge on the electroscope after transfer = 0.02 μC + 0.01 μC = 0.03 μC (positive).
The electroscope leaf deflection will increase and the net charge on the electroscope is now +0.03 μC = 1.87×10^{11} electrons depleted.

Questions

1 Use the triboelectric series to describe the outcome of rubbing:
(a) glass with silk;
(b) rubber with woollen cloth;
(c) insulated gold with cat fur.

2 Draw a diagram showing the electrostatic outcome of two Perspex rods rubbed with silk cloth and freely suspended in close proximity to each other.

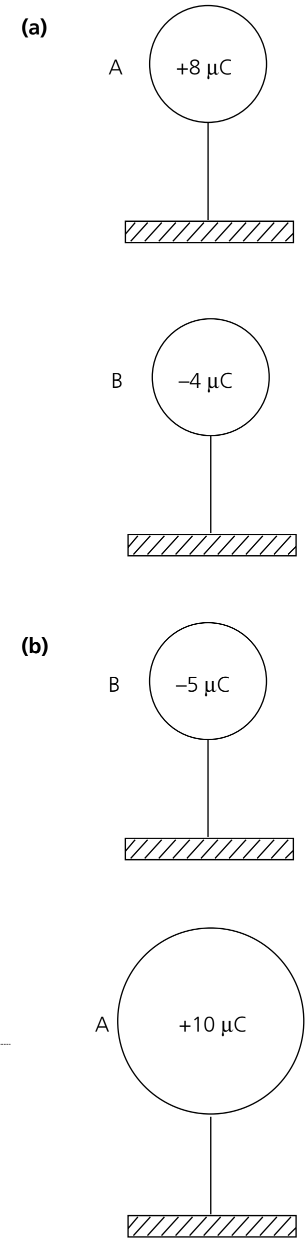
3 Draw a set of diagrams to explain the outcome of bringing the following within close proximity of, but not touching, a neutral electroscope:
(a) a negatively charged rod;
(b) a positively charged rod.

4 In a particular chemical reaction an iron atom (Fe) becomes a ferric ion (Fe^{3+}). Determine the charge on this ion in coulombs and describe what has occurred in terms of electron transfer.

5 Consider Figure 21.10**(a)**. A and B are identical insulated metal spheres with charges as shown. What is the charge on each sphere after they are touched together and then separated?

6 Consider Figure 21.10**(b)**. A and B are insulated metal spheres with charges as shown. If sphere A has twice the radius of sphere B, what will be the charge on each sphere after they are touched together and separated? (Remember, charge resides equally distributed over the surface of a spherical metallic conductor.)

Figure 21.10
For questions 5 and 6.



21.5 COULOMB'S LAW

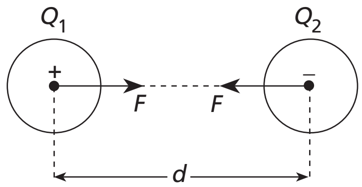
It has been seen that when electrically charged objects are brought into close proximity, there is a force between them that is either attractive or repulsive, depending on the nature of the charges. Charles-Augustin Coulomb in the late eighteenth century used a very sensitive electrostatic torsion bar balance system to investigate the nature of this force. From careful measurements made on the quantity of charge, the distances between charges and the forces acting on the charges, Coulomb was able to show that:

- the magnitude of the force was proportional to the product of the charges
- the magnitude of the force was inversely proportional to the square of the distance separating the charges
- the direction of the force was along a line joining the centres of the charges
- the magnitude of the force was dependent on the medium in which the charges were placed.

These points can be summarised mathematically as:

$$\boldsymbol{F} \propto \frac{Q_1 Q_2}{d^2} \quad \text{or} \quad \boldsymbol{F} = k\,\frac{Q_1 Q_2}{d^2} \quad \text{for charges in air}$$

Figure 21.11
Coulomb's law of attraction.



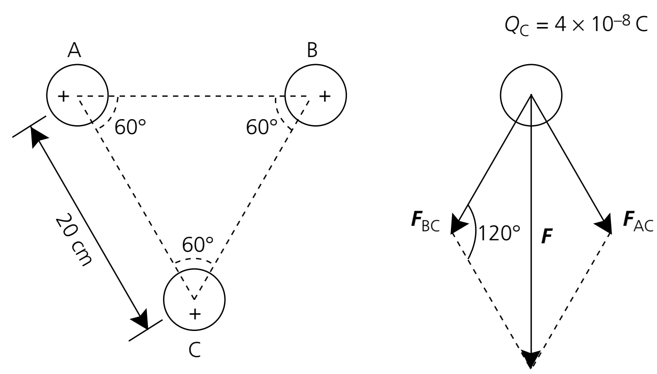
where Q_1 and Q_2 are the charges in coulombs (C); d is the separation of the charges in metres (m); and $k = \dfrac{1}{4\pi\varepsilon_0} = \dfrac{1}{4\pi \times 8.85 \times 10^{-12}} = 9.00 \times 10^9$ N m^2 C^{-2}. The constant epsilon-zero is called the 'permittivity of free space'.

This equation is referred to as **Coulomb's law** and it must always be remembered that the force involved is a vector quantity. For two charges in isolation the direction of the force will be repulsive if the charges are the same sign and attractive if the charges are opposite in sign (Figure 21.11).

The vector nature of this law is also very important if a system of two or more charges is considered. In this case it is necessary to determine the resultant electrostatic force in both magnitude and direction using vector addition techniques. In problems dealing with Coulomb's law it is often convenient to consider the charges as point charges. This is possible whenever the separation distance of the charges is very large compared with the size of the charges themselves.

Example

Figure 21.12
System of charges.



Consider the system of charges as illustrated in Figure 21.12. Deduce the magnitude and direction of the force on charge C. Each charge is $+4.0 \times 10^{-8}$ C and they are arranged in an equilateral triangle of side 20 cm.

Solution

The force on charge C will be the vector resultant of both forces labelled $\boldsymbol{F}_{AC}$ and $\boldsymbol{F}_{BC}$ as shown, which are equal to each other in magnitude.
Calculate the magnitude of each force vector:

$$\frac{9 \times 10^9 \times 4 \times 10^{-8} \times 4 \times 10^{-8}}{(0.2)^2} = 3.6 \times 10^{-4}\ \text{N}$$

Force ($\boldsymbol{F}$) may now be calculated using the cosine rule for triangles:

$$\begin{aligned}
\boldsymbol{F}^2 &= \boldsymbol{F}_{AC}^{\,2} + \boldsymbol{F}_{BC}^{\,2} - (2 \times \boldsymbol{F}_{AC} \times \boldsymbol{F}_{BC} \times \cos 120^\circ) \\
\boldsymbol{F}^2 &= (3.6 \times 10^{-4})^2 + (3.6 \times 10^{-4})^2 - (2 \times 3.6 \times 10^{-4} \times 3.6 \times 10^{-4} \times -0.5) \\
\boldsymbol{F}^2 &= 3.89 \times 10^{-7}\ \text{N} \\
\boldsymbol{F} &= 6.2 \times 10^{-4}\ \text{N}
\end{aligned}$$

The force is repulsive and will be directed vertically down the page or at an angle of 30 degrees to the line of the force $\boldsymbol{F}_{AC}$.
Note: this problem may also be solved by dividing the diagram up into smaller right-angled triangles and using simpler Pythagorean calculations.

Questions

7 Within an atomic helium nucleus, two protons are separated by a distance of 1×10^{-14} m. What is the size of the Coulomb repulsion force between them? Recall that the charge on a proton is the same as the charge on the electron.

8 Two point charges A and B each have charge +Q coulombs and are separated by a distance r metres. If the force acting between them is 6×10^{-4} N: **(a)** is the force attractive or repulsive; **(b)** what is the force if distance r is doubled; **(c)** what is the force if charge Q on both A and B is doubled; **(d)** what is the force if charge Q on both is halved, but so is the distance between them?

9 Four point charges A, B, C and D are arranged on corners of a square of side 25 cm. If A and B each have a charge of +1 μC while C and D each have a charge of +2 μC, what is the resultant force on a charge of +1 μC placed at the centre of the square?

NOVEL CHALLENGE

This one is only if you do chemistry as well. In his book, *The Ascent of Science*, Professor Brian Silver says that if you had two sugar cubes and transferred one electron of each billion electrons in one of the cubes to the other cube, the force between the two cubes would equal the weight of boxer Mohammed Ali (1000 N) when the cubes were placed 1 km apart. Sucrose (sugar) has the formula $C_{12}H_{22}O_{11}$ and a molar mass of 342. *Assuming a cube of it has a mass of 5 g, what figure do you get? We get nothing like 1000 N.*

21.6 ELECTRIC FIELDS

Like other fields existing in physics, such as a gravitational field, we can define a region containing an **electric field** as that in which any electrified object will experience a force. This makes it a similar concept to a mass experiencing a gravitational force of attraction to, say, the Earth. Electric fields exist in space around electrically charged objects and physicists define both the electric field magnitude or strength and its direction. Electric field strength, ***E***, is thus a vector quantity. The main difference between electric fields and gravitational fields is that the latter only has attractive forces, whereas electrical fields can provide both attractive and repulsive forces.

The magnitude of the electric field, or its electric field strength at any point, is defined as the force acting on a test unit charge placed at that point in the field. The direction of the electric field at any point is given by the resultant force direction acting on a test positive unit charge placed at that point in the field. Thus mathematically, electric field strength is defined with the following equation:

$$\boldsymbol{E} = \frac{\boldsymbol{F}}{q}$$

Electric field strength has standard units of N C^{-1} (newtons per coulomb).

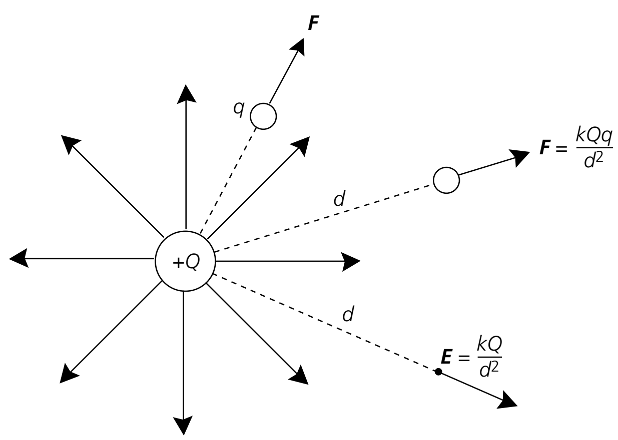


Figure 21.13
Electric field diagram surrounding a point charge +Q.

Consider a point source of positive charge, +Q, as shown in Figure 21.13. If a small test positive charge, q, was moved around the charge Q, then forces of repulsion would be felt in

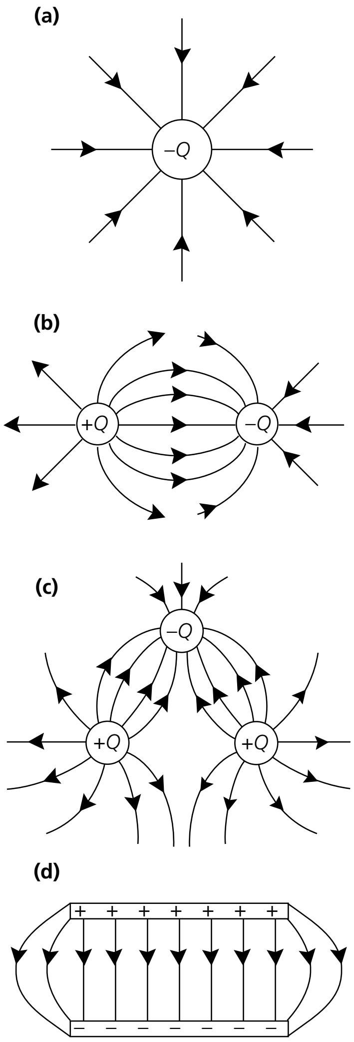


Figure 21.14
Electric field diagrams in cross-section.

radial lines directed outward. The further away q was moved along these lines the weaker the electric field strength would get. At all times the force magnitude would be given by Coulomb's law and thus the electric field strength, $\boldsymbol{E}$, at some distance d from the charge Q would be:

$$\boldsymbol{E} = \frac{\boldsymbol{F}}{q} = \frac{kQq}{d^2} \times \frac{1}{q} = \frac{kQ}{d^2}$$

and its direction is radially outward.

This equation therefore describes the electric field strength at a point d from a large point charge Q. The electric field strength $\boldsymbol{E}$ is defined as having a value of zero at an infinite distance from the source. Note that Figure 21.13 is representing the nature of the electric field in a two-dimensional cross-section only and the actual zone of influence is always in three dimensions around point charges. Figure 21.14 gives cross-sectional electric field diagrams showing the two-dimensional field lines present in several situations where there may be more than one charge influence involved. It is important to realise that:

- electric field lines never cross, in a diagram
- electric field lines are directed from positive charge to negative charge
- electric field lines will enter or leave any charged surface at right angles
- if electric field lines are close together or more densely packed per unit area, the force per unit charge is much higher, that is, there is greater electric field strength.

Notice that in Figure 21.14**(d)** the electric field lines between parallel charged plates is very uniform in nature. This uniform electric field will produce a constant force on any test charge held between plates independent of position. In 1909 an American physicist **Robert Millikan** used the very uniform electric field between charged parallel metallic plates to investigate the nature of electric charge itself. Millikan was able to balance the electric force on charged oil droplets sprayed between the plates with the droplets' own gravitational weight. By carefully measuring the mass of these oil droplets and changing their electric charge with X-rays, Millikan was able to calculate a value for the charge on an electron. He won the Nobel prize in physics in 1923 for his work on the electrostatics of elementary charges.

Research work in more recent years has concluded that particles called **quarks** have less electric charge than the elementary electron. Several different types of quarks are thought to exist in theory, making up protons and neutrons. Quarks are postulated to have charges of $+\frac{2}{3}$ e and $-\frac{1}{3}$ e but particle physicists have not yet been able to detect a free quark existing by itself in their research or experiments. (Refer to Chapter 29.)

Example

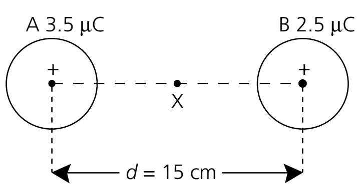


Figure 21.15
Diagram of point charges A and B.

Figure 21.15 shows two point charges A and B separated by a distance of 15 cm.

(a) Determine the magnitude and direction of the electric field at a point X midway between the two charges.

(b) Determine the point between the charges at which the electric field strength is zero.

Solution

(a) The electric field at X is the vector sum of the fields due to each point charge.

$$\boldsymbol{E}_{\mathrm{A}} = k\frac{Q}{d^2} = \frac{9.00 \times 10^9 \times 3.5 \times 10^{-6}}{(7.5 \times 10^{-2})^2} = 5.6 \times 10^6 \text{ N C}^{-1} \text{ toward B}$$

$$\boldsymbol{E}_{\mathrm{B}} = k\frac{Q}{d^2} = \frac{9.00 \times 10^9 \times 2.5 \times 10^{-6}}{(7.5 \times 10^{-2})^2} = 4.0 \times 10^6 \text{ N C}^{-1} \text{ toward A}$$

Hence the resultant: $\boldsymbol{E} = 5.6 \times 10^6 - 4.0 \times 10^6 = 1.6 \times 10^6$ N C^{-1} toward B

(b) The point where the electric field is zero is the point where the electric field $\boldsymbol{E}_A$ is numerically equal to the electric field $\boldsymbol{E}_B$, but opposite in direction. Let this take place at a distance s from A. Thus:

$$|\boldsymbol{E}_A| = |\boldsymbol{E}_B|$$

$$\frac{9.00 \times 10^9 \times 3.5 \times 10^{-6}}{s^2} = \frac{9.00 \times 10^9 \times 2.5 \times 10^{-6}}{(15 - s)^2}$$

$$3.5 \times 10^{-6}\ (15 - s)^2 = 2.5 \times 10^{-6} \times s^2$$

$$0.3s^2 - 30s + 225 = 0$$

$$s = 8.1 \text{ cm from A}$$

NOVEL CHALLENGE

Here's a good idea we overheard at a UFO conference. To launch a rocket, it was suggested that 1 gram of hydrogen ions be placed aboard a rocket and then another gram of H^+ ions be wheeled underneath in a wheelbarrow (say 1 metre away). One gram of hydrogen ions (H^+) contains 6×10^{23} particles. So the big positive charge in the rocket would be repelled by the big positive charge in the wheelbarrow (about a billion billion newtons) and this would be sufficient force to launch the rocket. *Calculate the force and the initial acceleration of a 100 tonne rocket. What is the major flaw in this design?*

21.7 ELECTRIC POTENTIALS

The gravitational field analogy can be used again when discussing the energy stored in an electric field. Recall that in gravitation, work was needed to lift a mass against the direction of the gravitational field and as this work was done, gravitational potential energy was stored in the mass. The energy stored was equivalent to the work done, W, and calculated using $W = m\boldsymbol{g}h$. Similarly, it requires work to be done to move electric charges of the same polarity together, as the force of mutual repulsion is overcome. Consider a point positive charge q moved from infinity towards another point positive charge Q. The force of repulsion is the Coulomb force and the work done in moving the charge q to a point in space at a distance d from the charge Q is given by:

$$W = q \times \boldsymbol{E} \times d$$

Compare this formula with the work done on a mass m in a gravity field $\boldsymbol{g}$ shown above.

This relationship for electrical work done is true even though the nature of the Coulomb force of repulsion in the field is not uniform. By using very small incremental distances and mathematical calculus techniques, it is possible to verify that the total work done is the same as the sum of all incremental force by distance calculations.

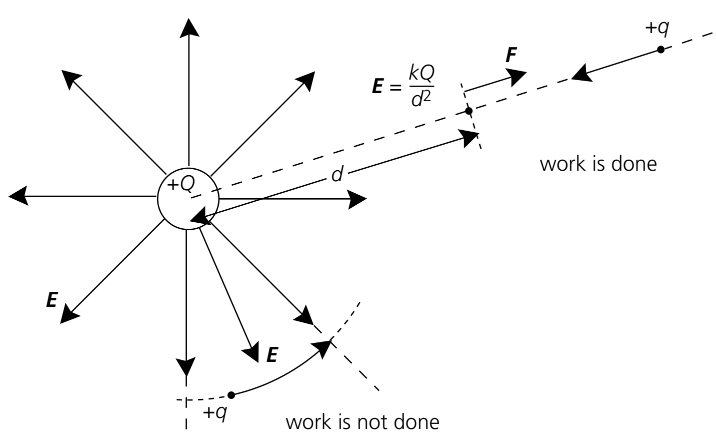


Figure 21.16
Work done within an electric field.

Also remember that because W = $\boldsymbol{F}\, d \cos \theta$, if the charge is moved perpendicular to the field lines then no work at all will be done. The work done in moving charge q will subsequently store electrostatic potential energy in the charge q. If the repulsive force between the charges is allowed to act by itself then the stored **electrostatic potential energy** will be converted back into kinetic energy of motion. This effect can be made use of in electrostatic charge accelerators, especially when the uniform field of parallel charged plates is present (Figure 21.16).

It is common to refer to the **electric potential** (V) of a point in space within any given electric field. The electric potential V at any point is the work done in moving a unit positive charge from infinity to that point. The electric potential V is thus the electrostatic potential energy stored per unit charge at any given point d from a charge Q. Thus:

$$V = \frac{W}{q} = \frac{q \times \boldsymbol{E} \times d}{q} = \boldsymbol{E} \times d = \frac{kQ}{d}$$

Electric potential at a point is a scalar quantity and is measured in units of joules per coulomb (J C^{-1}). One joule per coulomb is also known as a volt. 1 volt (V) = 1 J C^{-1}. Of course, moving unit positive charges from infinity to a point d within the field of a point charge Q is not particularly realistic, and hence the term **potential difference** is more useful in that it describes the difference in potential between two positions within the field, neither of which is at infinity (Figure 21.17). The potential difference is the work done in moving a unit charge from position one in the field to position two in the field.

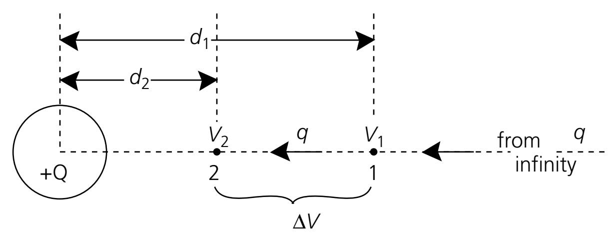


Figure 21.17
Electric potential difference between two points in a field.

$$\Delta V = V_2 - V_1 = \frac{W_{12}}{q} \Rightarrow W_{12} = q \times \Delta V = q \cdot (V_2 - V_1)$$

Potential differences become very useful in electric circuit work, which will be discussed in the next chapter. If both ends of an electrical conductor are held at different potentials (by connecting to the opposite poles of a battery, say!) then a potential difference is set up across the conductor and the subsequent electric field within the conductor will allow free electrons to flow. This is what is referred to as an electric current within the conductor.

Example

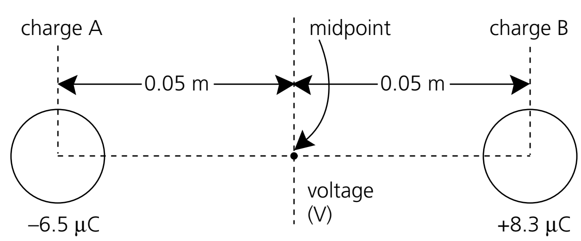


Figure 21.18
Potential difference between two charges.

Consider Figure 21.18, with charges A and B of −6.5 μC and +8.3 μC respectively, in space. If these charges are separated by a distance of 10 cm, find the potential at the midpoint between their centres.

Solution

The potential at the midpoint will be the scalar addition of the potentials due to each charge. Potential at midpoint due to charge A is:

$$V_A = \frac{kQ}{d} = \frac{9.00 \times 10^9 \times -6.5 \times 10^{-6}}{0.05} = -1.2 \times 10^6 \text{ V}$$

Potential at midpoint due to charge B is:

$$V_B = \frac{kQ}{d} = \frac{9.00 \times 10^9 \times 8.3 \times 10^{-6}}{0.05} = +1.5 \times 10^6 \text{ V}$$

$$\text{Hence } V_{Total} = V_A + V_B = 0.3 \times 10^6 \text{ V} = 3.0 \times 10^5 \text{ V}$$

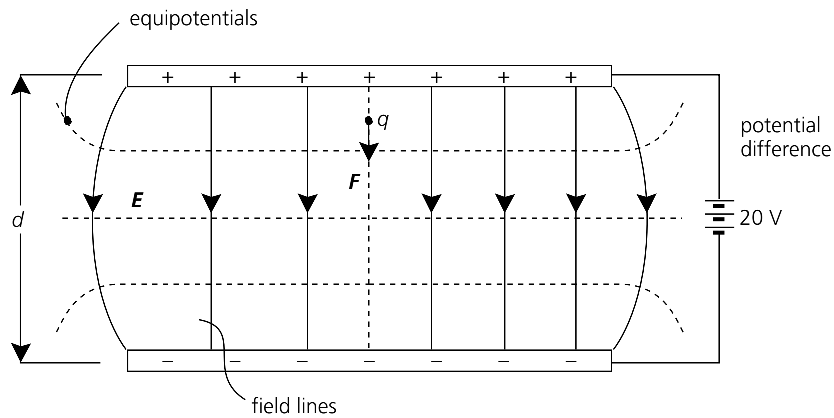


Figure 21.19
Uniform electric field between parallel conductive plates.

Recall in the previous section that the electric field between a set of parallel conductive plates (Millikan plates or a capacitor) is very uniform in nature. Consider now the electric potentials involved in this situation (Figure 21.19). If the plates are separated by a distance *d* and the potential difference between the plates is maintained by a battery of voltage *V*, this voltage is the work done in moving a test charge *q* from one plate to another.

$$V = \frac{W}{q} \quad \text{or} \quad W = q \times V$$

The test charge *q* would also experience a force in the field and work done against this force to move the charge across the plate separation would be given by:

$$W = \mathbf{F} \times d = q \times \mathbf{E} \times d$$

Thus within this uniform electric field:

$$q \times V = q \times \mathbf{E} \times d \quad \text{or} \quad V = \mathbf{E} \times d \Rightarrow \mathbf{E} = \frac{V}{d}$$

which defines the electric field strength in the alternative unit of volts per metre (V m^{-1}). Volts per metre are equivalent to newtons per coulomb (N C^{-1}). If lines are drawn in the region between the parallel plates that join points of equal potential, they would be parallel to the plates and spaced equal distances apart. These lines are called **equipotentials**.

The energy unit used in atomic physics is called the **electron-volt** (eV) and is defined using the fields of parallel plates. If an electron moves across a potential difference of 1 volt, then work done is:

$$W = q \times V = 1.6 \times 10^{-19}\ \text{C} \times 1.0\ \text{V} = 1.6 \times 10^{-19}\ \text{J}$$

Thus 1 eV = 1.6×10^{-19} J.

It can be seen that this energy unit is considerably smaller than the joule energy unit and as such is useful in atomic structure research. Particles leaving radioactive atoms can possess millions of electron-volts and accelerator machines used in high energy physics laboratories can develop particle energies of giga or tera electron-volts (GeV, TeV).

Example

In a Millikan-type experiment a suspended negatively charged latex sphere has a mass of 5.7×10^{-7} g and is held at rest between the plates with potential difference V of 280 volts. If the plates are separated by a distance of 4.0 mm:

(a) draw a diagram of the apparatus in cross-section and label the plate polarity correctly;
(b) calculate the electric field strength;
(c) calculate the charge on the latex sphere in both coulombs and elementary charges.

Solution

(a) Plates must be oriented with positive plate uppermost so that the electric field force balances the downward gravitational force.

(b) Using $\boldsymbol{E} = \dfrac{V}{d} = \dfrac{280}{4.0 \times 10^{-3}} = 7.0 \times 10^{4}\ \text{V m}^{-1}$ down.

(c) Since the gravitational force balances the electrical force:

$$q \times \boldsymbol{E} = m \times \boldsymbol{g}$$

$$q = \frac{m\boldsymbol{g}}{\boldsymbol{E}} = \frac{5.7 \times 10^{-10} \times 9.8}{7.0 \times 10^{4}} = 8.0 \times 10^{-14}\ \text{C}$$

$$\frac{8.0 \times 10^{-14}}{1.6 \times 10^{-19}} = 5.0 \times 10^{5}\ \text{electrons}$$

Questions

10 Draw the cross-sectional electric field diagram for a system of negative charges, all situated at the corners of an equilateral triangle.

11 What is the electric field strength 0.2 m from a point charge of −6 μC in both magnitude and direction?

12 A metal sphere of radius 35 cm carries a charge over its surface of 16 μC. What is the potential at its surface?

13 Two points in space are at electric potentials of +18 V and −6 V respectively. Calculate **(a)** the potential difference between these two points and **(b)** the work done in moving a charge of 5.5 μC from one to the other.

14 Two metal plates are placed vertically 30 mm apart and a potential difference of 300 volts is applied. (The top plate is positive.)

(a) Calculate the electric field between the plates.
(b) If a negative charge of 6 μC is placed in the field at a point 10 mm above the earth plate, what force acts on it?
(c) Calculate the energy gained by the charge as it is moved up to the positive plate.

21.8 APPLICATIONS OF ELECTROSTATICS

SR Activity 21.1 SPECIAL ELECTROSTATIC EFFECTS

Each of the situations described below involves an application of electrostatic cause and effect. Read each short paragraph and try to explain on paper, with diagrams, the physics involved in each application. Answer the questions following each section.

1 *Conductive tyres or discharge straps on vehicles* Aircraft and large fuel tanker trucks often have tyres made of conductive rubber so that electric charge built up during travel through the air will discharge quickly to the ground. If this precaution is not available, special conductive straps are attached between the ground and the plane or truck body before any fuel exchange takes place. Some people attach conductive straps to cars, which touch the ground as the car drives along. These straps are to earth the car so that electrostatic charge built up during travel will not cause annoying electric discharges to the passengers as they get out of the vehicle. Would it be better to connect the strap to the ground or to the plane first?

2 *Lightning arrestors on tall buildings* During thunderstorms clouds build up very large potentials due to charge separation within the cloud. This effect induces a large build-up of charge onto buildings and objects rising up from the ground, which could act as discharge points or points for lightning to strike. Lightning arrestors are a series of upwardly pointing metal rods or spikes, often ornately shaped, which are connected to a copper earthing strap that runs down the building to the earth. This system allows rapid discharge to the air of charge built up at the top of the building and thus helps to prevent lightning strikes (Figure 21.20).

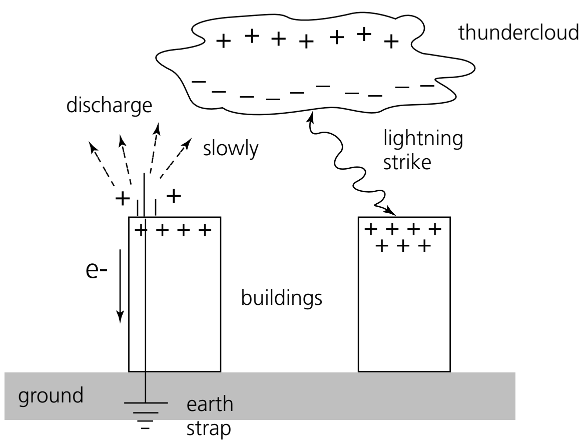
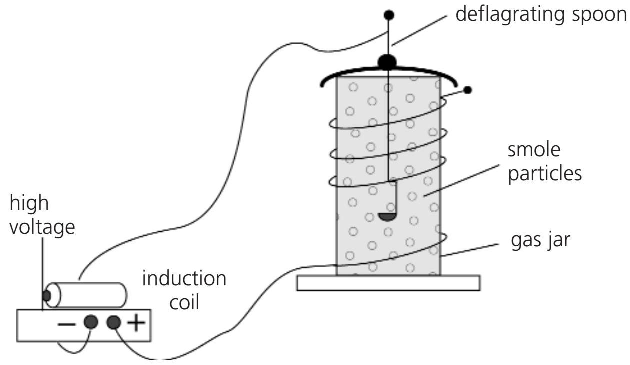


Figure 21.20 Applications of electrostatic principles: lightning arrestors.

Some people think that the arrestors carry lightning strikes to ground after the building has been hit by lightning. Explain whether or not this is true.

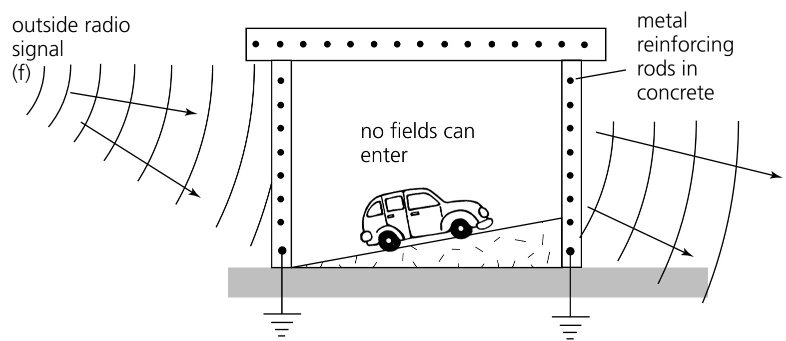
3 *Factory chimney precipitators* Smokestacks in modern industrial factories are fitted with a system of electrostatic precipitators. These consist of charged plates or helical coils around the top of the chimney stacks, which are held at a high positive electric potential. Smoke and ash debris formed in furnaces is often negatively charged, so that instead of being emitted to the air it is attracted to the plates and coils and precipitates out. This allows for collection of the solid material for periodic cleaning. You can model this effect quite easily with a gas jar full of smoke and several turns of copper wire wound around the body of the gas jar. If an induction coil is then connected across this coil and the central deflagrating metal spoon placed into the gas jar and turned on, the smoke very rapidly dissipates (Figure 21.21).
Explain how the smoke particles in the chimney might become charged in the real situation. How are the smoke particles charged in the lab model?

Figure 21.21
Model electrostatic precipitator.



4 *Faraday cages preventing radio reception* Have you ever had the experience of listening to your car radio and on entering a large city multi-storey car-park the radio goes weak or disappears altogether? This effect can often occur as you drive across a large enclosed bridge structure such as the Story Bridge in Brisbane or the Sydney Harbour Bridge. These reinforced concrete car-parks or metal bridges are earthed and no electromagnetic fields can enter them; recall that an electric field cannot exist within a hollow metal sphere. Faraday cages, named after Michael Faraday, do not have to be solid structures. An effective electrostatic field protective cage can be made out of metal wire mesh. Often these types of cages are used in the walls of laboratory rooms to help to prevent electrostatic or electromagnetic fields from interfering with sensitive electrical or computing instrumentation. (See Figure 21.22.)

Figure 21.22
Faraday cage effect.



Certain sections of radio and television receiver circuitry are often located inside a metal can housing on the local circuit board. Explain what section of the electronic circuitry is covered up, and why this might be the case.

Interesting applications of electrostatic fields are found in certain species of **electric fish** and rays. The South American eel-shaped fish (*Electrophorus electricus*) and the electric catfish (*Malapterurus electricus*) can deliver shocks from 450 to over 700 volts. Large torpedo rays (numbfish) can produce up to about 200 volts. Gymnotid eels and knife fishes generally produce much lower voltages in pulses from about 30 to 1700 hertz, which set up electric fields around their bodies. All these fish generate electricity in modified muscle tissue bundles called electroplates, located on both sides of their bodies. Stimulated by the autonomic nervous system these muscles generate large electric pulses instead of contracting as normal muscles do. The more powerful electric discharges can certainly stun or even kill prey directly; however, in the main, the specialised sensory nerves within the animal are sensitive to disturbances of the electric field surrounding it. Thus the fish uses the electric field it generates as a type of radar system to locate prey at night or in muddy waters.

A common machine in offices, schools and businesses today is the **electrostatic photocopier**. A charged plate within the machine receives a light reflected copy of the original page. The reflected dark image from the original causes the plate to remain charged in those areas that are copies of the original typing or diagrams. Carbon black is attracted to these charged areas of the plate and transferred to paper as it is pressed against it. The final step is to heat the carboned paper so that the powdered carbon is fused to the paper and made permanent. You will have noticed that if the photocopier is not working properly, the carbon black on the photocopy smudges easily or wipes off.

An **electron gun** is an assembly that forms the centre of instruments such as a cathode ray oscilloscope (CRO), a television picture tube (CRT) or computer monitor, as well as an accelerating device for particle accelerators (Figure 21.23). Electrons are boiled off the heated filament wire and enter the evacuated space between a pair of charged parallel plates. The voltage across these plates provides a field and an electrostatic force that accelerates the electrons from rest across the gap. The electrons are moving at very high speeds and continue on through an opening in the positive grid plate. Once the electron beam leaves the gun assembly it can be further controlled by electric or magnetic fields, such as occurs in a television tube, to provide the observable image on the screen. Formulae studied so far can be used to predict the kinetic energy and thus the velocity of the electron as it leaves the gun assembly. If the accelerating voltage is V and the charge on the electron is q_e then the work done by the accelerating force is converted into kinetic energy of the moving electron, thus:

$$q_e \times V = \frac{mv^2}{2} \Rightarrow v = \sqrt{\frac{2 \times q_e \times V}{m}}$$

This is referred to as the electron gun velocity formula, but it could be applied to any charge q of mass m accelerated through a potential difference V.

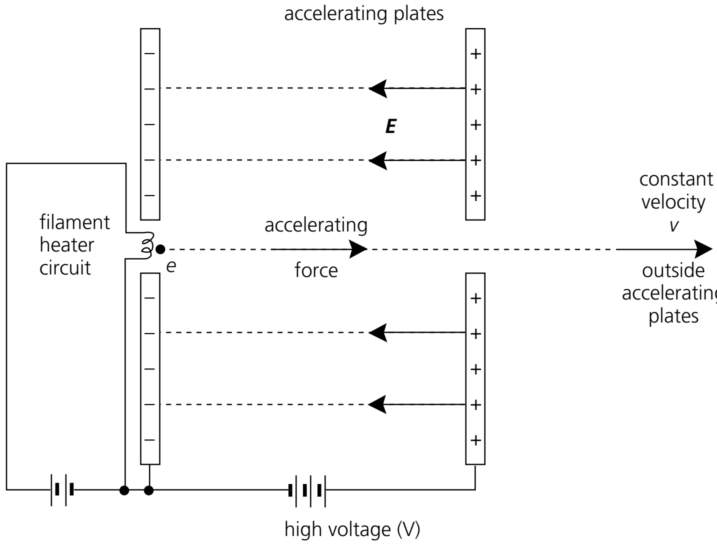


Figure 21.23
An electron gun assembly.

Example

If the accelerating voltage in an electron gun assembly is 350 volts, determine the kinetic energy of the electrons as they leave the gun and the velocity of projection into the evacuated space outside the gun plates.

Solution

Kinetic energy of the electrons equals the work done as they cross the potential difference:

$$W = q_e \times V = 1.6 \times 10^{-19} \times 350 = 5.6 \times 10^{-17}\ \text{J}$$

This work done equals the exit kinetic energy, so if the mass of the electrons is 9.11×10^{-31} kg, then using the formula derived:

$$5.6 \times 10^{-17} = \frac{mv^2}{2} = 9.11 \times 10^{-31} \times \frac{v^2}{2}$$

$$v = \sqrt{\frac{2 \times 5.6 \times 10^{-17}}{9.11 \times 10^{-31}}} = 1.1 \times 10^7 \text{ m s}^{-1}$$

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***15** Explain why a balloon rubbed with a piece of woollen cloth might stick to a wall notice-board without falling.

***16** State the charge on both materials if a piece of Perspex plate is rubbed with a silk cloth.

***17** Explain the movement of charge when a negatively charged Van de Graaff generator is momentarily earthed by touching it.

***18** Two charges repel with a force of 2.8×10^{-1} N. If one charge is $+6.5 \times 10^{-6}$ C and they are separated by 0.8 m in air, what is the value of the second charge?

***19** If a charge of 15 μC experiences a force of 750 N when placed in an electric field, deduce the strength of the field in correct units.

****20** An oil droplet experiences an electrostatic force of 5.6×10^{-14} N when placed into a uniform electric field of 4000 volts per metre. What is the magnitude of the oil droplet's charge?

****21** Two small metal-coated styrofoam spheres each of mass 2.80×10^{-6} kg are attached to nylon threads 45.0 cm long and hung from a common point. The spheres are then charged equally negative and the angle each supporting thread makes with the vertical is 16°. Calculate the charge on each sphere.

****22** A metal sphere carries a charge of 20 μC. If it has a diameter of 20 cm, what will be the potential, *V*, at its surface?

***23** Find the electric potential at a radial distance of 1.6 m from a point source of charge of value +8.2 μC.

****24** Two metal plates are oppositely charged and are separated by a distance of 4.5 mm in a vacuum. If a voltage of 520 V is connected across the plates, with the top plate positive, what is **(a)** the strength of the electric field between the plates; **(b)** the force on a test charge having an excess of 50 electrons on it placed between the plates?

****25** An electron in the gun of a CR tube is accelerated by a potential of 3.0×10^3 V. What is the kinetic energy of the electron in eV and J? What is the exit speed of the electron from the CR tube gun?

**26 A positively charged ion particle of mass 9.60×10^{-26} kg enters a uniform electric field of strength 20 N C^{-1} at right angles with an initial speed of 1200 m s^{-1}, as shown in Figure 21.24. If the ionic charge is 8.0×10^{-19} C:

(a) calculate the magnitude and direction of the force on the charged ion;
(b) calculate the acceleration of the ion particle in the field;
(c) calculate the time of travel of the ion particle in the field;
(d) calculate the final displacement and velocity in the direction of the field;
(e) describe the probable path of the ion particle in the field.

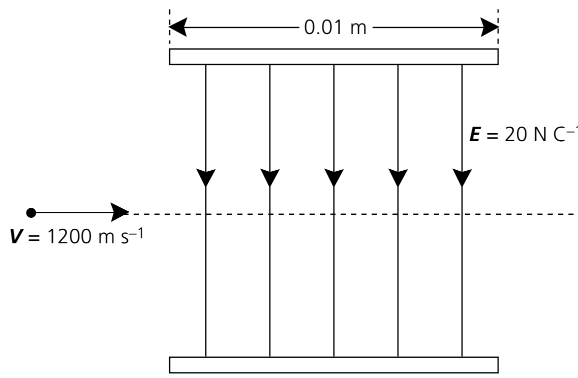


Figure 21.24
For question 26.

**27 The shaded area in Figure 21.25 represents an isolated charged metal object in cross-section. The surrounding lines are equipotentials. Use the diagram to analyse:

(a) near which point the electric field is strongest;
(b) the potential difference between A and B;
(c) the potential difference between A and C;
(d) the energy lost in moving a charge of $+4.5 \times 10^{-9}$ C from C to A;
(e) the energy gained in moving the same charge from A to D.

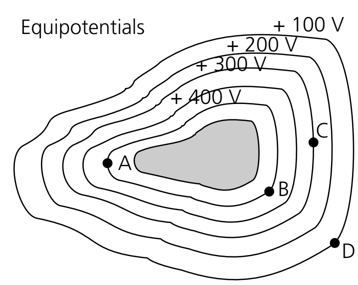


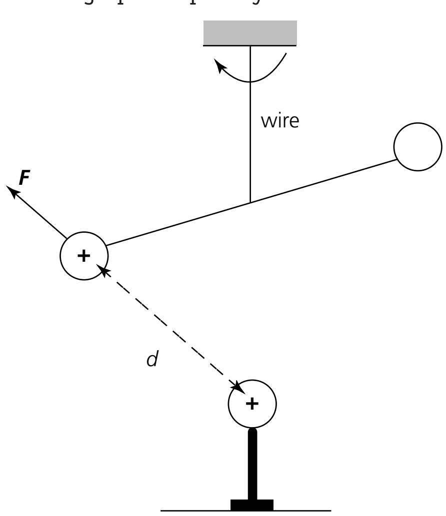
Figure 21.25
For question 27.

*28 What is meant by the phrase 'The Earth has zero potential'? Research this idea and report the outcome diagrammatically.

*29 Draw a sequence of diagrams to illustrate a technique for charging an electroscope negatively using an induction technique.

**30 In an experiment to replicate Coulomb's original experiment with electrostatic charges, a student set up a torsional balance as shown in Figure 21.26. The apparatus enabled the student to measure the twist in the suspension wire as the 'free' charge tried to rotate away from the fixed charge. From the value of the twist, the student derived a value for the electrostatic force of repulsion between the two spheres. Part of the student's data is shown in the data table. Unfortunately, two of the data points have been obscured by a chocolate smudge.

(a) Using graph paper, plot a graph to show the relationship between the electrostatic force and the separation distance.
(b) From your data, estimate the two readings obscured by the chocolate smudge.
(c) What relationship between the force and the distance is suggested by your graph? Explain your answer.



DATA TABLE

Av. Force	Separation
28.4 units	0.75 cm
16.0	1.00
10.2	1.25
	1.50
	1.75
4.00	2.00
1.80	3.00
1.00	4.00

Figure 21.26
For question 30.

(d) How would you check your hypothesis given in answer **(b)**?
(e) Carry out the suggestion you made in **(d)**.
(f) Suggest how the distance between the centres of the two spheres could have been measured.
(g) Explain why the student would have made repeated measurements instead of just one reading in each case.

Extension — complex, challenging and novel

*****31** Two charged particles Q (+2.5 μC) and q (+1.0 μC) are separated by varying distances, d, and their mutual force of repulsion, $\boldsymbol{F}$, is measured. Tabulated data are presented below:

$\boldsymbol{F}$ (N)	2.3	0.36	0.14	0.06	0.04	0.028
d (m)	0.1	0.25	0.4	0.6	0.85	0.9

(a) Which of these data points is in greatest error?
(b) Determine graphically a value for the Coulomb constant k.

*****32** Figure 21.27 represents a simplified diagram of the evacuated interior of a computer monitor's CR tube. Electrons are accelerated from the gun filament, pass through point X and enter the region between the deflecting plates YY_1, before finally striking the screen. If the electron mass is 9.1×10^{-31} kg:
(a) calculate the time of travel within the deflecting region YY_1;
(b) state the vertical displacement of the electron in the deflecting region YY_1;
(c) what is the electron velocity as it reaches the screen at the end of the CR tube?

Figure 21.27
For question 32.

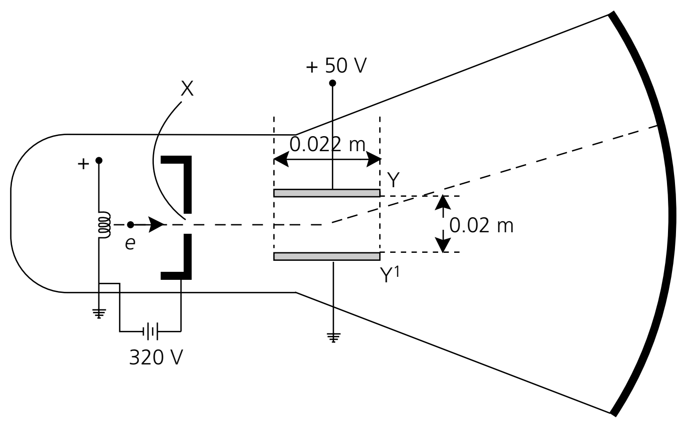
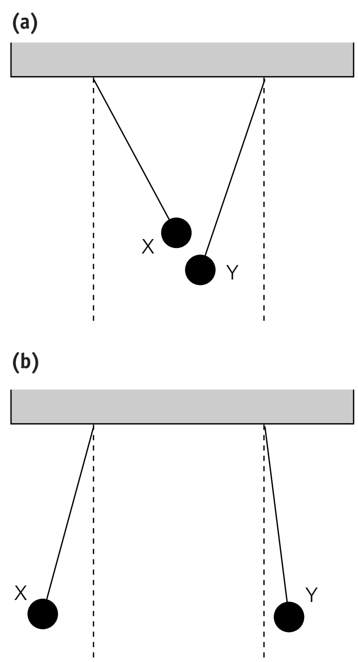


Figure 21.28
For question 33.



*****33** Two metal-coated spheres X and Y are suspended from light insulating threads of equal length. The spheres are of equal radii, and each carries an electric charge. Figure 21.28**(a)** shows the positions of the charged spheres at equilibrium. If the two spheres are touched together and then separated they come to rest in new equilibrium positions, as seen in Figure 21.28**(b)**. A student who was asked to explain these results makes the following 'correct' deductions:
(a) The sign of the charges on the two spheres was not the same.
(b) The magnitude of the charges on the two spheres was unequal.
(c) The mass of sphere X was less than that of sphere Y.
Use the information contained in the figures to justify each of the student's deductions.

CHAPTER 22

Electric Circuits

22.1 INTRODUCTION

Electrical energy is very important in our lives, as evidenced by the great inconvenience when it is not available — for instance, trying to cope without refrigeration when camping, or trying to cope without cooking appliances during electrical blackouts caused by industrial strikes or storm damage. The great explosion in technology in the twentieth century has been almost solely due to applications of electrical energy. The information age to which we belong would not be possible without methods of distributing large quantities of electrical energy for operating appliances or being able to store and transmit information across telephone, television and computer networks.

The electrical age began around 200 years ago when it was discovered how to store electrical energy and thus control it, rather than just deal with its electrostatic effects. The Italian scientists **Luigi Galvani** (1737–98) and Count **Alessandro Volta** (1745–1827) experimented with electrical effects in animal tissues, showing that the nervous system is electrically operated. Volta was able to produce the first example of an electric battery, which he called a 'pile', constructed from a series of pairs of dissimilar metal electrodes separated by moist cloth layers. Today we call this apparatus a voltaic battery or just a battery.

In this chapter we will examine the various effects of electric current, together with a model for its behaviour and the laws under which it flows in circuits. By the end of the chapter questions such as:

- what causes electric current to flow?
- what controls the direction of electric current flow?
- how is electric current measured?
- which is the more dangerous, voltage or current?
- how can we use electrical energy safely?
- how will electrical energy affect me in the future?

which you may have asked in the past, will be able to be answered satisfactorily.

Although electrical energy is widely used in modern society, it should never be treated lightly as it can become extremely dangerous when used inappropriately.

NEI Activity 22.1 ELECTRICITY AROUND THE HOME

1. Examine the following electrical devices found around the home and find the voltage marked on each: torch battery, car battery, calculator battery, watch battery.
2. Look at several light bulbs and determine the wattage rating marked on each.
3. Determine how many electric cable wires are coming into your house from the distribution pole in your street.

ELECTRIC CHARGES IN MOTION 22.2

In Chapter 21 we saw that an electric potential difference applied across a set of parallel plates causes an electric field with a resulting force acting on any electric charges within the field. In this chapter this idea will be taken further to define the nature of electric current in various types of conductors.

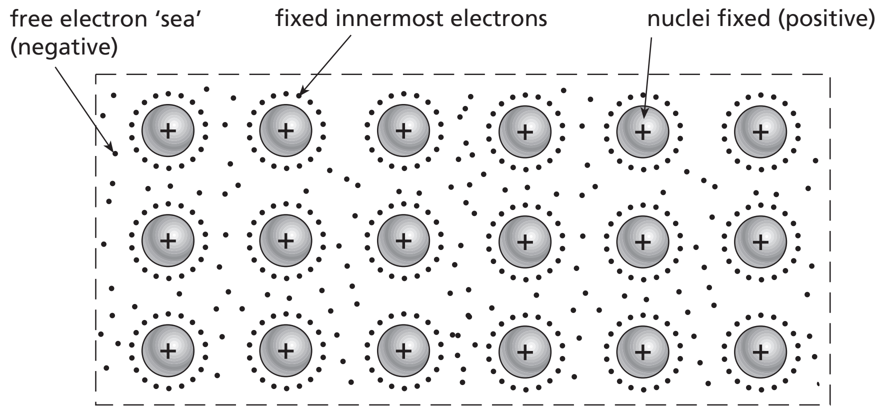


Figure 22.1
Metallic lattice. Positive nuclei in a sea of mobile electrons.

Generally metals are very good conductors of electric current. (Refer to Section 21.3.) This is because metal elements contain loosely bonded **valence electrons** in their outermost atomic electron shells. These are available for shared bonds with other nuclei. This atomic bonding pattern within blocks of metals creates a virtual 'sea of electrons' within the metal, allowing very easy motion of the electron particles under the influence of an applied electric field (Figure 22.1). The non-metal graphite is also a good conductor because of a similar pattern of loosely bound electrons within its solid crystalline lattice structure. Within metals it is therefore the negative electron particles that are free to move through a fixed nuclei lattice of positive charges.

Consider crystalline solids such as sodium chloride (common salt — NaCl). If this substance is dissolved in water, it dissociates into positive sodium **ions** (neutral atoms that have lost electrons) and negative chloride ions (neutral atoms that have gained electrons). Salt solution is referred to as an **electrolyte**. If an electric field is placed across this type of material, then charge movement of both positive and negative ions will occur. This is referred to as electrolyte conduction. A typical voltaic or electrolytic cell is shown in Figure 22.2, involving both electrolyte conduction and metallic conduction.

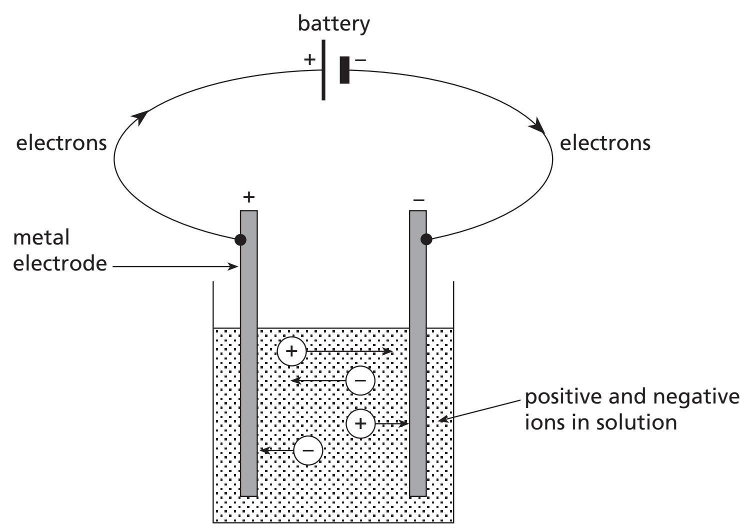


Figure 22.2
Electrolyte conduction.

Gaseous substances are normally insulators and will not conduct electricity due mainly to the wide spacing between nuclei and possible charge carriers. Gases can be made to conduct if the atoms are given enough energy by either heating or high voltage, or by irradiating with ultraviolet light or X-rays. Under these conditions the atoms of the gas are stripped of some of their electrons, the atoms become ionised, and charge motion due to both ions and free electrons can occur (Figure 22.3).

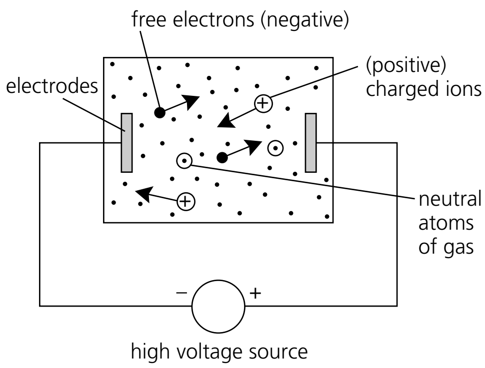


Figure 22.3
Gaseous conduction.

Let us return to conduction within typical metals. Consider a piece of copper conductor that has been drawn out into a fine wire owing to its very good **ductility**. The valence electrons within the metallic lattice are moving about at very high speed in random directions. If an external electric field is applied through the copper wire by means of a potential difference across its ends, then the free electrons will move under the influence of electric forces towards the higher potential. Remember, any one particular electron will experience a force $\boldsymbol{F} = q\boldsymbol{E}$ (Figure 22.4).

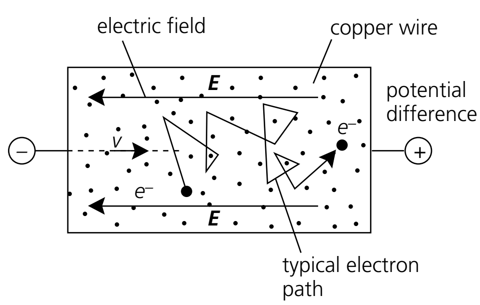


Figure 22.4
Drift velocity in metals.

Because the metallic lattice contains large numbers of nuclei, the electrons in motion undergo collisions that slow their progress. In general, the electrons drift at a particular terminal velocity characteristic of the conductor, which is known as the **electron drift velocity**, v. Typical metals have values for drift velocities of about 1×10^{-4} m s^{-1}. The flow of electric current along the wire occurs much more rapidly because an electrostatic repulsive pulse between neighbouring electrons occurs as soon as the electrons begin to move under the influence of the applied electric field. We would imagine an electron entering one end of the wire and, almost instantaneously, another electron being repelled from the opposite end of the wire. The actual electric current, I, flowing along the wire is the total number of electrons, q, passing any given point in the wire every second, t. If the rate of flow is constant then:

$$I = \frac{\text{charge}}{\text{time}} = \frac{q}{t}$$

The unit of electric current is the ampere or amp (A). One **ampere** of electric current is thus a flow of charge of 1 coulomb per second. 1 A = 6.25×10^{18} electrons per second. One ampere of current is quite large. In most electric circuits currents of microamps (μA) or milliamps (mA) are more common. Electric current is measured with an ammeter. (Refer to Section 22.5.)

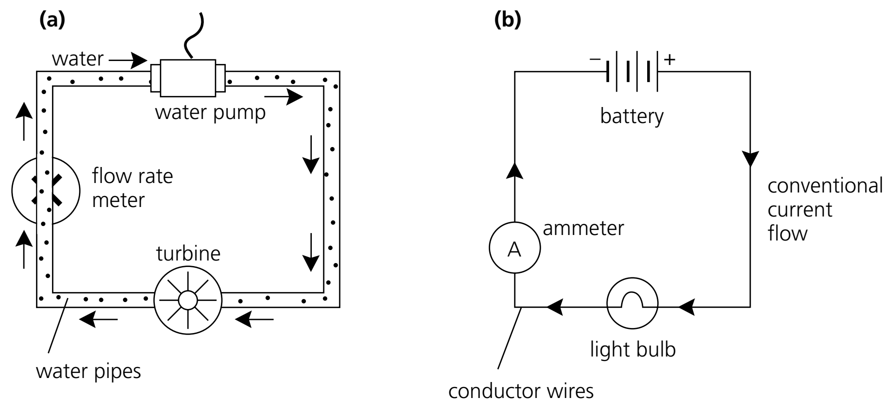


Figure 22.5
The water model.

It is very useful to consider an analogy such as a water model when considering the flow of electric charge. (See Figure 22.5.) The water mechanical pump is the equivalent of the electric battery. The water pipes are the equivalent of the electrical conductors and the water itself is analogous to the electric charges in motion, that is, the current. Note that as the water flows around the pipe circuit it can provide energy to run a water wheel, just as charge flowing around an electric circuit can provide energy to operate a light bulb. It is important to realise that in the water pipes water never gets used up, it just keeps getting recycled. The same thing occurs with electric charge. The electric charge doesn't get used up; it will keep flowing until the battery potential difference is reduced to zero as a result of energy transferred to the light bulb.

An electric **circuit** must be a complete closed loop path. In this case, the charges flow from the battery through the conductors to the light bulb and deliver the energy given to them by the battery. If the path is not complete, charge will not flow and the current stops. This is called an **open circuit**. If the battery terminals are connected directly together without the circuit containing a device such as a light bulb to restrict the amount of charge flowing, then a **short circuit** occurs. This is a very dangerous situation as the very large current that may flow can cause heating of the conductors, and subsequent fires. In fact, it is possible to cause sparking and welding of the metal conductors when very large batteries are short-circuited.

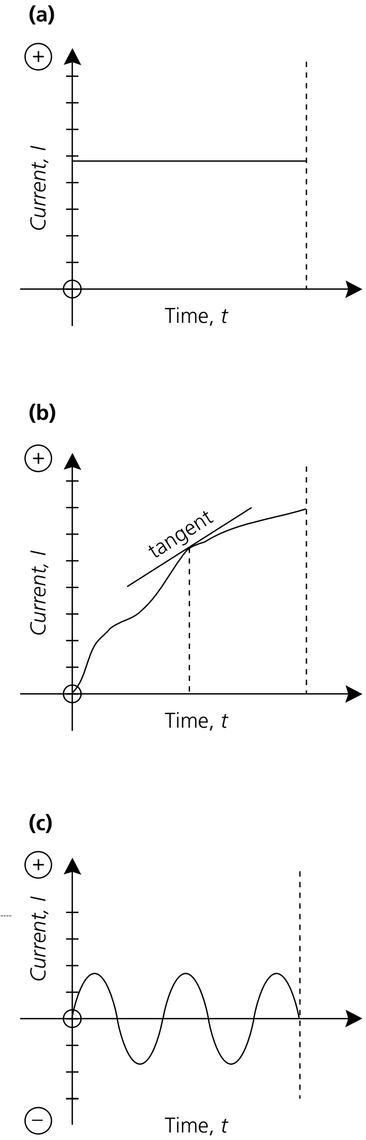
It is a historical fact that experiments were carried out with electricity and electrostatic charge long before the nature of atomic electrons was discovered. Benjamin Franklin had originally used the term 'positive charge' in electrostatics and early experimental work on electric current assumed that the charges in motion in conductors were positive charges and that they flowed along conductors from the higher positive potential to the lower negative potential (a little like water naturally flowing downhill). It is still common in physics and electric circuit analysis to refer to **conventional current** as the motion of positive charges from positive to negative. This is the convention used in this textbook. **Electron flow** is the direction of actual electron particle motion in a conductor from negative to positive. One amp of conventional current in one direction is the same as one amp of electron flow in the opposite direction. Recall that the convention for E, electric field direction, is that in which a positive test charge will move.

When electric charge flows from the source of charge around a circuit in the one direction, as in Figure 22.5, the type of current is called **direct current** (DC). Batteries and voltaic cells provide DC. The magnitude and direction of the flow is constant over time.

Ch22-Walding 4th 25/8/04 12:27 PM Page 470

Figure 22.6**(a)** illustrates this I,t relationship graphically. If the magnitude of the rate of flow of charge changes without the direction of flow changing in a circuit, then the instantaneous electric current can be found from the slope of the I,t graph as in Figure 22.6**(b)**. An electric generator device involving rotating coils of wire in a magnetic field (refer to Chapter 26) will produce electric currents that vary in both magnitude and direction many times per second. This oscillating type of current flow in a conductor is called **alternating current** (AC) and is represented graphically in Figure 22.6**(c)**. Industrial and household electricity is distributed via this type of current flow. (Refer to Sections 22.6 and 22.7.)

Figure 22.6
AC and DC waveforms.



Example

A particular type of metal has an estimated 1×10^{23} free electrons per metre of its length. If this metal carries an electric current of 1.4 amps, estimate the drift velocity of the electrons through the metal.

Solution

- Let the conductor contain n electrons per metre.
- Let the electron drift velocity be v m s^{-1}.
- In a time of t seconds, a total number of electrons = nvt will pass a given point.
- This represents a total charge $Q = nqvt$, if the charge on one electron is q coulombs.

But electric current is calculated:

$$I = \frac{Q}{t} \qquad I = \frac{nqvt}{t} = nqv$$

thus:

$$v = \frac{I}{nq} = \frac{1.4}{1.0 \times 10^{23} \times 1.6 \times 10^{-19}}$$

$$v = 8.7 \times 10^{-5} \text{ m s}^{-1}$$

which represents the drift velocity in the metal.

22.3 ELECTROMOTIVE FORCE AND VOLTAGE

Recall that in the previous section it was stated that a battery can supply direct current (DC). Let us take this idea further to see why this is possible. A **simple cell** consists of two dissimilar metals separated by a conducting solution. If several simple cells are joined together the combined arrangement is called a **battery**. For example, the normal 12 volt car battery usually has six individual cells connected together. The simple cell allows two dissimilar metals to be separated by a conducting electrolyte solution or paste. In a copper-zinc cell such as in Figure 22.7, chemical reactions cause the copper electrode to become positively charged and the zinc electrode to become negatively charged with a potential difference of about 1.0 V. The familiar carbon-zinc dry cell (carbon-positive, zinc case-negative) produces a potential difference of about 1.5 V.

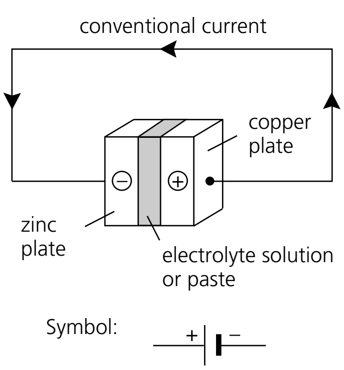


Figure 22.7
A simple cell.

NOVEL CHALLENGE

In the 1950s, car manufacturers changed from 6 V to 12 V batteries. Why do you suppose they did this? Are motorcycle batteries 12 V as well? Is the positive or negative of the battery connected to the car's body?

When a simple cell is constructed, the metal plates are separated by a salt solution, a weak acid solution or a conducting paste. The electric potential difference still exists between the positive and negative metal electrodes of the cell. This potential difference is measured in volts (V) and is usually referred to as the **electromotive force** (EMF) of the cell or simply its **voltage**. The term electromotive force is a historical reference to the original idea that a simple cell might force charges to flow in an external conductor. Experimentation with different combinations of metals in simple cell arrangements has identified a list, in order of the effectiveness in producing an EMF. This list is given in Table 22.1 and is known as the **electrochemical series**. Generally, the further apart two metals are located on the electrochemical series, the greater is the voltage (EMF) produced. Note the positions of familiar metals used in modern battery technology, such as lithium, manganese and nickel–cadmium.

Table 22.1 ELECTROCHEMICAL SERIES OF METALS

METAL		BEHAVIOUR
Lithium		metals
Caesium		donate
Barium		electrons
Calcium		
Magnesium		
Aluminium		
Manganese		
Zinc		
Iron		Potential difference
Cadmium		
Cobalt		
Nickel	e^-	
Tin		
Lead		
Copper		metals
Silver		accept
Gold		electrons

Photo 22.1
The battery in a car.

Once several cells are connected together and a battery is produced, the device can be used to provide electric current or a flow of charges in an external circuit connecting the positive and negative terminals, as shown in Figure 22.8. The battery itself is the energy pump that raises the charge to a higher electric potential at the positive terminal. Positive conventional current will flow from the positive terminal of the battery through the external circuit conductors and back to the negative battery terminal. Along this pathway the charge loses potential energy as it does work in various circuit elements. Potential energy is converted to other forms such as thermal, kinetic or light energy. Remember that in this process of energy conversion, the charge is not destroyed or used up, but its electric potential energy is reduced back to zero as it reaches the negative battery terminal. The battery will restore the electric potential energy of the charges back again to a high value. Figure 22.8 also illustrates this changing electric potential energy state of the charge carriers in the circuit. Remember, actual electron flow in the circuit is opposite to conventional current. In an electric circuit, electric potential is a measure of the electric potential energy per unit charge. More often than not, the most important feature in an electric circuit is the electric potential difference, which is commonly referred to as either a **potential rise** or a **potential drop**. The battery is

Ch22-Walding 4th 25/8/04 12:27 PM Page 472

a device that produces potential rises, whereas load elements such as light bulbs, resistive elements or motors can cause potential drop.

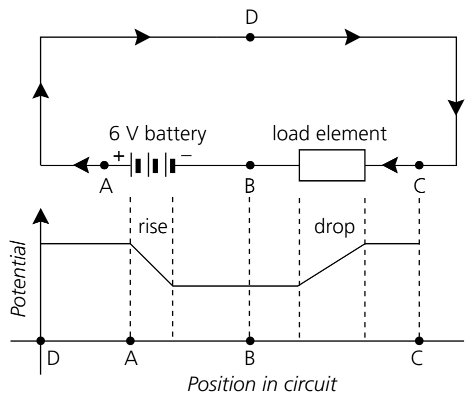


Figure 22.8
A simple circuit and potentials.

$$\text{Potential difference} = \frac{\text{work done to move charge}}{\text{amount of charge moved}}$$

$$V = \frac{\Delta W}{q}$$

where V is measured in volts (V); ΔW is measured in joules (J); q is measured in coulombs (C).

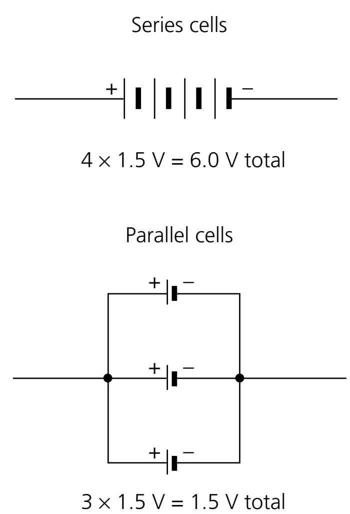
A potential rise of 1 volt means that a source of electric energy will give 1 joule of energy to each coulomb of charge that passes through it. A potential drop of 1 volt means that a load element such as a light bulb or resistor will remove 1 joule of energy from each coulomb of charge that passes through it. Recall from the electric current discussion in Section 22.2 that the current $I = \frac{q}{t}$, hence it is easy to show algebraically that $\Delta W = VIt$, which has the units of energy change.

Notice that the battery symbol, illustrated in Figure 22.8, actually consists of three separate cells connected together so that positive electrodes are directly connected to negative electrodes. This type of connection is called a **series connection** of the cells and the total EMF is the sum of the EMFs of each cell.

Series connection total EMF = sum of individual cell EMFs

Individual cells may also be connected together so that all positive electrodes are connected together and all negative electrodes are also connected together. This is called a **parallel connection** and the total EMF is then the same as each individual cell. In this connection a battery cannot supply more energy to each electron, but can in fact supply a greater quantity of electrons per unit time or a greater current flow in any external circuit. (See Figure 22.9.) Note that only equal values of EMF should be paralleled.

Figure 22.9
Cells in series and parallel.



Parallel connection total EMF = individual cell EMF

Example

A battery is known to contain four individual cells connected in series and is able to supply 3.6 J of energy to every 0.6 C of charge passing through it. What is the potential rise (EMF) produced by each cell?

Solution

Work done is equal to energy gained, hence $V = \frac{\Delta W}{q} = \frac{3.6\ \text{J}}{0.6\ \text{C}} = 6$ V. If the total EMF of the battery is 6 V, then each cell will produce EMF = 1.5 V.

Activity 22.2

NEI

A Big cells

Use a library, Internet searching or even the *Guinness Book of Records* to establish what is the largest existing lead–acid cell in the world, and what its electrical power output is.

B The human nervous system

SR

Read the following text information and answer the questions that follow.

The human nervous system contains important nerve-conducting pathways called neurons. Each neuron consists of a cell body containing a nucleus, and outgrowths called processes. The main one of these processes is the **axon**, which is responsible for carrying outgoing messages from the cell. This axon can originate in the central nervous system (CNS) and extend all the way to the body's extremities, effectively providing a highway along which messages travel to and from the CNS.

Dendrites are smaller, secondary processes that grow from the cell body and axon. On the end of these dendrites lie the axon terminals, which 'plug in' to a cell where the electrical signal from a nerve cell to the target cell can be made. This 'plug' (the axon terminal) connects into a receptor on the target cell and can transmit information between cells. The 'all-or-none law' applies to nerve cell communication; they use an ON/OFF signal (like a digital signal) so that the message can remain clear and effective throughout its travel from the CNS to the target cell or vice versa. This is a factor because, just like electricity signals, the signal fades out and must be boosted along its journey; if the message is either 1 or 0 (i.e. ON or OFF) the messages are absolute.

Nervous cells are classified into **inter-neurons** (neurons lying entirely within the CNS); **afferent neurons** (also known as sensory neurons — specialised to send impulses *towards* the CNS and *away from* the peripheral system); and **efferent neurons** (carrying signals *from* the CNS *to* the cells in the peripheral system).

When it was discovered over a century ago that nerve-impulses involved electric charges, it was assumed that a nerve impulse was simply an electric current flowing through a nerve, just as electric currents flow through conducting wires. Measurements of actual electric currents in nerves proved that the conduction process could not solely be due to conduction along the nerve fibre as its resistance was far too high and the speed of conduction was far too slow. It was proposed by Julius Bernstein in Germany in 1902 that nerve conduction primarily involved an electrochemical process. Bernstein suggested that the permeability of the nerve cell membrane varies for different ions in solution, especially sodium and potassium ions, and that the selectivity of the nerve membrane maintains the separation of ions and thus the electric potential. With subsequent modifications to the original Bernstein theory, the transmission of nerve impulses through neurons following stimulation is thought to occur as follows.

NOVEL CHALLENGE

Make up five questions that would test a person's understanding of electric current. Think about whether your questions are just about recall of facts or are really testing the person's understanding.

The membrane of a resting neuron is polarised; that is, the inside is negative relative to the outside. The concentration of sodium ions is greater outside, while the concentration of potassium ions is greater inside the neuron. Any stimulation causes the membrane to undergo a change allowing sodium ions to rush into the cell, which causes the inside to become positively charged relative to the outside. Very quickly, the membrane becomes permeable to potassium ions and they now rush out of the

Ch22-Walding 4th 25/8/04 12:27 PM Page 474

cell, which restores the inside of the cell again to a negative state. It is this rapid reverse polarisation of the membrane at successive points along the cell that constitutes a transmitted nerve impulse. Following the passing of the electrical impulse the ionic balance of the cell is restored back to its usual resting state by a biological ion exchange pump. It can be seen that the separation of charge and the subsequent electric potentials are vitally important in biological systems and not just in non-living systems. Within these biological systems the electric potential differences are of the order of 50–90 mV and are known as **action potentials**.

When multiple cells depolarise, either simultaneously or sequentially, they generate an electrical waveform which can be detected by external electrical circuits. For example, the depolarisation of cardiac cells produces the ECG (electrocardiogram). These millivolt signals can be detected electronically by either bipotential or instrumentation amplifiers.

Questions

1. Why is the human nervous system like an electrical circuit? Do you think Ohm's law might apply to the circuitry? Refer to Section 22.4 to help you answer this.
2. Draw a diagram of what a typical human neuron might look like from the description above. Compare this with one you will find in any good Biology textbook. Oxford texts are best!
3. If nervous transmission is an 'all-or-none' system, what are the voltage amplitudes of the electrical switch signals that are travelling the body?

22.4 RESISTANCE AND RESISTIVITY

When a potential difference is applied across a metallic conductor, the electrons do not move very rapidly along the conductor. The electrons are accelerated by the applied electric force field due to their small mass; however, they very quickly collide with the positive metallic lattice ions in the conductor and lose energy. This rapid acceleration and subsequent collision leads to the average electron drift velocity, as discussed in Section 22.2. The magnitude of current through a conductor is proportional to the drift velocity of electrons through it. The effect of the collisions within the lattice is to reduce the current. This is the same as occurs in a stream that contains a lot of rocks, trees and other debris that reduce the rate of flow of water along it. Every time an electron collides with one of the metallic lattice ions, it loses energy, which is transferred to the lattice as heat and vibrational energy. This means that the temperature of the conductor increases. This opposition to the flow of electric current that any conductor produces is called its **electrical resistance**. The smaller the value of current that flows as a result of any given applied voltage, the larger the resistance.

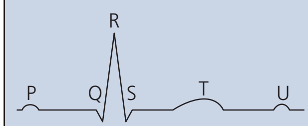
Several factors determine the resistance of a conductor. Firstly, the longer the length L of a conductor, the greater the number of collisions occurring, while the likelihood of a collision is decreased if the conductor's cross-sectional area A is increased. Different conductors will have varying lattice types. For example, if the atomic lattice is tightly packed, more collisions are likely. The type of material from which any conductor is made controls the property called the resistivity, rho (ρ). Thus the overall electric resistance R is given by

$$R = \rho\frac{L}{A}$$

where ρ = resistivity measured in Ω m; L = length measured in m; A = cross-sectional area measured in m^2.

PHYSICS FACT

To obtain a standard ECG (as shown in the figure below), a patient is connected to the machine with three electrical leads (one to each wrist and another to the left ankle) that continuously monitor heart electrical activity using an instrumentation amplifier.



Each peak in the ECG is identified with a letter from P to U that corresponds to a specific electrical activity of the heart: The P-wave represents the electrical excitation (or depolarisation) of the atria, which leads to the contraction of both atria. The QRS complex represents the depolarisation of the ventricles, which initiates the ventricular contraction. The contraction starts shortly after Q and marks the beginning of the systole. Voltage peaks are of the order of 1.0 millivolts. The T-wave represents the return of the ventricles from excited to normal state (repolarisation). The end of the T-wave marks the end of the systole. The U-wave is usually very small and represents the repolarisation of a collection of specialised muscle fibres that make up the pacemaker system, which is responsible for spreading the electrical signal throughout the ventricle. Obviously, by counting the number of QRS complexes that occur in a given time period, one can determine the heart rate of an individual, but an ECG can give a lot more information. For example, since the ECGs obtained from different individuals have roughly the same shape for a given lead configuration, any deviation from this shape indicates a possible abnormality or disease.

For you to investigate

Draw the waveform for a person with: tachycardia, atrial fibrillation and ventricular fibrillation. Explain what each of these conditions means.

The measurement unit for resistance is the **ohm**, named after **Georg Simon Ohm** (1787–1854), a German physicist. Its symbol is the Greek letter omega (Ω). Electric meters known as ohmmeters, as well as multimeters, can be used to measure any resistance value for a particular conductor; however, a simple experimental circuit can also be used, as shown in Figure 22.10.

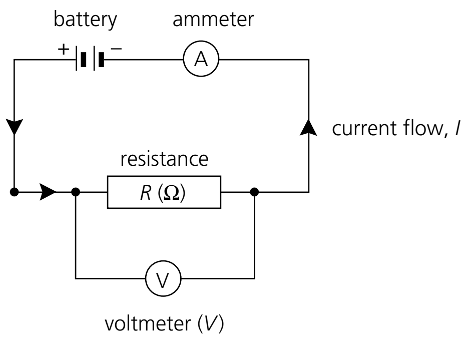


Figure 22.10
Measuring resistance.

This circuit measurement depends on the relationship between applied voltage and subsequent electric current through the resistance, which is known as **Ohm's law** and which is discussed more fully in Section 22.5, but is stated as:

The current flowing through a conductor is directly proportional to the potential difference applied across its ends, provided temperature and all physical conditions remain constant.

The ratio of V to I is defined as the electrical resistance, R. Thus, mathematically, I is proportional to V:

$$\frac{V}{I} = R \quad \text{or} \quad V = IR \quad \text{or} \quad R = \frac{V}{I}$$

A conductor would have a resistance of 1 ohm (1 Ω) if a potential difference of 1 volt (1 V) across its ends produces a current of 1 amp (1 A) flowing through it.

Table 22.2 ELECTRIC RESISTIVITY AND TEMPERATURE

MATERIAL	RESISTIVITY, ρ (Ω m)	TEMPERATURE COEFFICIENT, α (°C^{-1})
Silver	1.5×10^{-8}	4.1×10^{-3}
Copper	1.7×10^{-8}	4.1×10^{-3}
Aluminium	2.6×10^{-8}	3.9×10^{-3}
Iron	8.9×10^{-8}	6.2×10^{-3}
Platinum	9.8×10^{-8}	3.7×10^{-3}
Mercury	94×10^{-8}	0.88×10^{-3}
Nichrome	100×10^{-8}	0.4×10^{-3}
Carbon	5×10^{-5}	-5×10^{-4}
Silicon	600	-700×10^{-4}
Fused quartz	$\approx 10^{17}$	

Like many physical properties, resistivity not only depends on the material involved but also on the temperature. Table 22.2 lists the electric resistivity, ρ, properties of various materials as well as their temperature coefficient of resistivity, α, values. The resistivity of pure metals increases linearly with temperature because a temperature increase causes the lattice ions to vibrate with greater amplitude. This increases the likelihood of electron collisions and decreases the current through the conductor. The expression for the increase in resistivity with temperature for any conductor is:

$$\rho_T = \rho_0\ (1 + \alpha\Delta T)$$

and since resistivity is proportional to the resistance, R, then we can also write:

$$R_T = R_0\ (1 + \alpha\Delta T)$$

where R_T = conductor resistance at a temperature of T°C; R_0 = conductor resistance at a temperature of 0°C; α = temperature coefficient of resistivity °C^{-1}; ΔT = temperature change in °C.

A practical application of the resistance change with temperature is a **resistance thermometer**.

NEI Activity 22.3 NOT YOUR NORMAL THERMOMETER!

Platinum is a metal with a very high melting point and reasonably high resistivity value. Using the library, or encyclopaedia, research the construction and method of operation of a platinum resistance thermometer. Describe how its variable resistance characteristic could be measured in a practical situation where it might be used to determine the operating temperature of a furnace.

Example

What is the electrical resistance at 0°C of a piece of copper wire whose length is 1.2 m and whose cross-sectional area is 17.2 mm^2? How would the electrical resistance change if the copper wire temperature was raised to 25°C?

Solution

From Table 22.2, the resistivity for copper is $1.7 \times 10^{-8}\ \Omega$ m at 0°C. Using:

$$R_0 = \rho\frac{L}{A} = \frac{1.7 \times 10^{-8} \times 1.2}{17.2 \times 10^{-6}}$$

$$R_0 = 1.18 \times 10^{-3}\ \Omega$$

and using the value of the coefficient from Table 22.2 for copper:

$$R_T = R_0\ (1 + \alpha\Delta T)$$
$$R_T = 1.18 \times 10^{-3}\ (1 + 4.1 \times 10^{-3} \times 25)$$
$$R_T = 1.3 \times 10^{-3}\ \Omega \text{ at } 25°\text{C}$$

Notice in Table 22.2 that some substances like silicon and carbon actually posses a negative temperature coefficient and thus will decrease their resistance as the temperature increases. This feature can lead to difficult handling methods when trying to control their temperature in working electronic circuits.

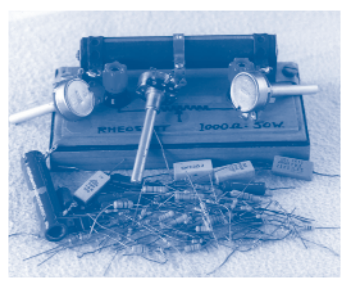
Practical resistors vary in design and size and are very common in electronic devices. They will be further discussed in terms of this application in Chapter 23; however, resistive elements are used wherever electrical energy needs to be converted into heat energy, such as in domestic electrical appliances like room heaters, stove elements and hot water systems. The resistive element in these applications is usually made of an alloy containing nickel and

NOVEL CHALLENGE

A A wire is stretched to twice its length. What happens to its resistance? Justify your answer mathematically.

B Can a copper wire and an aluminium wire of the same length (say 1.0 m) have the same resistance? There are at least two answers and you should justify these mathematically.

Photo 22.2
Various resistors.



chromium metal called nichrome wire. Notice its very high resistivity value in Table 22.2. Quite often, long lengths of resistive wire are wound on special insulating formers and, in conjunction with a sliding contact, produce a device whose resistance can be varied. It is known as a **rheostat**. This type of device is used to control voltages in electric circuits and to act in conjunction with electric motors and dimmer switches. The common volume control knobs on radio, television and stereo equipment are always simple variable resistor (rheostat or **potentiometer**) components. Tungsten wire forms the filament of modern incandescent light bulbs. The high resistivity of tungsten, especially when allowed to heat up within the controlled environment of a halogen gas inside the glass bulb, causes very bright light to be emitted with good efficiency.

Human skin is a very good insulating material, luckily, but the tissues and fluids just beneath the skin contain a large number of ions and hence conduct electricity very efficiently with low resistance. The variable skin resistance can be measured with sensitive equipment and is often the basis of the American Justice System polygraph or lie-detector test.

Another application is in diagnostic medicine. By producing a small electric current between two points on the body surface it is possible to measure the electric resistance. Usually one electrode is attached to a patient's leg and another is moved over the body surface with a voltage applied between the electrodes. The electrical path resistance varies especially near abnormalities such as nerve damage or tumour tissue locations. This technique is quite useful in the detection of cancer, for instance.

Internal resistance

Finally in this section, let us look at the ideas of **internal resistance** of a battery and connections of resistances in series and parallel. In a battery, chemical energy is continuously being converted into electric energy when the battery is in use. During this process, internal heat is produced and the amount of heating is dependent on the current being drawn from the battery. Thus the battery behaves as if it had an internal resistance, *r*.

Figure 22.11
Battery internal resistance.

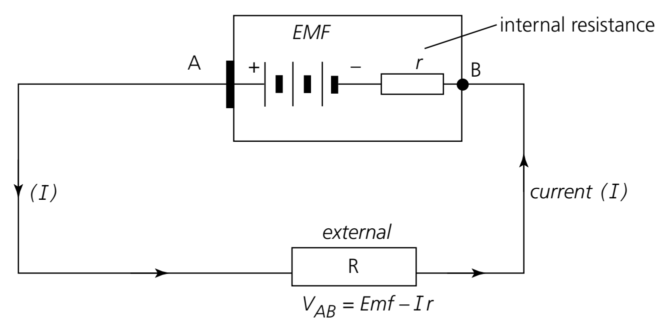


Figure 22.11 shows the circuit diagram for a battery supplying current to an external resistance, *R*, and includes the battery's internal resistance. Because the current, *I*, passes through the battery as well, a potential drop of *Ir* is caused, which subtracts from the battery EMF. Thus the terminal voltage of the battery, as measured by a voltmeter, would be:

$$V_{AB} = EMF - I \times r$$

Notice that, if the battery is not supplying any external circuit current, the terminal voltage equals the EMF. When the car is started, a very large current, in the order of 100 A, must be supplied by the battery for a short period of time. If the car is started with the lights on, the lights will usually dim considerably as the terminal voltage is reduced by *Ir* across the battery itself, producing a much lower voltage across the headlights. All new batteries have an internal resistance that is quite small, say, 0.05 ohm for a typical D cell. As the battery gets older, its internal resistance builds up to such an extent that it is no longer able to deliver any useful current, due to the fact that its terminal voltage reduces to zero. We commonly call this a 'flat' battery.

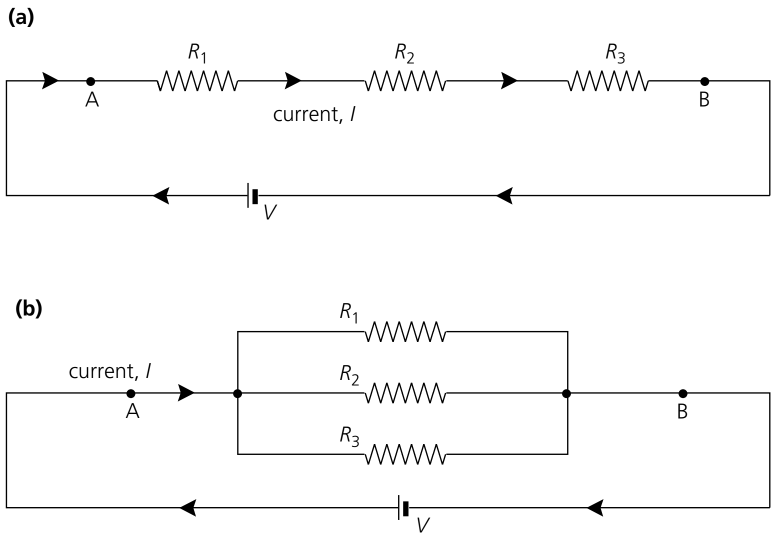


Figure 22.12

Consider Figure 22.12**(a)** which shows three resistors connected in series so that any current flowing from A to B must pass through each in turn. This effect simply adds the individual resistances to create a total of all three. It is the same as increasing the effective length of a single resistor. Thus:

$$R_{tot} = R_1 + R_2 + R_3$$

Consider Figure 22.12**(b)**, which shows three resistors connected in parallel, like the rungs of a ladder. Current flowing from A to B in this situation has three paths to take. In a sense, the total cross-sectional area of the conductor for current I is being increased and thus the total resistance is reduced. This leads to the addition of the reciprocals of each resistance to give a relationship expressed as:

$$\frac{1}{R_{tot}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Both of these connection rules will be proved in greater detail using the electrical circuit laws of Ohm and Kirchhoff in the next section.

EI Activity 22.4 CHRISTMAS AND CONDUCTIVITY

1. Inexpensive Christmas tree lights have the individual bulbs connected in series. Predict what would happen if one of the light bulbs blew. Try this out by carefully removing a bulb from a working set. Also try this with more expensive light sets where the individual resistive bulbs are connected in parallel. What is your prediction now?
2. Use the *Guinness Book of Records* to find the highest temperature at which practical applications of the principle of superconductivity (zero effective resistance) will occur.

Example

Calculate the total resistance of a pair of 25 Ω resistors connected in parallel to a battery whose EMF is 12 V. Deduce the current, measured by a DC ammeter, that will flow from the battery.

Solution

Calculate the total effective resistance:

$$\frac{1}{R_{tot}} = \frac{1}{25} + \frac{1}{25} = \frac{2}{25}$$
$$R_{tot} = 12.5\ \Omega$$

Use $V = I \times R_{tot}$ to find current as measured by the ammeter:

$$12 = I \times 12.5$$
$$I = \frac{12}{12.5} = 0.96\ \text{A}$$

Questions

1 Describe the difference in behaviour between static electricity, direct current (DC) electricity and alternating current (AC) electricity in terms of flowing charges.

2 What is the voltage necessary to move 15 coulombs of charge through a conductor, if the energy required is 80 joules?

3 A piece of metal conductor is estimated to contain 3×10^{22} electrons per metre of its length. If it carries a current of 1.5 A, determine the average drift velocity of the electrons in the conductor.

4 Calculate the increase in potential produced by a cell if every 2.6 coulombs of charge passing through is supplied with 3.9 joules of energy.

5 A voltage of 120 V is applied to a bulb whose resistance is 200 Ω.
(a) What is the current through the bulb?
(b) How much charge flows through the bulb every hour?

6 A pigeon stands on a 100 kV high-tension wire that carries 50 A. If the line resistance is $2.0 \times 10^{-4}\ \Omega\ \text{m}^{-1}$, calculate the voltage across the bird if its feet are 3.0 cm apart. What can you deduce about the likelihood of the pigeon being electrocuted?

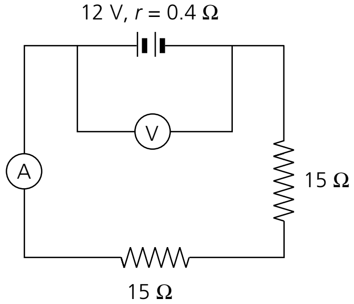
7 You are asked to design the electrical parts of an electric toaster. Describe the nature of the electrical conductors you might use.

8 The resistance of a certain metal conductor, A, is found to be 0.36 Ω at 25°C. If you found another conductor, B, made of the same material but different in characteristics as shown below, calculate in each case the new resistance value of B compared with A.
(a) Conductor B is three times longer than A.
(b) Conductor B is only half the cross-sectional area of A.
(c) Conductor B has just been taken from an oven operating at 350°C.

9 Refer to Figure 22.10. A student set up the experimental apparatus and found the following results as tabulated. Plot a current versus voltage graph and determine the value of the resistance at all points. Describe what you find. Is the resistor 'ohmic' in its characteristics? Explain.

Current (A)	0.0	0.013	0.25	0.04	0.05	0.06	0.08
Voltage (V)	0	2	4	6	8	10	12

Figure 22.13
For question 10.



10 Consider Figure 22.13, showing an EMF source and its internal resistance connected to two series resistors of value 15 Ω. Calculate the readings on the circuit meters shown.

22.5 ELECTRIC MEASUREMENT AND CIRCUITS

In this section we will look at:

- important electric measurement meters
- the laws of DC electric circuits and how to draw them using correct symbols
- methods of analysing the circuits to calculate values of current, voltage and resistance.

Let us first deal with the basic laws of electric circuits. Recall that in the previous section, an experimental circuit was discussed that allowed the measurement of resistance. (Refer to Figure 22.10.) If this circuit has a variable source of EMF, a data table of current flow, I versus applied voltage, V obtained, and a graph of the results drawn, then an important set of conclusions can be drawn (Figure 22.14).

- We find that a linear relationship exists between current, I, and voltage, V, for most types of resistors or resistive elements.
- The graph of current versus voltage is a straight line that passes through the origin.
- The slope of this line is a constant value.
- If a different value of resistance is used, the same type of relationship is found but the graph has a new slope.

These conclusions were first reached by Georg Simon Ohm (1787–1854), a German physicist, and they are summarised as a general property of materials, called Ohm's law:

> The current flowing through a conductor is directly proportional to the potential difference applied across the ends of the conductor, provided temperature and other physical factors are kept constant.

The measured ratio of voltage to current is defined as the conductor resistance, R.

From the definition of resistance, we obtain the equivalent forms $V = IR$ and $I = \frac{V}{R}$. The equation $V = IR$ is often used in circuit calculations to evaluate resistance and link voltage to current. It may represent Ohm's law.

All conductors that obey Ohm's law are called **ohmic conductors** whereas conductors that do not are called **non-ohmic conductors**. The best examples of non-ohmic conductors are modern semiconductor devices such as transistors, diodes and thermistors and these will be discussed in Chapters 23 and 24. A graph of current versus voltage for any non-ohmic conductor will not be a straight line, but the gradient of the tangent at any point can be used to determine the instantaneous dynamic resistance at any specific voltage or current. This is useful in more advanced AC calculations and circuit analysis.

The next two circuit laws were formulated by **Gustav Robert Kirchhoff** (1824–87), also a German physicist, while studying electrical networks. The first of these laws is based on the law of conservation of electric charge and applies to junction points in a circuit; that is, points where three or more wires join together. It is usually referred to as **Kirchhoff's junction law** and is represented in Figure 22.15**(a)**.

This law is expressed as:

> The sum of all currents entering any circuit junction is equal to the sum of all currents leaving that junction point.

or symbolically:

$$I_1 = I_2 + I_3$$

The second is called **Kirchhoff's loop law** and is based on the law of conservation of energy as applied to complete closed circuit paths or loops. (See Figure 22.15**(b)**.) This law is expressed as:

> The algebraic sum of all voltage changes encountered in any complete closed circuit loop is equal to zero.

Figure 22.14
Ohm's law.

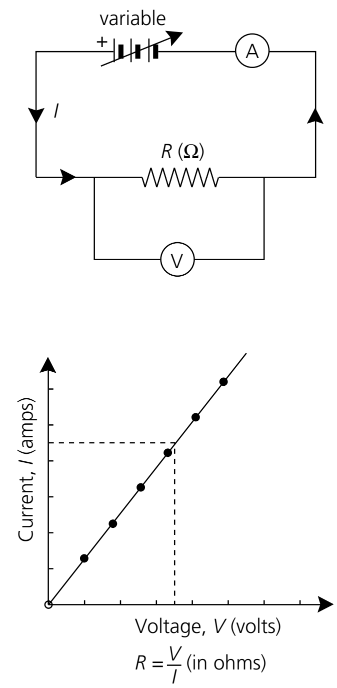
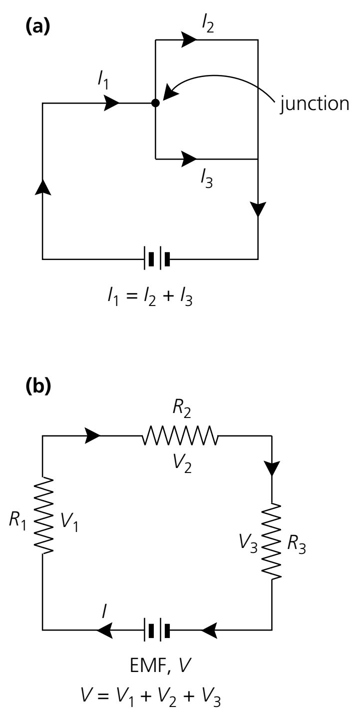


Figure 22.15
Kirchoff's circuit laws. Junction law **(a)**; loop law **(b)**.



or symbolically:

$$V = V_1 + V_2 + V_3$$

Remember that in most simple circuits with constant current flowing, electric charge gains electric potential energy in the battery and loses it within each external load element such as resistors. Hence, this voltage loop law is the same as saying that the voltage rise of the battery is equal to the sum of the potential drops across each load resistor.

At this point it is important to realise that in the analysis of most DC circuits these three laws are always used in combination, as in the next example.

Example

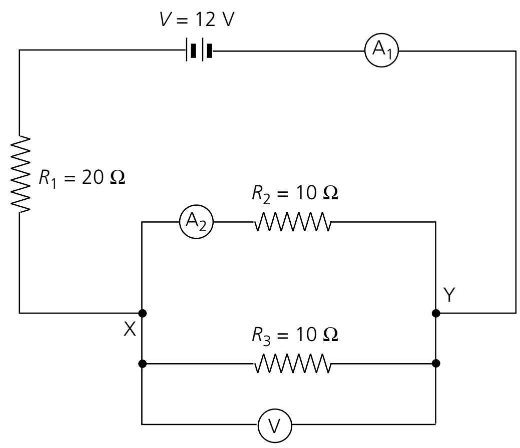


Figure 22.16
A simple network circuit.

Consider the circuit drawn in Figure 22.16 containing a simple network of three resistors connected to a DC battery of 12 V. Use circuit laws to calculate the readings on the electric meters in the circuit.

Solution

Reading the circuit, we need to find:

(a) the equivalent resistance of the parallel pair XY;
(b) the total equivalent resistance in the circuit;
(c) the total current flowing from the battery, A_1;
(d) the voltage drop across the equivalent resistance XY;
(e) the current flowing through resistor R_2 measured by meter A_2.

Thus, for the parallel pair:

$$\frac{1}{R_{XY}} = \frac{1}{R_2} + \frac{1}{R_3}$$

$$\frac{1}{R_{XY}} = \frac{1}{10} + \frac{1}{10}$$

$$R_{XY} = 5 \text{ ohms}$$

Note: the circuit could now be redrawn with only this single equivalent resistor.

Thus, total circuit resistance:

$$R_{tot} = R_1 + R_{XY} = 20 + 5 = 25 \text{ ohms}$$

Hence total current flowing from battery:

$$I_{tot} = \frac{V}{R_{tot}} = \frac{12}{25} = 0.48 \text{ A} = 480 \text{ mA}$$

Because of Kirchhoff's loop law:

$$V = V_{XY} + I_{tot} \times R_1$$
$$V_{XY} = V - (I_{tot} \times R_1)$$
$$V_{XY} = 12 - (480 \times 10^{-3} \times 20)$$
$$V_{XY} = 2.4 \text{ volts}$$

Notice that the sum of voltages around the circuit is 2.4 V + 9.6 V = 12 V.
Thus the current flowing through the resistor R_2 is given by:

$$V_{XY} = I_2 \times R_2$$
$$2.4 = I_2 \times 10$$
$$I_2 = 0.24 \text{ A} = 240 \text{ mA}$$

This is the reading on ammeter A_2.

Electric meters

With the development of the early electricity industry in the 1880s, engineers needed a simple, reliable and, above all, very fast way of measuring electric currents and voltages. The methods of the physics laboratory requiring delicate apparatus, controlled environments, careful calibrations and lengthy calculations were not suitable for the rough conditions of the industrial engineer. William Ayrton and John Perry, engineers from the Finsbury Technical College in London, devised new robust and portable instruments, which they called the ammeter and the voltmeter. It is interesting to note that the British physicists of the day were not at all impressed with these new engineering instruments. It was their opinion that the only quantities that could be measured directly were mass, length and time. They regarded all other quantities as having to be derived from these 'absolutes' by the ingenuity and skill of the experimenter and certainly not able to be read from a 'scaled instrument'. To the physicists of the Victorian era these new instruments were a threat to the moral development of students. How times have changed; today, electric meters are an integral part of any physics laboratory.

When electric measurements are made on a circuit it is obviously important that the electric meters used should only alter the circuit's behaviour in a very minor way. Any electric meter is going to have some internal resistance and this will need to be taken into account in the way in which the meter is used. The two most useful electric meters, as already seen in this chapter, are the **ammeter** and the **voltmeter**, for measuring current and voltage respectively. Both of these meters require current to operate, yet they must have negligible effects on the currents and voltages within the circuit itself. Let us see how. Both ammeters and voltmeters contain a sensitive assembly known as a **galvanometer**. This assembly contains a fine wire coil that has a pointer needle attached to it and is free to rotate within a magnetic field. The galvanometer uses the electromagnetic motor principle in which a current-carrying coil will rotate in a magnetic field. This rotation is balanced by a return spring and the needle deflection will register directly the amount of current flowing through the galvanometer coil. (See Photo 22.3.) The internal resistance of galvanometers can vary but is usually quite low (20–100 Ω) and the maximum current that the fine wire coil can carry is also very low, of the order of microamps or milliamps when producing a **full scale deflection** (FSD) across the measurement pointer scale of the instrument.

Photo 22.3
Galvanometer assembly.

In order to make a galvanometer operate as an ammeter it must be placed in series into the main circuit and also contain a current bypassing **shunt resistor** of sufficient value to prevent internal damage to the galvanometer (Figure 22.17). The parallel shunt resistor allows most of the measured current to bypass the galvanometer and not damage it. The shunt is often a piece of resistance wire.

Ch22-Walding 4th 25/8/04 12:27 PM Page 483

Figure 22.17
A galvanometer used as an ammeter.

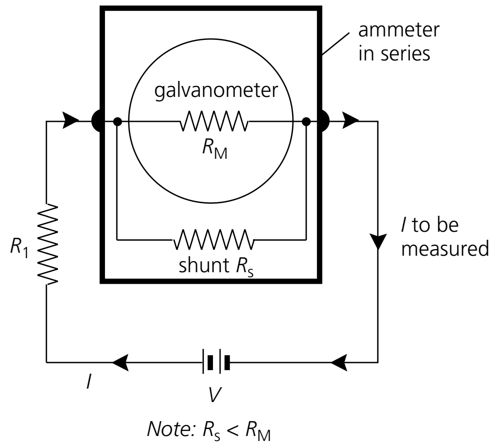
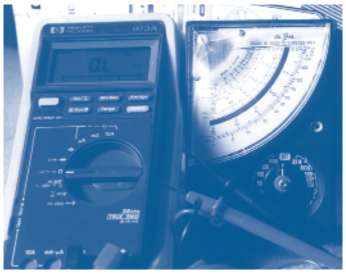


Photo 22.4
Multimeter instruments.



Example

What value of shunt resistor would be required by a galvanometer, whose internal resistance, R_M, is 25 Ω and whose FSD current, I_M, is 1.0 mA, if it is required to form part of an ammeter that will measure up to 6 A in total?

Solution

Let the current to be measured, I_{tot}, be 6 A.
Because the shunt resistor is in parallel with the galvanometer, then $V_{shunt} = V_{meter}$.
Therefore, using Ohm's law:

$$I_S \times R_S = I_M \times R_M$$

But $I_{tot} = I_M + I_S$ by junction law:

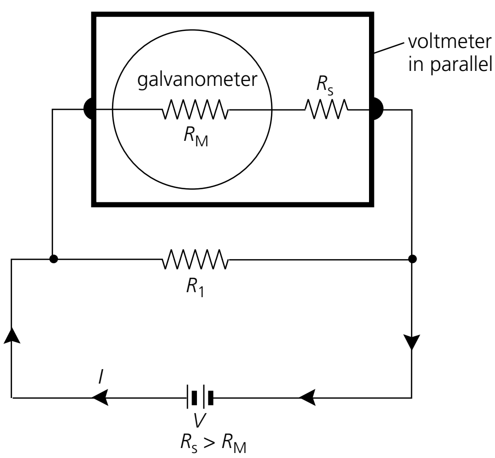
$$R_S = \frac{I_M \times R_M}{I_{tot} - I_M} = \frac{1.0 \times 10^{-3} \times 25}{6.0 - 1.0 \times 10^{-3}}$$

$$\text{shunt resistor } R_S = 4.2 \times 10^{-3}\ \Omega$$

Photo 22.5
CBL2 data-logger and graphics calculator.

A voltmeter is a galvanometer placed in parallel to the circuit component across which the voltage is being measured. Since the galvanometer typically has a low internal resistance, most of the current flowing in the circuit would flow through it and cause damage. To prevent this, a voltmeter always contains a very high value series resistor, so that the resistance of the voltmeter becomes large compared with that of the circuit component being measured and the current flow in the circuit is hardly altered. (See Figure 22.18.)

Figure 22.18
A galvanometer used as a voltmeter.



Example

What value of internal series resistor would be required by a galvanometer whose internal resistance, R_M, is 25 Ω and whose FSD current, I_M, is 1 mA, if it is to form part of a voltmeter that is required to measure up to 12 V in total?

Solution

As the series resistor and R_M are in series and together must only allow a total current flow of 1 mA through the voltmeter, then:

$$V = I_M (R_S + R_M)$$

If $V = 12$ V, then

$$12 = 1.0 \times 10^{-3} (R_S + 25)$$

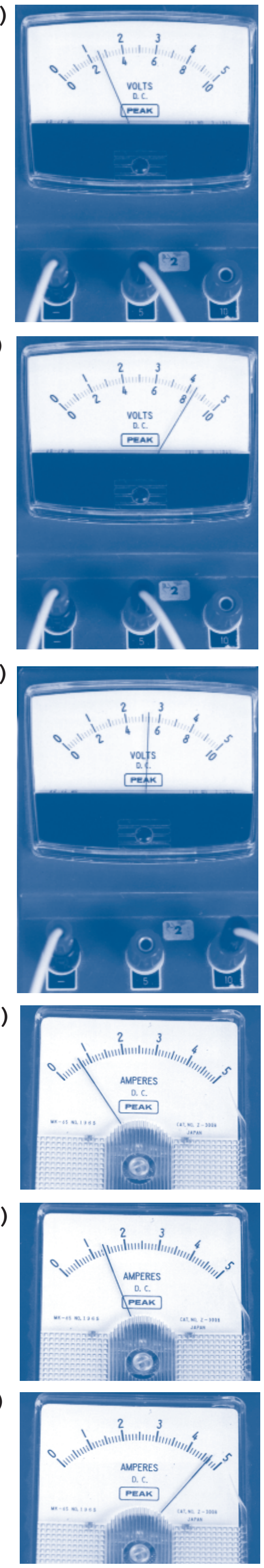
$$R_S = \frac{12}{1.0 \times 10^{-3}} - 25$$

Series resistor $R_S = 12\ 000\ \Omega = 12\ \text{k}\Omega$ would be used.

Often electricians or electronics technicians make use of a combined meter called a **multimeter**. (See photo.) This instrument is a multi-scaled device that is usually capable of measuring resistance, voltage and current in both DC and AC modes. In recent times, the output display is a digital display rather than the analog needle movement type of voltmeter and ammeter described in this section. The use of a multimeter has more importance in electronics and will be further discussed in Chapter 23. Electricians who need to test the effective resistance of insulation around conductors to check safety requirements use an instrument called a 'megohm tester', which is able to measure the resistance at a particular high voltage.

Photo 22.6
Meter scales.

(a) (b) (c) (d) (e) (f)



Activity 22.5

SR

A Read the meter scales

Use the set of photographs (Photo 22.6) showing various electrical instrument scales. Read the measured quantities accurately. List what each instrument is, and its scale reading in correct units.

EI

B Battery discharge project

You are required to test and complete an experimental report on the discharge characteristics of several different types of batteries. To do this you will need to design and set up a discharge circuit that allows the data-logging of the terminal voltage under load of the various battery types. You could, for example, compare such types as normal zinc–carbon batteries with Energizer alkaline MAX, Energizer e2-Titanium, Nickel–cadmium rechargeable and lithium metal hydride batteries. The final choice is up to you.

The test rig should be able to allow the connection and slow discharge of the batteries through a normal light bulb circuit and you should program the CBL2 data-logger to take voltage sensor readings as the battery discharges through the light bulb circuit. Your experimental design should include all circuit diagrams and constructional methods. You should display all final results in graphical form for easy comparison by using the features of the data-logger.

If you have both voltage and current sensors for the CBL2 or an equivalent data-logger, then you may be able to directly compare voltage and current discharge characteristics for your battery set. You could also compare your findings with technical data that is available on the Internet sites of companies such as Eveready, Energizer, Panasonic or TDK.

Circuit symbols

We have been using several common electric circuit symbols already. Electric circuit diagrams are the standard method of representing actual circuits in practice. Some standard electric symbols used are shown in Figure 22.19. More will be met in the next chapter. It should be noted that a rectangular style is used when drawing electric circuit diagrams. This is for ease of reading the connections between various components but, in practice, the actual working circuit may not follow this rectangular style, especially if forming part of a printed circuit board in a consumer electronic device such as a television set or computer. Finally in this section, we will analyse a more complex electric circuit, making use of all circuit laws, electric meters and methods of connection discussed so far. This process is very important as you must be able to read unfamiliar electric circuits and carry out the necessary calculations to solve for unknown or required circuit component values. You must also be able to use the laws of circuit behaviour to predict voltages and currents at any point in a circuit. The following example illustrates the general steps that might be followed, but remember that there is usually more than one way to the solution.

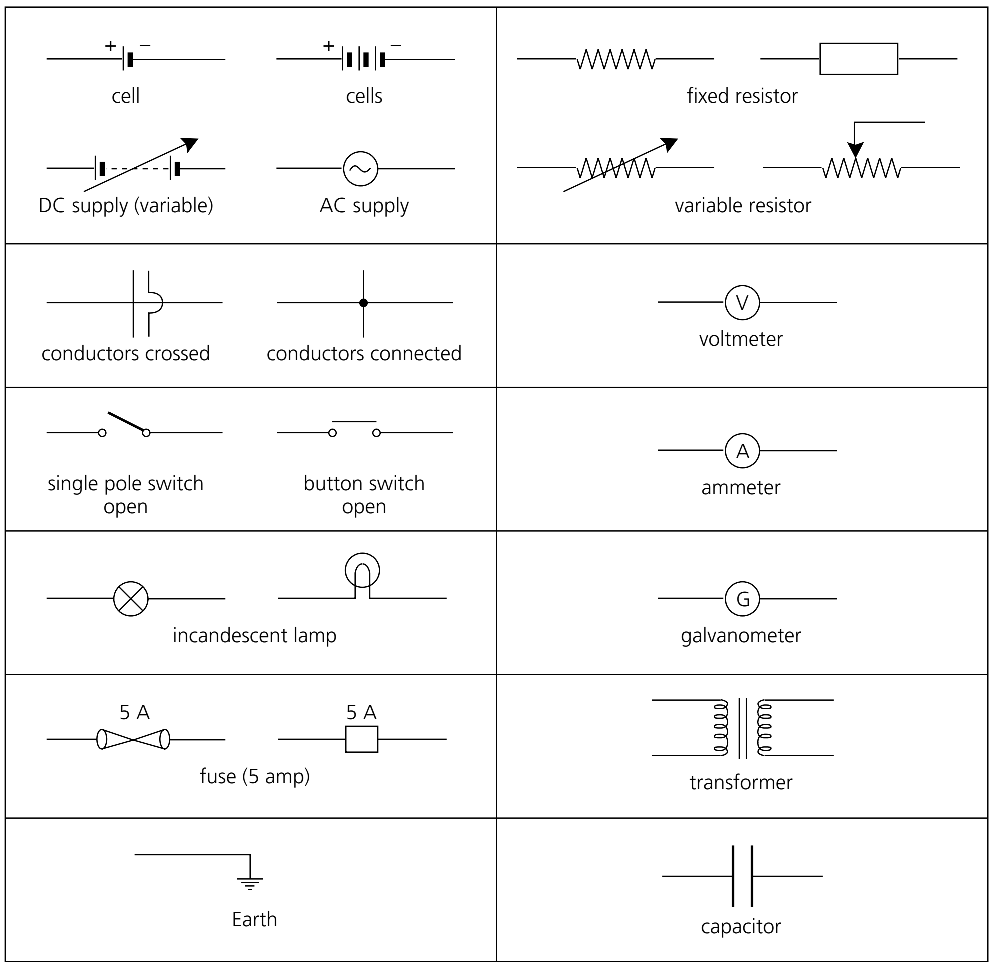


Figure 22.19
Circuit component symbols.

Ch22-Walding 4th 25/8/04 12:27 PM Page 486

Example

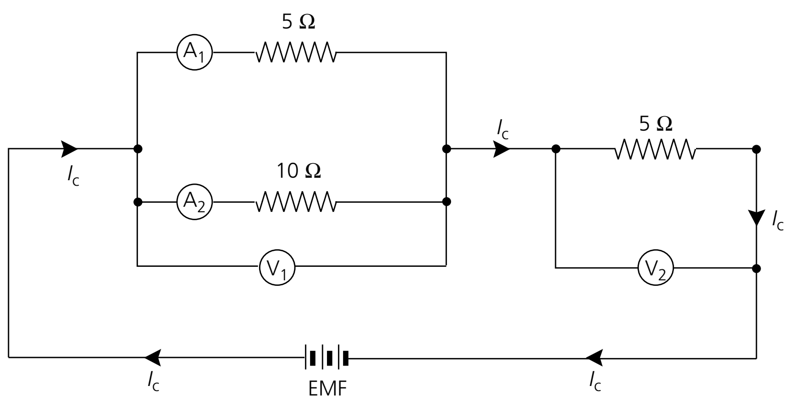


Figure 22.20

Consider the circuit shown in Figure 22.20. Calculate the circuit current, I_c, flowing from the battery, as well as the current reading on ammeters A_1 and A_2 and the voltage reading on V_1 and V_2. What is the battery EMF if it contains three cells each of 1.5 V?

Solution

Note the following points about this circuit:

- Battery has three cells each of 1.5 V, therefore EMF = 4.5 V.
- 10 Ω and 5 Ω resistors are in parallel and this combination is in series with the 5 Ω resistor.
- The current readings $A_1 + A_2$ will equal I_c because of Kirchhoff's junction rule.
- The voltages across the 5 Ω and 10 Ω resistors will both be equal to voltage V_1.
- The sum of voltages $V_1 + V_2$ will equal the EMF, 4.5 V, due to Kirchhoff's loop law.

Step 1 Calculate equivalent resistance of parallel combination:

Use $$\frac{1}{R_P} = \frac{1}{5} + \frac{1}{10} \Rightarrow \frac{1}{R_P} = \frac{3}{10} \Rightarrow R_P = 3.3\ \Omega$$

Step 2 Calculate equivalent circuit total resistance in series with battery:

$$R_{tot} = R_P + 5 = 3.3 + 5 = 8.3\ \Omega$$

Step 3 Calculate total current flow, I_c, using Ohm's law:

$$I_c = \frac{V}{R_{tot}} = \frac{EMF}{R_{tot}} = \frac{4.5}{8.3} = 0.54\ A$$

Step 4 Now consider only the 5 Ω resistor. Apply Ohm's law to find the voltage V_2:

$$V_2 = I_c \times R = 0.54 \times 5 = 2.7\ V$$

Step 5 Calculate the voltage V_1, using the loop law:

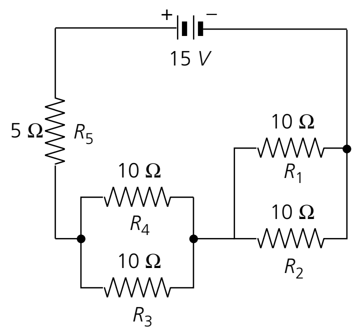
$$V_1 + V_2 = EMF \Rightarrow V_1 + 2.7 = 4.5 \Rightarrow V_1 = 1.8 \text{ volts}$$

Step 6 Voltage V_1 = 1.8 V is the voltage drop across each resistor 5 Ω and 10 Ω in the parallel arm. Hence, calculate currents A_1 and A_2:

$$\begin{array}{lll} V_1 = I_1 \times 5 & 1.8 = I_1 \times 5 & I_1 = 0.36\ A \\ V_1 = I_2 \times 10 & 1.8 = I_2 \times 10 & I_2 = 0.18\ A \end{array}$$

Notice finally that the sum of the currents I_1 and I_2 equals the circuit current $I_c = 0.54$ A as required by Kirchhoff's law. It is also possible to redraw equivalent but simplified circuit diagrams at each step to further aid understanding of the analysis.

Figure 22.21
For question 12.



Questions

11 A student set up an electric circuit with two 25 Ω resistors in parallel, connected to a battery of EMF 12 V. She wishes to calculate total circuit current and the individual currents through each resistor. Draw a circuit diagram she would use and calculate values.

12 Consider the electric circuit shown in Figure 22.21. Calculate the current flowing through each resistor and the voltage drop across each resistor using the laws of circuit analysis. Fully describe your steps and redraw the appropriate equivalent circuits at each step.

ELECTRICAL ENERGY AND POWER 22.6

The most important aspect of electrical energy use in modern society is the ease with which this form of energy can be converted into a whole range of other energy forms, such as heat, light, mechanical energy, electromagnetic energy — radio, television. Electrical energy is generated in several ways, at a simple DC level with devices such as cells and batteries, through to AC generators of different types. Domestic and industrial AC electricity supply is generated often by coal, oil or gas burning power stations or even by thermal, wind power or hydroelectric power stations. The basis of all forms of AC electric generators is the spinning coil induction turbine. Once the electrical energy is produced, AC transformers can change the voltage so that it may be efficiently distributed via conducting cables around the countryside to factories and homes. These devices, techniques and issues will be discussed in detail in Chapter 26.

NOVEL CHALLENGE

A writer in *New Scientist* magazine (July 1998) described how in 1938 he stayed in a country house in England and helped wind up a 1 tonne steel ball suspended on a chain into the roof space. During the evening, the ball was allowed to fall slowly, turning a generator to keep a light glowing all night. He said that this was impossible as there was not enough gravitational potential energy in the ball to do this. Verify his claim by working out how long a 60 W lamp would glow if the ball was raised 5.0 m. Assume the energy conversion was 100% efficient (unlikely!). In actual fact it turns out that the steel ball did not turn a generator but turned an enclosed 44 gallon drum partly submerged in petrol. As the drum turned, petrol evaporated and was burnt in a gas lantern.

NEI Activity 22.6 BIG GENERATORS

1. Use library research or send away to your local electricity authority to find out about the types of power stations that produce electricity for your school and home. If you have a solar or wind-powered electric generator at your house or school, find out how it works and write a short descriptive report.
2. Research from an encyclopaedia, Internet searching or the *Guinness Book of Records* the highest voltage ever produced and where it was accomplished.
3. Try to find out from similar sources the location of the largest electric generator operating in the world and what its output is.

When an electric current flows through a resistor, thermal heat is produced, as was discussed in Section 22.3. Electrical energy is being converted to thermal energy within the resistor and this forms the basis of any electric appliance designed to produce heat, such as radiators, electric stove elements, hot water systems, electric blankets and electric kettles. In an electric light bulb, this resistive heating of the filament wire even begins to produce light energy. Electrical energy is often converted into mechanical energy; for example, in any appliance that contains an electric motor. Whenever electrical energy conversion is occurring, electric charge, Q, is being moved through a potential difference, V. This requires an electric force doing work, W, given by $W = Q \times V$.

The rate of energy transfer or the rate at which electrical work is done is called **electrical power**. Thus:

$$P = \frac{W}{t} = \frac{Q \times V}{t} = V \times I$$

Power is the product of the potential drop or voltage across an appliance times the current flow through the appliance. This formula is appropriate in both DC and AC voltage and current situations. The unit for electrical power is the **watt** (W) with 1 watt being equivalent to a rate of energy transfer of 1 joule per second: 1 W = 1 J s^{-1}.

Using Ohm's law, it can be readily seen that alternative forms of the power formula can be derived, namely:

$$P = V \times I = I^2 \times R = \frac{V^2}{R}$$

Many domestic and industrial electric appliances state the **power rating** on their compliance plates. For example, a television set that is rated at 110 W will consume electrical energy at the rate of 110 joules per second. It is this energy usage that consumers have to pay for as it is supplied by the electricity authority. (Refer to Section 22.7.)

Example

(a) Calculate the power dissipated by an electric drill operating from the normal 240 V AC supply and drawing an operating current of 1.6 A.

(b) Calculate the monthly energy used by a television set whose power rating is 110 W and is operating daily for 6.5 hours.

Solution

(a) Use $P = V \times I$ = 1.6 A × 240 V = 384 W.

(b) Energy used daily:

$$W = P \times t = 110 \times 6.5 \times 3600 = 2.6 \times 10^6 \text{ J}$$

But if the set is used for one month of, say, 30 days, then total energy used = 7.7×10^7 J.

The common unit for electrical energy usage in domestic and industrial situations is the **kilowatt-hour** (kW h). The electricity authority commonly refers to the kW h as a 'unit' of electricity and it represents the amount of electrical energy used by a device rated at one kilowatt over a period of one hour.

Questions

13 Calculate the power rating of a light bulb operating at 240 V and 0.6 A.

14 What is the resistance at normal operating conditions of the following appliances run from the 240 V AC mains: **(a)** 50 W television set; **(b)** 1 kW hair dryer; **(c)** 100 W light bulb?

15 A heater is connected to the normal mains supply. If the resistance element has a value of 8 ohms, how much electrical energy is supplied to the heater in 5 minutes?

16 It requires 4.2 kJ of energy to raise the temperature of 1.0 kg of water by 1°C. If an electric hot water system is rated at 6.5 kW for 240 V AC and it holds 250 kg of water, calculate **(a)** its coil resistance while operating; **(b)** the energy required, in kW h, to heat the water from 15°C to 80°C.

NOVEL CHALLENGE

The formula P = V^2/R implies that if the resistance (R) of a light bulb is decreased, the power consumption (P) will increase and hence it will glow brighter. But the formula $P = I^2 R$ implies that if R is decreased, P is also decreased and the bulb will get dimmer. They can't both be right. What is your answer to this apparent anomaly?

HOUSEHOLD ELECTRICITY AND ELECTRICAL SAFETY

22.7

NOVEL CHALLENGE

One that most people get wrong: 100 W bulbs glow brighter than 40 W bulbs when connected in parallel across a 240 V source. *If these two bulbs were connected in series, how would their brightness compare?*

NEI

Activity 22.7 YOUR METER BOX

Look into the meter box at your house, and (without touching anything!) determine:

1. how many meters it contains;
2. how many fuses, switches or circuit-breakers appear to be used;
3. the descriptions associated with each fuse or switch;
4. which way the spinning discs inside each of the largest meters turn.

Electrical energy produced by power stations is AC or alternating current. The current changes direction in any household AC circuit at a frequency of 50 Hz and in Australia is provided at a voltage of 240 V. (Refer back to Figure 22.6**(c)**.) This 240 V is converted at local street pole transformers from much higher AC voltages on the main high-tension (high-voltage) distribution grid. Our domestic 240 V AC electricity is potentially very dangerous and significant precautions must be taken in any household installation in order to protect consumers from faults that may occur with appliances.

The electrical cabling involved in household electricity usually contains three colour-coded insulated conductor wires. The first is the brown or red covered **active wire**. This is often called the live wire and carries the current to the appliance when it operates. The electric potential of the active wire varies between positive and negative, and is particularly dangerous as it would produce a fatal shock if touched because the potential (voltage) involved would force current to flow through your body to earth. The second wire is covered in blue or black insulation and is called the **neutral wire**. The neutral wire allows return current to flow from the appliance in operation. This neutral wire is at earth potential (zero volts) because it is earthed at the local electricity supply sub-station, and is thus far less potentially dangerous than the active wire. The third wire is covered in green and yellow, or just green, insulation and is called the **earth wire**. This is a safety connection made to a thick metal stake entering the ground at some point around your house.

NOVEL CHALLENGE

In an experiment to measure the efficiency of a microwave oven, 1 L (1000 g) of water was placed in an icecream bucket and its temperature measured. It was then cooked on 'high' for 2 minutes and its temperature measured again. Use the formulas $Q = mc\Delta T$ and $P = W/t$ to prove that the power output of the microwave $(P) = 34.8 \times \Delta T$. *Hint:* let $Q = W$. *If a 750 W microwave raised the temperature of 1 L of water by 20°C, calculate the efficiency of the oven.*

Activity 22.8 YOUR EARTH SYSTEM

Locate and describe the position and type of the main earthing point in the electrical system wiring at your house. Is it located under cover or out in the open? Why is it located at this position? In the past a mains water pipe has often been used as the main earthing point. This is less common today. Explain why this might be the case.

In most household situations both the earth and neutral cables are connected together at the fuse-box to form what is known as the **multiple earth neutral** (MEN) system. If an appliance has a metallic outer surface then it is connected internally to the earth wire. This is because, if the active wire insulation breaks down or an internal fault occurs, causing the outer metallic case to become live at 240 V potential, the current will flow through the earth wire and not through the higher resistance of someone touching it. Earthed appliances are generally much safer for this reason. Electrical power points in the walls are similarly connected with a three-pin socket for active, neutral and earth (Figure 22.22). The electrical switch must be placed in the active line, in order to turn an appliance on or off at the wall socket.

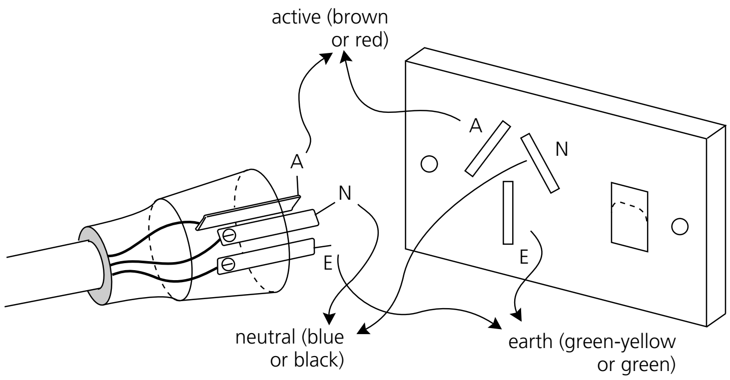


Figure 22.22
Electric plug and socket.

240 VOLT COLOUR STANDARDS	FLEX (POWER LEADS)	WALL CABLE
Active	brown	red
Neutral	blue	black
Earth	green-yellow	green-yellow

You may have noticed that some appliances have a power lead that contains only figure-of-eight twin conductor wire. This type of appliance has a **double insulated** rating, which means that any internal metal parts are not only insulated with the normal protective wire insulation but also the outer case of the appliance is plastic and cannot become live in the event of a fault condition or primary internal insulation breaking down. Compare the diagrams in Figure 22.23, which also shows the commonly used symbol for a double insulated appliance — a concentric pair of squares.

NOVEL CHALLENGE

The Australian flat-pin plug (see Figure 22.22) is similar to the plug used in mainland China, but dimensional differences prevent the Chinese plugs from being used in Australia. Argentina uses the same plug as Australia but the Active and Neutral are reversed, and the plugs are banned from use in Australia. What is the problem with having A and N reversed?

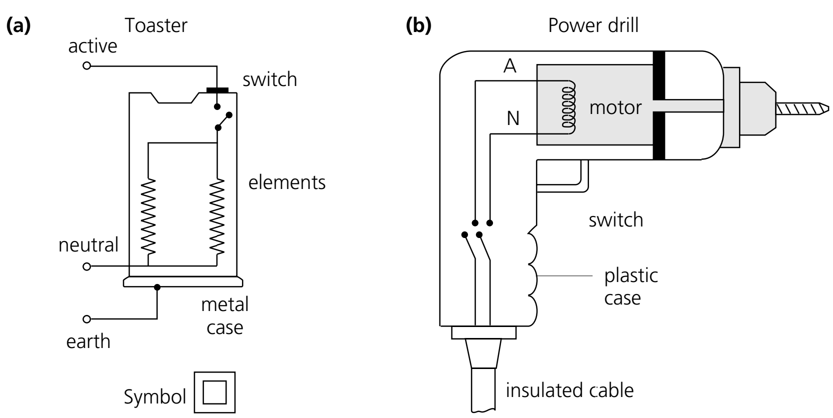


Figure 22.23
Earthing **(a)** and a double insulated appliance **(b)**.

If a fault occurs in the electrical wiring of a house, or if an appliance becomes faulty, a short circuit, which allows a high current to flow very easily, may be produced. This subsequently causes rapid heating of the conductors. This heating effect would be high enough to cause melting of insulation and a fire if it were not for fuses placed in the active lines of the household circuits. A **fuse** is a small piece of resistive wire alloy that is designed to melt and

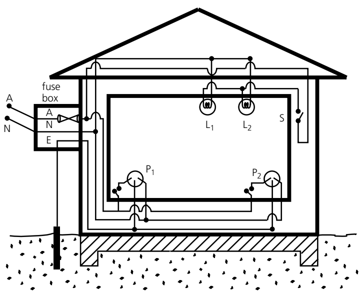
Ch22-Walding 4th 25/8/04 12:27 PM Page 491

break with excessive overheating or at a particular current rating such as 8 or 16 amps. Once the fuse breaks a completed circuit is no longer present and any further current is stopped. Once the fault condition is diagnosed by an electrician or the faulty appliance repaired, the broken fuse can be replaced and the circuit is again complete. Fuses are also important because they prevent a circuit being loaded with too many appliances in parallel and thus causing excessive currents to be drawn.

Photo 22.7
Household meter box.

A modern replacement for the fuse is the **electromagnetic circuit-breaker** (see Photo 22.7), which is a small electromagnetic solenoid switch that will also break the circuit if a fault current develops. The advantage is that it can be quickly manually reset. Figure 22.24 gives some idea of how a typical household room might be wired with two power points and two lights with switches. Notice that the room lights are not usually earthed and that both the power points and the lights are in paralleled connections.

Figure 22.24
Simplified household AC wiring.



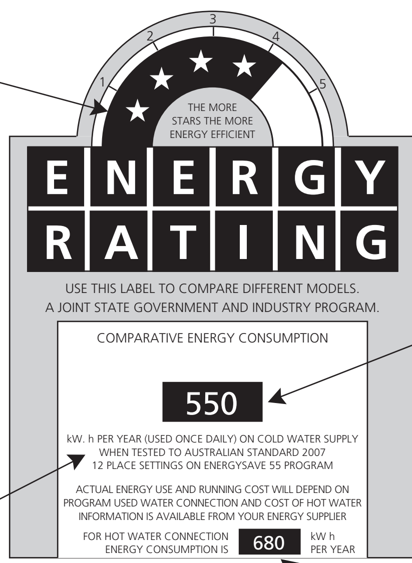
One of the most successful techniques for maintaining electrical safety in the household is to install an **earth leakage circuit-breaker** (ELCB), also called a **residual current device** (RCD) (see Photo 22.7). This device is usually permanently installed in the meter–fuse-box of the house. It will electronically sense the very small differences in the electric current balance between incoming active line and outgoing neutral line that will occur in any wiring or appliance fault condition that allows current flow to earth. These devices are available in different trip current ratings (the most common tripping at 30 mA) and are very sensitive and fast-acting (10–20 ms), so that electrical safety is maintained. In some situations around the house even portable RCD units are becoming popular.

Household appliances should always be chosen and used on the basis of their electrical operating efficiency. Many larger appliances in homes and industry, such as refrigerators, washing machines, dryers, dishwashers and range stoves, carry an **Australian energy rating** system label, which contains vital testing information relating to a national standard for electrical efficiency. See Figure 22.25 for an example of this label from a dishwasher.

Figure 22.25
The dishwasher energy rating label explained.

How do I choose the most energy efficient dishwasher?
• The simplest way to compare the energy efficiency of dishwashers is look for the stars. The more stars on the Energy Rating Label the more energy efficient the dishwasher. More stars means the dishwasher uses less electricity to achieve the same level of performance.

How does a dishwasher get an Energy Rating?
• To determine the Energy Rating manufacturers must have their appliances tested to an Australian Standard. The program setting used in the tests is stated on the label and this program will satisfy the needs of an average wash.



What do the figures on the label mean?
• The dishwasher Energy Rating Label shows two energy consumption figures in red boxes. They will tell you how much energy the dishwasher will use per year, if it is operated once a day.

• The figure in the large box is based on the manufacturer's recommended water connection.

• The figure in the small box is based on an alternative water connection. (Usually hot water only or a dual hot/cold connection). Here the operating costs will vary depending on your hot water system.

Appliances usually fall into two broad categories: the high power devices that mostly generate heat and the low power devices that include lighting and consumer electronic items such as radios, TVs and computers. Table 22.3 lists the electrical ratings of some appliances, showing the typical operating range.

Table 22.3 ELECTRICAL POWER AND ENERGY CONSUMPTION

APPLIANCE	TYPICAL POWER (W)		AVERAGE USE (h day^{-1})		ENERGY USE (AVE) (W h day^{-1})	
	MIN.	MAX.	MIN.	MAX.	MIN.	MAX.
Kitchen:						
• Lights	11	100	1.00	3.00	11	300
• Refrigerator	100	260	6.00	12.00	600	3 120
• Microwave	650	1 200	0.17	0.25	111	300
• Toaster	600	600	0.03	0.08	18	48
Laundry:						
• Lights	11	100	0.25	1.00	3	100
• Iron	500	1 000	0.17	0.42	85	420
• Washing machine	500	900	0.22	0.33	110	300
• Clothes dryer	1 800	2 400	0.20	0.54	360	1 300
• Water pumps	300	500	0.25	1.00	75	500
Lounge:						
• Lights	15	100	1.00	4.00	15	400
• Television	25	200	0.50	5.00	13	1 000
• Video recorder	100	100	0.5	5.0	50	500
• Stereo	60	80	0.5	3.00	30	240
• Vacuum cleaner	100	1 000	0.13	0.25	13	250
• Oil radiator heater	1 000	3 000	8.00	14.0	8 000	42 000
• Strip heater	500	1500	0.5	1.00	250	1 500
Bedroom:						
• Lights	11	100	0.5	2.00	6	200
• Radio	10	40	0.33	3.00	3	120
Garage:						
• Lights	11	100	0.17	2.00	2	200
• Power tools	200	800	0.17	0.17	34	136
• Hot water (storage)	3 000	20 000	5.00	8.00	10 000	24 000

Example

If the tariff cost from the local electricity supplier for domestic light and power is 9.18 cents per kilowatt-hour, use the information in Table 22.3 to calculate the total monthly cost of operating a television set of 120 W at maximum average daily use.

Solution

From the table, average daily use in 5 hours, at power rating P, = 120 W for the set:

$$\text{Energy} = P \times t = 5 \times 120 = 600 \text{ W h per day}$$

For one month of 30 days = 18 000 W h = 18 kW h, at cost of 9.18c = $1.65 total.

NOVEL CHALLENGE

Have you seen the new compact fluorescent bulbs designed to replace the normal incandescent ones? They cost about $20 but last 8000 hours and produce light equivalent to a 75 W incandescent bulb but only consume 20 W of power. Incandescent bulbs cost 75 c and last about 1200 hours. Which is better value?

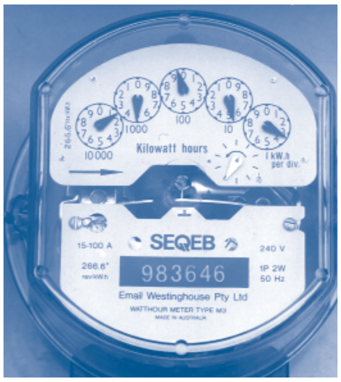
NOVEL CHALLENGE

Consider the statement: 'Electricity from a battery is 100 times more expensive than the electricity from your power point'. Is this true? Design an experiment to find out how much electricity you can get out of a 1.5 V AA battery and compare it to the 240 V mains supply (10 cents per 'unit').

NOVEL CHALLENGE

Incandescent light bulbs have a life proportional to $1/V^{14}$ seconds (where V is the applied voltage). Hence, if you run a 240 V bulb at 80% of its rated voltage you will increase its lifetime. By how many times will its life be increased?
We think 23 times; but how did we get that?

Photo 22.8
Kilowatt hour meter ENERGEX.



Activity 22.9 APPLIANCES

1 Try to find out if the larger appliances in your house have an energy rating label. From it calculate the average cost of operating them as normal for one year.

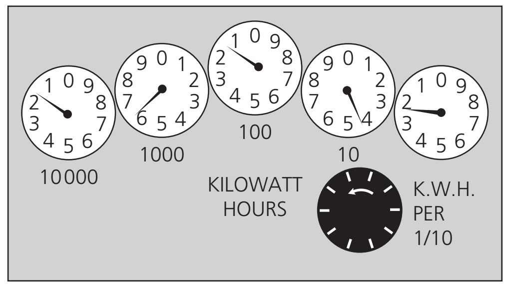
2 Figure 22.26 gives a diagram of a typical set of dials on the meter box electrical kilowatt–hour meter. Refer also to Photo 22.8. Can you work out how to read the dials? Try reading your own household meters and keep a weekly record of the power consumption in order to compare the readings with your electricity bill.

3 RCD- or ELCB-type devices only protect against certain types of electrical faults. How do these devices actually work and what specific types of hazards do they protect us against?

Figure 22.26
How to read a meter.

How to read a meter
Some customers like to check their meter readings from time to time.

- The dial type meter is the most common type installed. Stand directly in front of the meter so that you can see the exact position of the pointer. Start at the right-hand dial and record the number the pointer has just passed on each dial.
- If you wish to check your average daily consumption, take readings at the same time of day, several days apart and divide the difference in readings by the number of days.



The reading from the dials above is 16142

Also, remember that meters belong to ENERGEX (or other authorities in other States). Interfering with them is illegal and staff are trained to spot any evidence of tampering.

Electric shock

Finally in this chapter we will take a look at some of the effects of DC and AC electricity on the human body, including **electrocution**. The severity of an **electric shock**, that is, bringing the body into contact with an EMF source, depends on the current flow, duration, frequency, skin moisture, surface area of contact, pressure exerted, temperature and the path taken by current through the body. A current passing through vital organs such as the brain or heart is the most dangerous. The biological effects of electricity result from both the DC electrical resistance of the body and the AC electrical impedance (frequency-dependent resistance).

NOVEL CHALLENGE

A 240 V jug rated at 2000 W takes 1 min 45 s to heat 2 cups (500 g) of water from 23°C to boiling (100°C). Calculate its percentage efficiency. If 1 L of water was being boiled from the same temperature, propose whether the time would be exactly twice or more, or less, than twice. Hmmm — think of the energy losses. Check our web page for some real data!

Table 22.4 EFFECTS ON THE HUMAN BODY OF 240 V, 50 Hz AC FOR 0.5 s

CURRENT (mA)	EFFECT ON THE BODY
0.5	threshold of perception
1.0	able to be felt; tingling sensation
4.0	pain felt; rarely causes damage
10.0	threshold of 'let-go'; just able to release
20.0	muscles paralysed; unable to release
50.0	severe electric shock; burns; ventricular fibrillation threshold
150.0	breathing difficult; major damage
200.0	death likely
500.0	serious burning; breathing stops; death inevitable

Voltages as low as 32 V AC or 115 V DC can be dangerous. Table 22.4 lists several identifiable levels of electric shock. In general, it is true that an AC voltage is more dangerous than an equal DC voltage because it will trigger stronger muscular contractions. Fortunately the human skin is a fairly good insulator, which provides a barrier against dangerous electric currents. The effective resistance between two points on opposite sides of the body, when the skin is dry, is in the range 10 000 to one million ohms; however, if the skin is wet, the resistance may be less than 1000 ohms. For voltages greater than about 50 V, the human skin begins to break down as an effective barrier and the body's internal resistance becomes more important in determining the current flow through the body.

INVESTIGATING

During times of heavy power demand, the voltage of the electricity from the power station drops by up to 2%. *Is this to save money, to save power or for some other reason?*

Table 22.5 AVERAGE TOTAL BODY RESISTANCE FOR 95% OF POPULATION

TOUCH VOLTAGE (V)	AC RESISTANCE (Ω)
25	6100
50	4400
75	3500
100	3200
110	3000
240	2100
500	1600
1000	1500

Table 22.5 lists the total average body resistance for 95% of the population at various touch voltages. From these tables it can easily be seen that at 240 V the body current typically is about 100 mA and, depending on the contact time, can be fatal most of the time. The most dangerous path for current is from one limb to another, across the chest, as this is most likely to affect the heart. Exposure to electric shock, especially via this pathway, can bring about **cardiac fibrillation**, or rapid and uncontrolled beating of the heart, which can starve the brain of oxygen, quite quickly causing permanent damage or even death.

If fibrillation of the heart begins, ambulance officers at the scene of an electrical accident will begin a procedure known as **defibrillation**, which involves usually two steps. Firstly, the rapid fibrillation must be stopped. This is done by placing the plates of a defibrillator on the chest on either side of the heart. A 5 kV DC pulse lasting about 1–50 ms passes through the heart causing it to stop temporarily. The defibrillator recovers within 2–3 s and is ready then to deliver a second pulse. The second step involves a second pulse identical to the first, applied again through the heart, if it hasn't restarted its normal sinus rhythm naturally. If this is unsuccessful, a third, even stronger, pulse is applied. In the event that a defibrillator is not available, cardiopulmonary resuscitation (CPR) should be administered until hospitalisation.

SR ACTIVITY 22.10 CARDIAC DEFIBRILLATION

You are required to examine the following information, provided on the Biomedical Electronics pages of the website of the Australasian Society of Cardio-Vascular Perfusionists Inc. Read and interpret the material carefully, also using assistance from Chapters 21 and 22, and complete the questions that follow.

Principles of operation of DC defibrillator

A Energy used for cardioversion and defibrillation

Electrical output of defibrillators is expressed in terms of energy. Joules (J) or watt-seconds (W s) describe the power (watts) and the length of time for which it is applied (s).

Thus energy (joules) = power (watts) × duration (seconds)

Note: watt = current (amps) × voltage (volts)

1 watt = 1 J s^{-1}

Defibrillators are set according to the amount of energy stored; this depends on both the charge and the potential. Capacitance is the measure of the ability of an object to hold an electrical charge; SI unit is coulomb (C).

$$\text{coulombs} = \text{amperes (A) x seconds (s)}$$

$$\text{potential (V)} = \frac{\text{power (W)}}{\text{current (A)}}$$

$$\text{power } (W) = \text{energy (J) per second}$$

$$\text{current (A)} = \text{charge (C) per second}$$

$$V = \frac{Js^{-1}}{Cs^{-1}} = \frac{J}{C}$$

$$J = C.V$$

$$\text{Stored energy } (J) = \frac{1}{2} \times \text{capacitance (C)} \times \text{potential (V)} = \frac{1}{2} \times C \times V^2$$

Example: with paddles potential of 5000 V applied across two plates of a capacitor, produces a store of electrons of 160 mC of charge.

B Operation and circuit

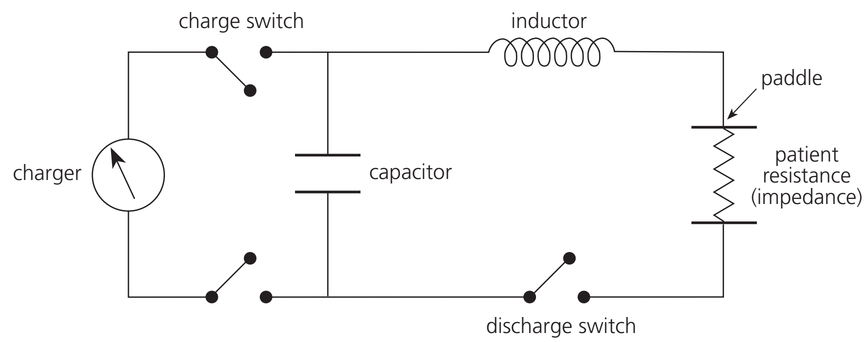
Defibrillation energy is temporarily stored in a capacitor. The large capacitor is charged to the selected energy and then discharged through the paddles applied to the chest. The energy stored in the capacitor is released as a current pulse (e.g. 35 A for 3 ms) resulting in a synchronous contraction of the heart after which a refractory period and normal beats may follow.

An inductor is included in circuit to ensure that the electric pulse has an optimal shape and duration.

During discharge, the inductor absorbs some of the energy so that not all is discharged to the patient.

Defibrillators are calibrated in terms of delivered energy, not stored energy.

Figure 22.27
For Activity 22.10B.



C DC defibrillator pulse shapes (waveforms)

The defibrillation waveform is a major factor in determining efficacy of defibrillation.

A damped sine wave defibrillator consists of a capacitor, inductor and electrodes.

Placing of an inductance coil in series with the capacitor, the resultant waveform is half sinusoidal in configuration; a slight variation is the underdamped sine wave, in which the sine wave reverses slightly, and which may reduce the defibrillation threshold.

The duration of the current for adult is 5–10 ms. Current intensity depends on the set stored energy on the defibrillator.

Patient impedance, or the resistance to current that is offered by the chest, is called chest impedance; an average figure is 75 Ω.

Variations in the patient's impedance cause the delivered dose of current to vary widely.

Factors that influence impedance between the defibrillator paddles (resistance) are:

- delivered energy
- paddle (electrode) size and composition
- interface between paddle and skin (gel used to reduce this)
- paddle pressure (increased pressure decreases impedance)
- time interval between discharges
- number of discharges (increased number decreases impedance)
- phase of patient ventilation
- distance between paddles.

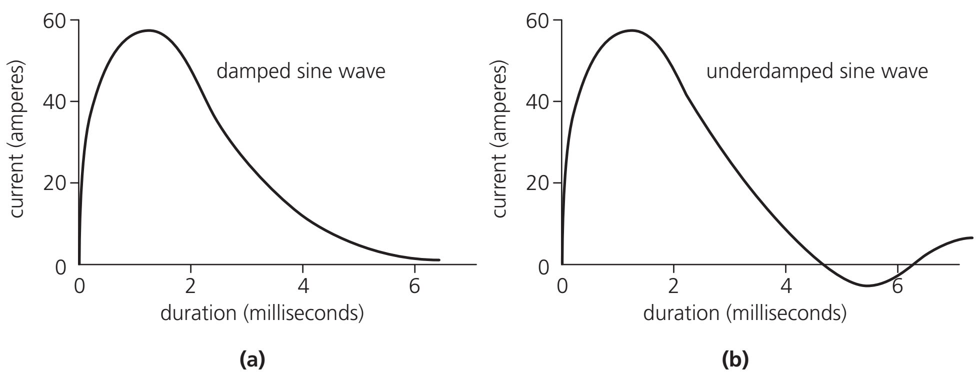


Figure 22.28
For Activity 22.10C.

D Energy levels required for internal and external defibrillation

1 External — transthoracic defibrillation

(a) Children (weighing 2.5–50 kg) = 1–5 J/kg
(b) Adults (body weight in adults does not seem to be a major factor determining energy requirements)

Initial setting 200 J followed rapidly by 300 J and 360 J if needed; animal studies suggest that these doses are also valid during hypothermia.

2 Internal — direct open chest defibrillation
Adults: initial setting 5 J with increments up to 20 J.

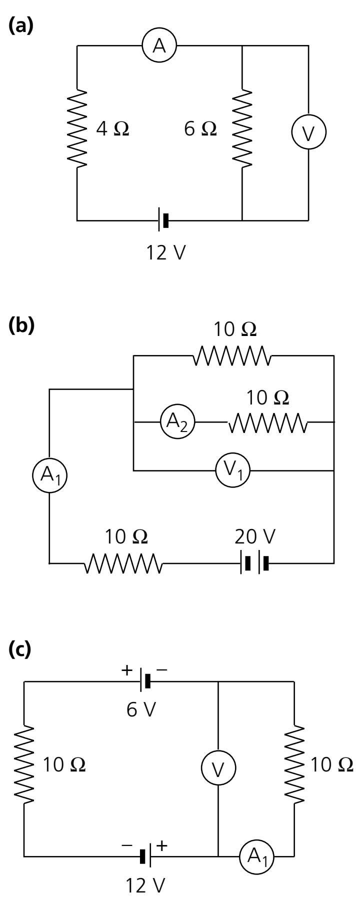
Questions

1 Calculate the energy (in joules) delivered by the paddles in the example in section A above.
2 Why does the inductor in the circuit change the shape of the waveform?
3 What are the factors that cause current delivery to the patient to change?
4 Why do you think body weight in adults is not so important as in children?
5 Why is the energy for external defibrillation much higher than for internal defibrillation?

Australian farmers make use of electric fences. A typical **electric fence** produces 7500 V DC pulses, lasting 0.2 ms at intervals of about one second. These voltage spikes are usually produced from a 12 V battery. Electricity authorities specify a 10 000 V maximum and the unit must be able to deliver at least 5000 V under a typical load of 500 Ω. A good electric fence unit connected to a clean fence should be able to maintain 7500 V over 20 km of fencing. Farmers check the fence voltage with a voltmeter and can usually tell from specified points, such as open gates, if the system is working properly. Animals that occasionally get caught in the fence will die of stress, not from the electric shock given. Usually, after the first shock onto a cold wet nose, the cattle or other stock learn very quickly not to approach the electrified fence.

Probably the most dramatic effect of electricity on the body is the use of an **electric chair** in the American criminal justice system. It was in 1890, in New York, that William Kemmler became the first criminal in history to be put to death by electrocution. Apart from a period in the 1970s, several hundred people a year have been executed in the 'chair', between 1890 and the present in the USA. Traditionally the prisoner is securely fastened to a solid chair by straps holding the chest, groin, arms and legs. The electrodes are moistened copper plates attached to each calf and a band around the head. Jolts of 4–8 A at between 500 and 2000 V AC are applied for a half-minute at a time. A doctor inspects the body to see if death has occurred or if another jolt should be administered. At Sing Sing prison in New York, an initial voltage of 2200 V at 7–12 A is used at half-minute intervals over a period of two minutes. Current flow in each leg and the head is monitored. Body temperatures rise to above 50°C. The use of the electric chair has dropped in recent years as lethal injection is being adopted by more US States. The use of capital punishment in all forms is currently banned in Australia.

Figure 22.29
For question 23.



It is important to realise the biological effects of even small voltages on our bodies. Electrical safety must always be utmost on our minds when we work at home or in the laboratory. Electrical energy, while vitally important in modern society, deserves our utmost respect. It is far too dangerous for anyone to be complacent about its potential.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***17** What is the electric current in the following cases?
- **(a)** A 20 C charge passes an ammeter in 5 s.
- **(b)** A 5 C charge passes an ammeter in 20 s.
- **(c)** A 200 C charge passes an ammeter in 3 minutes.

***18** What is the energy gained by a charge of 18 C when it passes through a source of EMF of 12 V?

***19** If a 1.5 m length of resistance wire has a resistance of 1.9 Ω and the wire has a diameter of 0.9 mm, calculate the resistivity of the wire.

***20** With the aid of a circuit diagram, show how you might measure the operating resistance of a single light bulb. You have a battery, ammeter and voltmeter.

***21** The accumulator of a car produces 12 V. If the car lights at the sides and rear are each rated for 12 V, but the two interior lights are only rated at 6 V, how should the lights be connected — in series or in parallel?

***22** How does the resistance of a 60 W, 240 V light bulb compare with that of a 25 W, 240 V light bulb? Which has the thicker bulb filament?

***23** In each circuit of Figure 22.29, find the readings on all voltmeters and ammeters, as well as the total circuit equivalent resistance.

***24** Explain why birds can safely touch overhead power cables but humans standing on the ground cannot.

**25 In the circuit of Figure 22.30, calculate the following:
(a) The battery voltage.
(b) The circuit current at the points labelled X, Y and Z.
(c) The reading on the voltmeter, V.

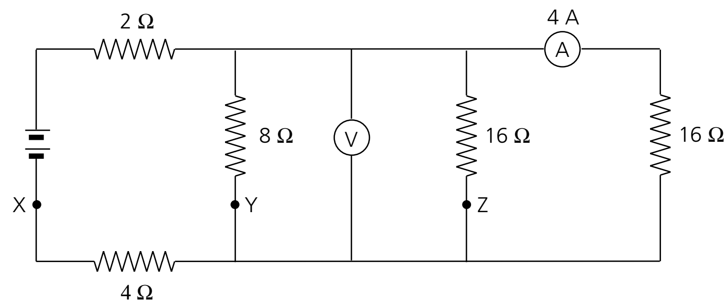


Figure 22.30
For question 25.

26 An automatic washing machine is labelled 240 V, 960 W. Calculate **(a) the operating current in normal use; **(b)** the operating resistance in normal use.

**27 A set of decorative Christmas lights consists of bulbs labelled 12 V, 1 W. If the set is designed to operate from the AC mains, determine:
(a) how many bulbs are in the set;
(b) what the total power consumed by the set of lights is;
(c) what the average voltage drop across each bulb is;
(d) what the current in each bulb is and the resistance of each.

*28 Explain the physical and operational differences between a fuse and a circuit-breaker as safety devices in household circuits.

*29 Explain the precautions that are required if any metal-framed appliance is to be connected to the household mains supply of electricity. Complete a circuit diagram.

**30 Figure 22.31 shows a graph of potential difference versus current for two different electrical devices A and B.
(a) Which device is an ohmic resistor?
(b) If A and B are connected in series and a current of 200 mA passes through them, what is the total potential difference across A and B?
(c) If A and B are joined in parallel, what PD across them would produce a current in A equal to half the current in B?
(d) What is the resistance of device A and how does this compare with the resistance of B with a voltage of 15 V applied?

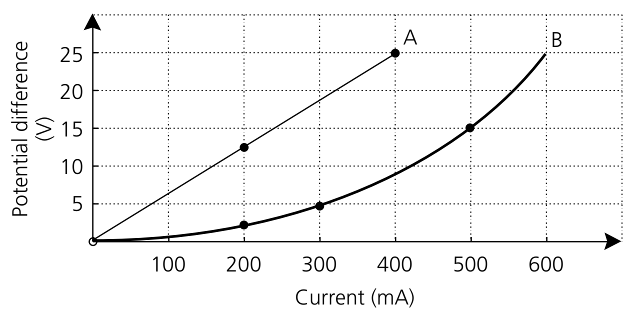


Figure 22.31
For question 30.

**31 The resistance of a 1.5 m length of conductor was measured as a function of temperature and the following data values obtained:

Temp (°C)	−50	0	50	150	250
Resistance × 10^{-2} Ω	2.6	3.35	3.9	5.3	6.6

(a) Plot a graph, resistance versus temperature. Use a line of best fit.
(b) From the line gradient, find the value of (α).
(c) Develop a conversion formula from resistance to temperature.

***32** Draw the circuit diagram that represents the following description. Two paralleled 50 Ω resistors are connected in series with three larger resistors of 100, 200 and 250 Ω. This assembly is connected to a 20 V battery and an ammeter to measure total circuit current. A voltmeter and an ammeter must be used to measure current flow through and voltage drop across one of the 50 Ω resistors, when a main circuit switch is closed.

****33** Design and draw the diagram for a two-way model circuit that will operate a lamp from two different locations in a house, say, from upstairs or downstairs.

****34** Electric kettles range in power from 1000 W to 2500 W. The rating is usually stamped underneath. Table 22.6 has been taken from *Choice* magazine and compares the performance of 10 different cordless kettles.

Table 22.6 CORDLESS KETTLES

BRAND: MODEL	MASS (kg)	CAPACITY (mL)	BOILING TIME (1 L WATER)	POWER CONSUMPTION (W)
Kambrook KU300	0.64	1500	3 min 15 s	1890
Kambrook KU400	0.64	1430	3 min 15 s	1910
Kenwood JK800	0.73	1640	2 min 50 s	2190
Linda Superboil LJ6	0.60	1500	3 min 5 s	2040
Moulinex A94	0.70	1760	3 min 10 s	2010
Ronson 8523	0.71	1680	2 min 55 s	2210
Russel Hobbs 3110	0.92	1620	2 min 50 s	2240
Sunbeam KE019	0.71	2050	2 min 40 s	2290
Tefal Freeline	0.72	1660	3 min 0 s	2120

(a) Plot a graph to determine if there is any relationship between power consumption and the boiling time for 1 L of water. Describe the relationship.

(b) Estimate the boiling times of a 1600 W kettle and a 2400 W kettle.

(c) The efficiency of a kettle can be calculated by the formula:

$$\text{efficiency} = \frac{\text{power out}}{\text{power in}} \times 100\%.$$

The power in is shown in the table as 'Power consumption'. The power out can be calculated from data about heating up the water. Assume that the water started at a temperature of 25°C. Calculate the efficiency of each kettle and make a rank order of them. Are the high powered kettles the most efficient? (*Hint:* the efficiency of the Linda superboil is 84%.)

Extension — complex, challenging and novel

*****35** The headlights on a car operate typically at 60 W and the parking lights typically at 5 W. Assuming there are two main headlights and four parking lights, what length of time will it take to discharge a 60 Ah battery, if the lights are left on?

Ch22-Walding 4th 25/8/04 12:27 PM Page 500

*****36** In the circuit of Figure 22.32, CD is a 2.0 m length of nichrome wire of value 1.8 Ω. The galvanometer connected between A and B registers no current. Calculate the value of the unknown resistor R.

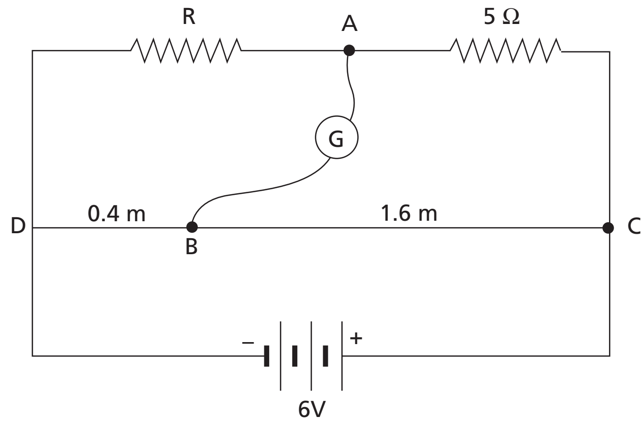


Figure 22.32
For question 36.

*****37** The diagram of Figure 22.33 is known as a Wheatstone bridge and is used to measure unknown resistances. An unknown resistance R_x is placed into the circuit and resistor R_3 is varied until the galvanometer reads zero current. This is called a null balance.

(a) Explain what occurs when a null balance is attained in the circuit, in terms of voltage.

(b) Prove that the formula for the unknown resistor is $R_x = \dfrac{R_2}{R_1} \times R_3$.

(c) Calculate the value of an unknown resistor when R_1 = 710 Ω, R_2 = 317 Ω, R_3 = 2.24 kΩ.

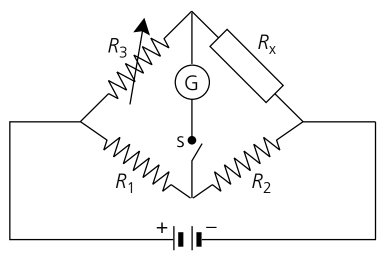


Figure 22.33
Wheatstone bridge circuit.

*****38** Figure 22.34 shows a complex circuit. Calculate the following values:

(a) total circuit resistance;

(b) voltage drop across the 60 Ω resistor.

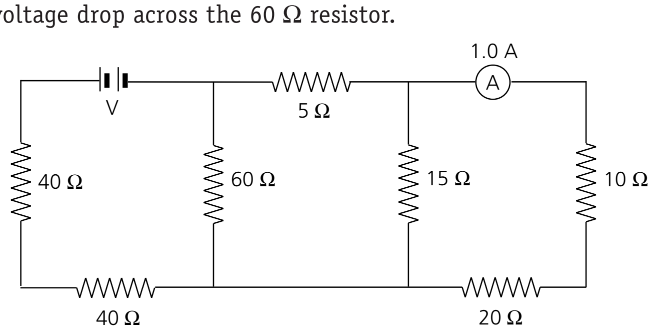
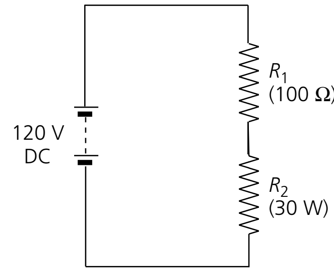


Figure 22.34
For question 38.

Figure 22.35
For question 39.



*****39** A 100 Ω resistor is connected in series with a second, unknown, resistor, R_2, and a 120 V battery, as shown in Figure 22.35. If the battery has negligible internal resistance and the unknown resistor dissipates 30 W of power, calculate its resistance value.
(*Hint:* there are actually two values that are possible in this case!)

CHAPTER 23

Electronics

INTRODUCTION 23.1

It was seen in the previous chapter that electricity is very important in modern society. Electronics is a relatively new branch of general electrical studies, which makes use of small-scale components and circuits. The development of electronics in the latter half of the twentieth century has led to many of our modern consumer electrical products and technology. Unlike the general use of electricity that dates back well over 150 years, modern electronics has expanded only since the invention of simple semiconductor devices in the early 1950s. The improvement in understanding of solid state physical devices has made possible micro-miniaturisation and very fast electronic circuit operations. These features are the basis of modern computer technology and the real start of what has been termed the 'information age'.

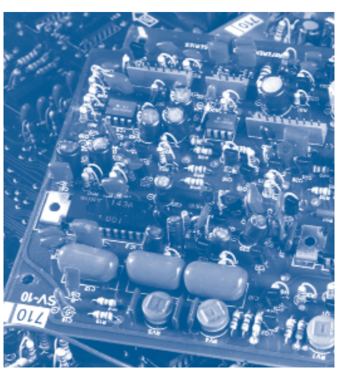
Electronic engineering today involves the use of many thousands of specialised components and circuits that have all developed from the early discoveries in semiconductor solid state physics. In this chapter we will look specifically at a set of basic electronic circuit components and their behaviour. It is from this basic set of components that electronic systems can be developed. This will be further discussed in Chapter 24. At the completion of this chapter, you will be able to answer questions like the following:

- Why do we have portable batteries but still need household electricity supplies?
- What is the difference between DC and AC voltages and currents?
- Is colour-blindness a problem for people working in electronics?
- How can some electronic devices be made so small that they can hardly be seen?
- Are electronic circuits, especially those of larger devices like TV sets and computers, really as complex as they look?

MEASUREMENT AND TEST INSTRUMENTS 23.2

Multimeters

Photo 23.1
Electronic circuit boards.



In the previous chapter, the basic electric measurement meters called the voltmeter and the ammeter were introduced. These instruments are most often used in fixed circuit applications or as stand-alone instruments to give a continuous voltage and current reading. In practical electronics a more versatile and portable measurement instrument is required. This is the **multimeter**. Its readout can be an analog needle over a scale or, more usually, a digital display readout. (See Photo 22.4 on page 483.) Modern multimeters provide numerous measurement quantities over a wide range of values, such as DC/AC voltage, DC/AC current, resistance and conductivity testing, frequency, diode conductance and capacitance. Unlike the older analog models which required the correct range of the quantity to be selected via a selector switch, newer digital models are often **auto-ranging**, which means that the input circuitry automatically senses the correct measurement range and produces an output that matches this range. For example, without altering the instrument, apart from selecting AC voltage, a technician is able to measure millivolts AC in a radio circuit, through to household domestic 240 volt AC mains.

It is always important to be aware of the measurement being made with a multimeter and check both the quantity being measured and the approximate range of the values before making any connection to a circuit. In any unknown measurement exercise it is good protection policy to select the instrument's highest scale setting and then adjust downward until an appropriate instrument scale reading is achieved. It is also often necessary on some multimeter models to adjust the input probe connectors on the face of the instrument and move them to different sockets when changing from low values to high values. The manual that comes with the multimeter will explain this. Recall also from the previous chapter that any measurement instrument should affect the circuit under test as little as possible. Especially with analog multimeters, this means that the instrument's **sensitivity range** should be as high as possible, with a value of 20 kΩ/V being regarded as the most useful in professional electronic measurement. Digital multimeters have the advantage that their sensitivity range is already very high due to the electronic integrated circuit components that comprise the internal circuitry.

Photo 23.2
CBL2 data-logger.

When using a multimeter to measure either DC or AC voltage, the probes should be placed in parallel or across the component in the circuit in exactly the same way as a conventional voltmeter. When using a multimeter to measure either DC or AC current, the probes will need to be placed in series with the component, as with a conventional ammeter. This may necessitate disconnecting one end of the component from the circuit and using the multimeter probes to remake the connection. When using a multimeter to measure resistance, it is often necessary, especially on analog models, to re-zero the meter. This means shorting the probes together and using the 'ohms adjust' knob to make sure that the needle of the multimeter actually reads zero resistance. There is a variety of different multimeters and it is always important to read the manual properly and take care in operating the instrument.

Data-loggers and sensors

You may be lucky enough to have the use of a set of measurement instruments that could make the automatic measurement of electrical and other physics quantities very easy. This is the laboratory **data-logger**; when coupled with electronic or other sensors and connected to a graphics calculator, this allows not only the recording of data over time intervals from milliseconds to months, but also the analysis of that data almost immediately. Common sensors such as current and voltage probes allow normal electrical circuit quantities to be logged, but a range of non-electrical attachments such as accelerometers, pressure sensors, temperature probes and rotation sensors allow an even greater range of physics measurements in your experiments. Photo 23.2 (a) shows an example of the Texas instruments CBL2 data-logger connected to the TI-83 Plus graphics calculator and sensors.

Photo 23.3
Power supplies.

Power supplies

A bench-top **power supply** is needed for electronic testing as it supplies the necessary DC/AC operating voltages for the components and circuit under test. Quite often these power supply units in the laboratory are referred to as power packs or rectifiers. How they are constructed will be further discussed in Section 23.6 and Chapter 31. One of the most important considerations in the use of bench-top power supplies is to avoid the voltage connector leads, the positive, negative or the AC leads, coming into contact and short-circuiting. Not only will this usually produce dangerous sparking, but the power supply itself may be damaged. The short circuit created will generally exceed the maximum current rating for which the power supply is designed and internal damage may occur if the unit is not fuse-protected or does not contain some form of current overload circuit-breaker. It is also important that you check that the voltage settings on the power supply are as low as possible before turning the instrument on. In general, changes made to any electrical circuit should only be made with its power supply connections turned off. This avoids very high voltage transient spikes from causing possible damage to the circuit components under test.

Ch23-Walding 4th 25/8/04 12:29 PM Page 503

Cathode ray oscilloscope

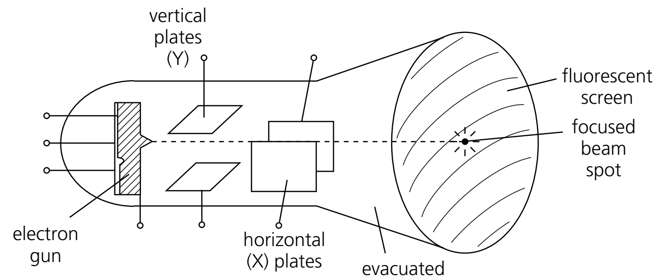


Figure 23.1
Cathode ray tube assembly.

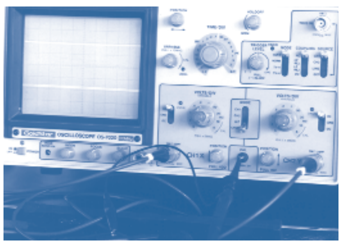
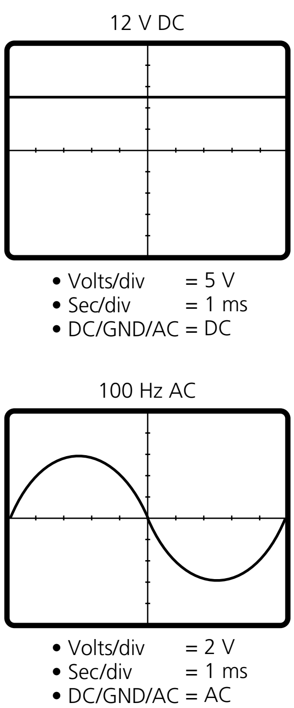


Photo 23.4
A dual beam oscilloscope.

Invented in 1897 by a German physicist, Karl Ferdinand Braun, the **cathode ray oscilloscope** (CRO) is an instrument used widely in electric and electronic circuits to display and make measurements on voltage waveforms. The CRO uses a very fast electron beam striking the face of a **cathode ray tube** (CRT), which is being deflected by rapidly changing magnetic and electric fields. The moving electron beam passes over the calibrated scale on the face of the CRT and allows measurements of voltage wave shape, amplitude and frequency. As well, it allows a simple visual inspection of the way the voltage may vary under different circuit settings. Within the evacuated cathode ray tube, an electron gun emits a narrow beam of electrons, which travel down the tube and strike a fluorescent screen at the front. Light is emitted and a bright spot is formed on the screen. (See Figure 23.1 and Photo 23.4.) This electron beam passes through a set of vertical deflection plates carrying a voltage proportional to the input voltage being measured. This can be expanded by the use of the 'vertical gain' control. At the same time the beam passes through a set of horizontal deflection plates across which is placed a regularly changing **timebase voltage**. The timebase voltage can be selected by a control on the front of the CRO and allows the user to set the time the beam takes to sweep from one side of the screen to the other. This is called the **sweep time** and can vary from microseconds to seconds. The combination of **vertical amplifier gain** and timebase produces a moving spot that perfectly matches the input voltage being tested, with vertical scale divisions representing voltage amplitude and horizontal divisions representing time. Figure 23.2 represents a typical CRO tracing for a constant DC voltage of 12 V as well as a 100 Hz sine wave AC voltage if the timebase is set to one millisecond.

Figure 23.2
CRO tracings.



Some of the common controls on a typical oscilloscope instrument are:

- Intensity and Focus — these controls allow the brightness of the electron beam as well as the spot sharpness to be adjusted on the screen.
- Channels A and B — these are the connector points for the input probes. Each probe usually has a ground clip as well as the test clip, so that the input signal can be measured with respect to the ground or zero potential. On dual-beam or dual-trace oscilloscopes, two separate voltage signals can be connected and displayed on the respective A and B inputs.
- Horizontal and Vertical positions — these controls allow the overall positioning of the spot or trace to be adjusted. For convenience, especially on dual-beam oscilloscopes, two separate voltage waveforms may need to be adjusted horizontally or vertically so that they can be more easily compared. There are separate controls for each channel.
- Volts/div — this stands for volts per 1 cm division in the vertical direction. This is the vertical amplifier gain control and allows the complete input waveform to be displayed no matter what its amplitude might be.
- Sec/div — this stands for seconds per 1 cm division in the horizontal direction. This is the timebase horizontal deflection adjustment. The selector switch will set the time interval representing each horizontal division.
- Trigger level — this setting determines the point at which the beam begins its sweep across the screen. It allows synchronisation of the sweep timebase waveform and the input test waveform so that the signal trace is stable on the screen and does not drift about. The trigger may be either an internal instrument signal or an external signal.
- DC/GND/AC — this selector switch is set to the type of input waveform being measured. For example, if set to AC, any DC component of the input test voltage signal will not cause extra vertical deflection.

AC waveform analysis

Consider Figure 23.3, which illustrates the voltage waveform as seen on an oscilloscope connected across resistor R. In the DC case the voltage waveform is easily described as a single fixed voltage; however, in the AC case there exist several alternatives in describing the magnitude of the voltage. Note that in the AC case, the frequency of the waveform would almost certainly be 50 Hz as supplied by any standard power supply.

Whereas direct current (DC) only flows in one direction, alternating current (AC) reverses direction at a given frequency as determined by the alternating voltage. Often the AC voltage can be mathematically described as a sinusoidal voltage (sine wave) represented by the expression:

$$V = V_0 \sin\omega t = V_0 \sin(2\pi f)t$$

where V = instantaneous voltage; V_0 = voltage maximum; f = frequency; t = time.

Let us look carefully at the methods of describing the AC waveform magnitude or amplitude. It is obvious that the average of all instantaneous voltage points over one complete AC cycle will be zero volts, as the positive area equals the negative area if the waveform is perfectly symmetrical about the zero axis. One method called the **peak voltage**, V_P, is to read the scaled vertical amplifier from the zero axis up to the top of the sine wave. In Figure 23.3, if the vertical amplifier was set at 5 volts/div then the reading would be V_P = 10 V. A second method is to record the **peak to peak voltage**, V_{PP}, off the vertical amplifier scale from the top of a crest to the bottom of a trough. Thus $V_{PP} = 2V_P$ = 20 V. The most common method is to employ the reading known as the **root mean square voltage**, V_{RMS}.

It can be shown mathematically that:

$$V_{RMS} = \frac{V_P}{\sqrt{2}} = 0.7\ V_P$$

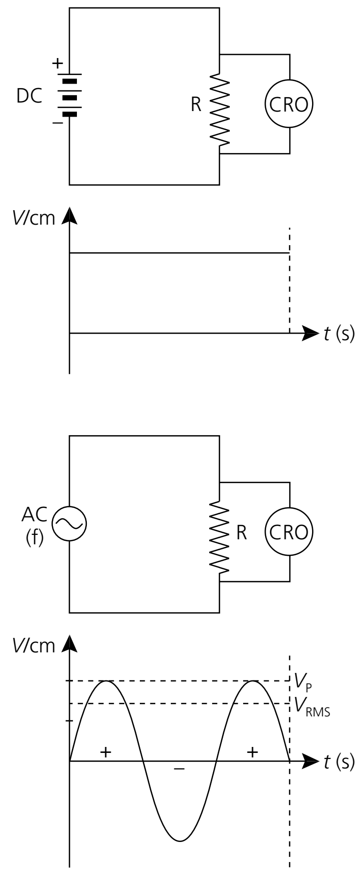
In the AC waveform of Figure 23.3:

$$V_{RMS} = \frac{10}{\sqrt{2}} = 7.07\ V$$

The reason why the V_{RMS} method is most often used to describe AC voltages is that the average power dissipated in a resistor by any RMS voltage over one complete cycle will be identical to the power dissipated in the same resistor by an equivalent DC voltage. In terms of resistive electrical power dissipation, V_{RMS} (AC) = V_{DC} and thus a method of direct comparison is obtained.

An AC voltmeter or multimeter set to read AC voltage will usually display its output directly on an RMS scale. An oscilloscope will most conveniently display the waveform as a measured V_P or V_{PP} value. In some electronic circuits, the actual voltages present at particular points as viewed on an oscilloscope will often contain both DC and AC components. In these instances, the oscilloscope can be used to measure the DC offset voltage as well as the RMS ripple voltage.

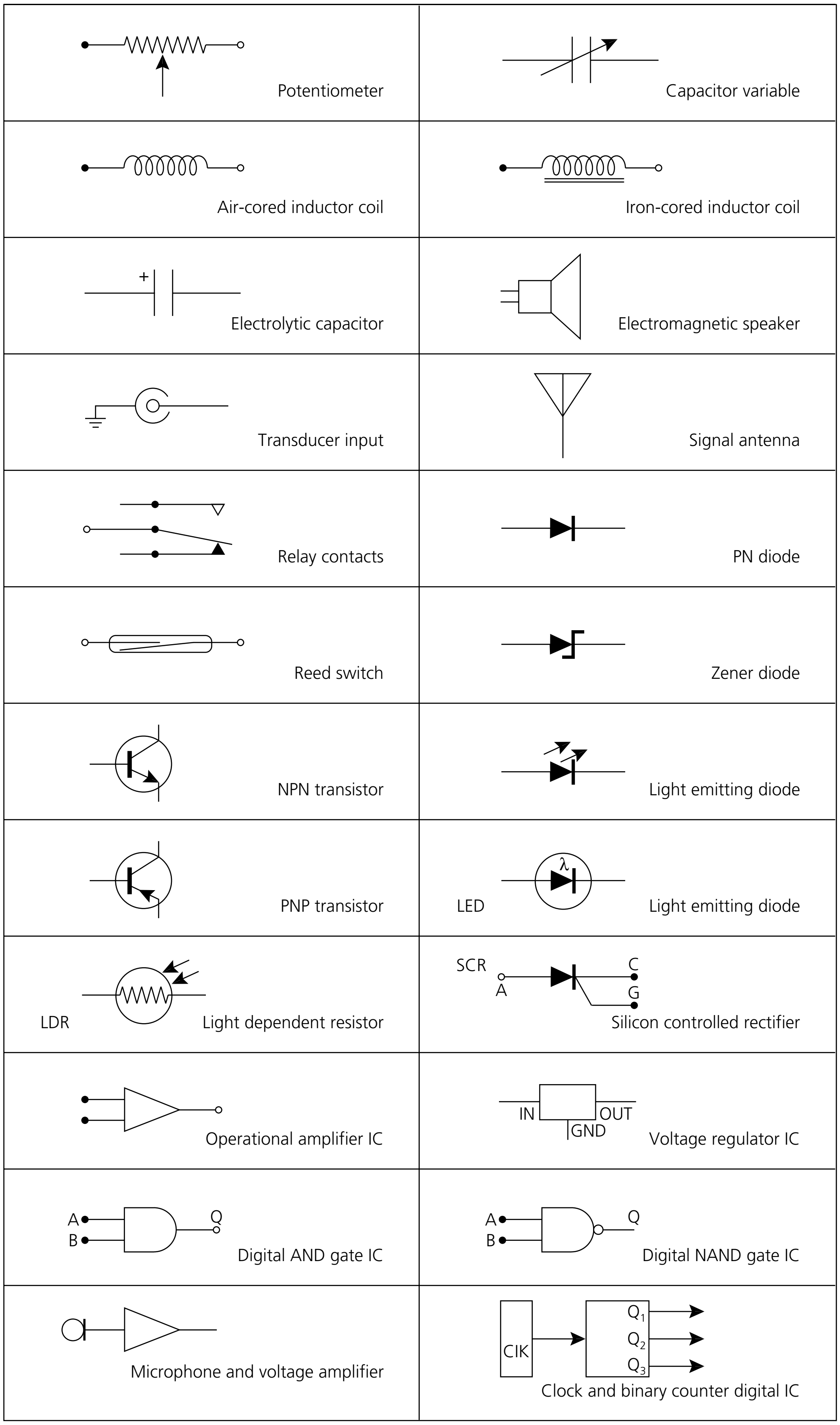
Figure 23.3
AC waveform analysis.



Circuit symbols

Recall that in the previous chapter in Figure 22.19, several electric circuit symbols were introduced. At this point it would be appropriate to introduce a further set of specific electric circuit symbols for devices that will be met from this point onward. Any table such as this can never be complete and reference to other textbooks and electronics magazines will build up your knowledge and ability to recognise a large number of electronic circuit symbols. (See Figure 23.4.)

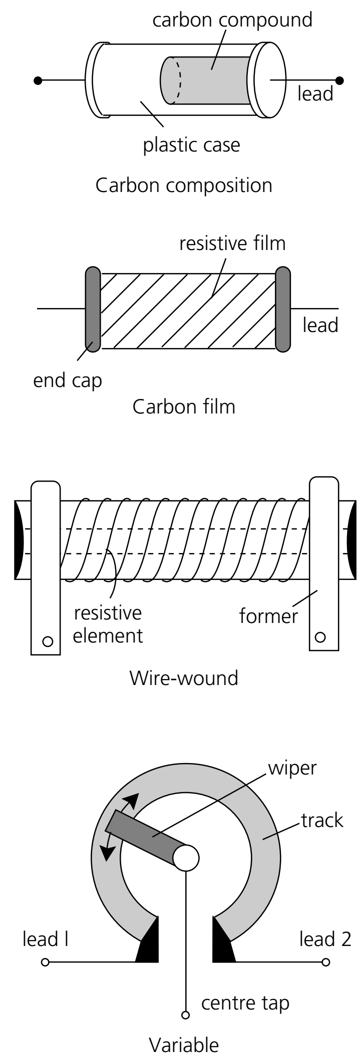
Figure 23.4
Electronic circuit symbols.



23.3 RESISTORS IN ELECTRONICS

At this point it would be useful to revise the basic properties of resistance, as described in Section 22.4. In this chapter we will look closely at the practical **electronic resistor** component and its uses. Resistors are simple devices used to control the flow of electric current as well as act in voltage divider networks. Resistors are generally of two types: fixed and variable. They differ considerably in physical size due to the total electrical power they are required to dissipate in a circuit. A resistor dissipates energy when a current flows through it and the resistor consequently heats up. It must be able to lose this heat energy to the air or cooling elements without being damaged. The rate of energy dissipation, or its power rating, is determined by its physical size and shape. The greater the resistor's surface area, the greater is its **power rating**. Commonly in electronics, manufacturers mass-produce resistors that are able to dissipate one-quarter watt ($\frac{1}{4}$ W) and ($\frac{1}{2}$ W), (1 W) or (5 W). Resistors made from wire-wound elements are able to dissipate even greater power. When used in electronic circuits, the power rating of a resistor must exceed its actual operating power dissipation, otherwise physical burn-out will occur — a possible cause of electrical fires in poorly designed electronic devices.

Figure 23.5
Resistor types.



Example

In a simple DC circuit, with a voltage supply of 9 V, what value series resistor will limit the current flow to 90 mA? What is this resistor's power rating in the circuit and what would be the best resistor type to purchase for the circuit construction?

Solution

Given $V = 9$ V, $I = 90 \times 10^{-3}$ A, and using:

$$V = I \times R$$
$$9 = 90 \times 10^{-3} \times R$$
$$R = 100\ \Omega \text{ as the value of the resistor}$$

Power rating $P = V \times I$, $P = 9 \times 90 \times 10^{-3} = 0.81$ W.
Hence the best resistor would be a 100 Ω, 1 W type.

Fixed resistors

In electronics, resistors are usually made of some form of carbon mixture. The simplest type of resistor is the **carbon composition** resistor, made of finely divided graphite carbon mixed with a powdered insulating medium such as crushed clay in a defined proportion. This mixture is pressed into a plastic case with two connector leads embedded into the mixture and supported by the case (Figure 23.5). A more recent construction is the **carbon film** resistor in which a graphite carbon film is deposited on the outside of a ceramic rod. A spiral pattern is wound around the rod so that the effective length of the resistor film is often considerably longer than the rod itself. Special metal end caps are welded to the connector leads and pressed over the ends of the rod to complete the electrical path. **Metal film** resistors are similar to carbon film except that the resistive element is made from a metal oxide such as tin oxide. The advantage is that closer tolerance to the stated resistor value in ohms can be maintained, and the oxide film has a much longer lifetime than carbon film; however, they are more expensive to make. A **wire-wound** resistor can be made in which a resistive wire element (nichrome) is wound around a small core called a 'former'. The main advantage with this type is the higher power rating that can be achieved especially if the former is hollow and air- or fluid-cooled. Generally, manufacturers use an outer coating of paint or epoxy compound over the resistor in order to print onto the component the descriptive **colour code** or simple ohmic value, power rating and tolerance percentage. Electronics technicians become quite expert at recognising resistor values directly from their printed colour codes.

Variable resistors

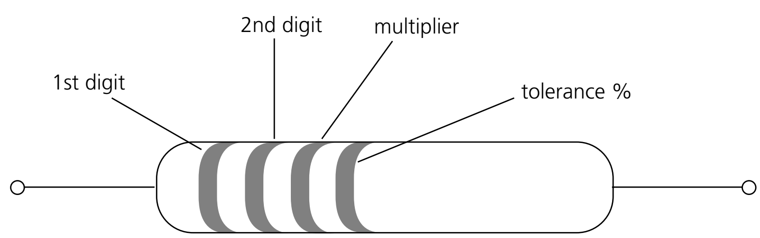
Not all electronic resistors are fixed in value; a host of different types of variable resistors are manufactured. A **variable resistor** is generally a two-terminal device that is used in circuits to vary the electric current flowing; however, three terminal devices called **potentiometers**, or pots, are also common and these are generally used to vary voltage or potential in applications such as volume controls on radios and rheostats in light dimmers. The term 'pot' is now generally used in electronics to describe any variable resistor, no matter what its primary function. The resistive element of these devices usually is either a carbon deposit on an insulating surface, or a resistive wire wound around a suitable former. A moving arm or **wiper** is drawn across the resistive element and connects to a metal contact, with the total resistance of the device changing from zero ohms at one end to the maximum ohms value at the other end.

Often the resistive element design follows a **linear** relationship, where the resistance is proportional to how far the wiper arm has moved, called an A-type pot. In other cases the resistance change is **logarithmic** in nature. This is called a C-type pot and is most often used as a volume control in an audio amplifier as it matches the human ear's logarithmic response to sound volume changes.

Preferred resistance values and colour codes

All resistors have three basic parameters: wattage, resistance and tolerance or accuracy. Physical size and type determine the wattage. Some manufacturers use a cream body colour for carbon film resistors and a blue or green body for metal film types. Manufacturers use a standardised band colour code printed on smaller resistors where physical lettering is not

Figure 23.6
Resistor colour code.



Black - 0
Brown - 1
Red - 2
Orange - 3
Yellow - 4
Green - 5
Blue - 6
Violet - 7
Grey - 8
White - 9

Gold - 5%
Silver - 10%
None - 20%

Example
Red Red Green Silver
is
2.2 M Ω ± 10%

possible. The resistor colour code illustrated in Figure 23.6 and in the centre colour pages is one of the best known in electronics and with a little practice in reading the code it is easily committed to memory. The coloured bars always start closer to one end of the resistor and this is the bar that is used first in order to deduce the value in ohms of the resistor. If ever there is any doubt about the resistor colour code, then a multimeter should be used to determine a resistor's actual ohmic value.

The bands of the colour code as marked on a resistor give both the ohmic value and its tolerance as follows:

- Band 1 — first figure of ohmic value.
- Band 2 — second figure of ohmic value.
- Band 3 — multiplier for the first two figures.
- Band 4 — tolerance band colours, where brown is 1%, red is 2%, gold is 5%, silver is 10%, and none represents 20%.

Note that some resistors have a five band colour code and in this case the first three bands give the first three digits in the value, the fourth in the multiplier and the fifth is the tolerance.

Because it would be impractical to manufacture every possible value of resistance, there exists a **preferred value** set of resistors used in circuit construction. The most common preferred value range is known as the E12 series (12 values per 100). An E24 series also exists (24 values per 100). These preferred value sets apply to each of nine decade ranges or multipliers from $\times 10^{-2}$ to $\times 10^{6}$ (Table 23.1).

Table 23.1 RESISTOR PREFERRED VALUE SERIES

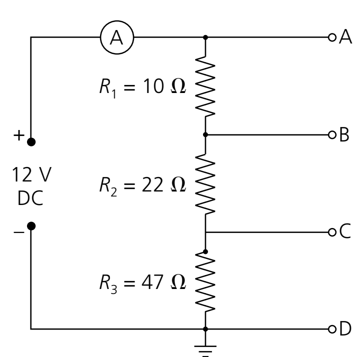
E12 SERIES (Ω)			E24 SERIES (Ω)			
10	33		10	18	33	56
12	39		11	20	36	62
15	47		12	22	39	68
18	56		13	24	43	75
22	68		15	27	47	82
27	82		16	30	51	91
Decades		100–1000				
0.01–0.1		1000–10 000				
0.1–1		10 000–100 000				
1–10		100 000–1 000 000				
10–100		1 000 000–10 000 000				

Thus, any manufactured resistor must begin with a number from the E12 or E24 series and be a multiple from 10^{-2} or 10^{6}. It should be noted at this point that in some resistor applications in electronics, critical values of resistance are not important, as a circuit will often perform quite successfully with anything up to about 20% tolerance in resistor values.

Resistors as voltage dividers

Recall that the total resistance in series of several resistors is the sum of each individual resistance. If the circuit of Figure 23.7 is set up, the respective voltage drops across each resistor in the series chain can be used as a separate voltage supply. That is, the 12 volt DC supply can be split into three separate voltages, V_{AB}, V_{BC}, V_{CD}. Ohm's law can be used to calculate the respective voltages and maximum currents that can be supplied using such a **voltage divider** circuit. Any electronic device, such as a small motor, connected to any output of the divider cannot draw a very large current compared with that flowing around the divider circuit, otherwise the effect on the voltage divider will be quite large. The general rule is that as the current drawn from a voltage divider circuit increases, the output voltage decreases. Refer to the following example.

Figure 23.7
Voltage divider circuit.



Example

Consider the circuit shown in Figure 23.7. Calculate:

(a) the current drawn from the 12 V supply as registered on the ammeter;
(b) the value of the voltages V_{AB}, V_{BC}, V_{CD};
(c) the most appropriate connection points for a 3 V, 10 mA motor.

Solution

Total circuit resistance = $R_1 + R_2 + R_3 = 79\ \Omega$.

(a) To calculate the ammeter current, use:

$$I = \frac{V}{R_{tot}} = \frac{12}{79} = 150\ \text{mA}$$

(b) To calculate the respective voltages, use:

- Voltage $V_{AB} = I \times R_1 = 150\ \text{mA} \times 10\ \Omega = 1.5$ V.
- Voltage $V_{BC} = I \times R_2 = 150\ \text{mA} \times 22\ \Omega = 3.3$ V.
- Voltage $V_{CD} = I \times R_3 = 150\ \text{mA} \times 47\ \Omega = 7.0$ V.

(c) As the required 10 mA maximum is considerably less that the 150 mA flowing through the divider network, the motor will not load the circuit and the most appropriate connection points to operate the motor are voltage V_{BC}.

In order to produce a continuously variable voltage divider it is necessary to substitute a potentiometer for the fixed resistor chain. **Volume controls** on a radio or TV work this way, with the potentiometer being a 10 kΩ or 20 kΩ type. The movable wiper of the pot produces a continuously variable output voltage from zero to the maximum available at the divider input. These devices are usually controlling very small voltages and currents present within a radio receiver or small amplifier.

EI Activity 23.1 TEACHER RESISTANCE

Obtain from your teacher numerous examples of different resistors. Use the colour code to determine their nominal resistance value, or read it directly and compare this with the actual value as determined by a multimeter set to read resistance. Be careful to correctly set the multimeter to its correct range and check that it is zeroed properly by shorting its measurement probes and checking for zero ohms. Use the 'ohms adjust' control if necessary.

Questions

1. List four different types of electronic test instruments and explain their functions.
2. If the waveform being observed on an oscilloscope is too small and the peaks too close together, explain the adjustments necessary to redisplay the waveform with greater clarity.
3. Convert 15.6 V_{RMS} to a peak to peak value.
4. A resistor of value 470 Ω is required to control a current of 250 mA. What minimum power rating is appropriate? What is its colour code markings?

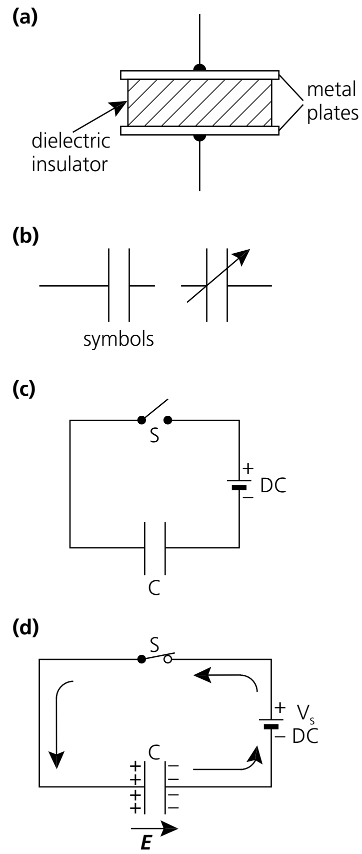
CAPACITORS, INDUCTORS AND RELAYS 23.4

Resistors dissipate energy in the form of heat and hence they cannot store energy. Two components that are capable of storing energy in electrical circuits are the **capacitor**, which stores energy in an electric field, and the **inductor**, which stores energy in a magnetic field.

Capacitors and DC

A capacitor can be likened to a very fast storage tank for electric charges. It consists basically of two conducting plates separated by an insulating material called the **dielectric**. This assembly is often housed in a protective outer coating with either the plates held flat as in a disc, or with the metal foil plates and dielectric rolled up to form a cylinder. In both cases, either via markings or colour coding, the outer casing holds information such as capacitance value, tolerance and working voltage. Connecting leads allow the flow of current to and from the capacitor's plates (Figure 23.8 and Photo 23.5).

Figure 23.8
Capacitors: dielectric **(a)**; symbols **(b)**; uncharged **(c)**; charged **(d)**.



According to literature, the first capacitor, called a Leyden jar, was discovered almost simultaneously by Dean von Kleist of the cathedral of Camin in Germany in October 1745, and Peter von Muschenbrock, professor in the University of Leyden, in January 1746. As described by them, it was a glass jar or vial with inner and outer electrodes made of various substances — water, mercury, metal foil, etc. The modern miniature glass dielectric capacitor differs in form and structure from the 250-year-old Leyden jar, but the principle of operation is the same.

If a DC voltage is applied to a capacitor, an electric current will carry charge to the plates, so that one plate becomes positively charged and the other is left negatively charged. When a capacitor is fully charged, several conditions exist:

- There is equal and opposite charge on the plates.
- The voltage across the plates equals the supply voltage, V_s.
- An electric field exists within the dielectric.
- Any further flow of direct current (DC) is blocked.

The **capacitance**, C, of this system is defined as the amount of charge in coulombs, Q, stored on each plate when the potential difference voltage, V, across the plates is 1 V. Capacitance $C = \frac{Q}{V}$ is in coulombs per volt; 1 C V^{-1} = 1 farad (F). The unit called the farad is named after Michael Faraday (1791–1867), the great British physicist who was the first to develop the idea of electric and magnetic fields.

Typical capacitors in electronics have values ranging from one microfarad (1 μF) to one thousand microfarads (1000 μF). The capacitance of any capacitor is dependent on the type of dielectric material it contains and, because the dielectric material is usually very thin, all capacitors have a maximum **working voltage** rating. **Capacitor leakage** refers to the amount of charge that is lost during capacitor operation when a voltage is applied across its terminals. Generally with modern electronic capacitors, leakage is not an important factor in circuit design.

Photo 23.5
Capacitors and inductors.

Using calculus, it is possible to calculate, in joules, the total energy, W, stored in a fully charged capacitor. If Q is the charge stored on each capacitor plate, C is the capacitance and V is the working voltage across the plates, then:

$$W = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C}$$

Generally capacitors are used in one of three ways in electronic circuits.

- In conjunction with resistors, they can be used in timing circuits, making use of the length of time it takes for a particular voltage across the capacitor plates to appear.
- To bypass or filter rapidly changing AC frequencies and block the flow of DC.
- To eliminate voltage fluctuations arising from the conversion of AC voltage to DC voltage.

These last two applications are further discussed in Sections 23.5, 23.6 and Chapter 31.

Types of capacitors

Different types of capacitors vary widely in their electrical characteristics. Let's look at the most common types.

NOVEL CHALLENGE

One ‘bit’ of computer memory consists of a capacitor which stores a half a million excess electrons. In 1970, they required 2 million electrons per ‘bit’. In 2025 it is predicted that they will be down to 1 electron per bit. Wouldn’t it make more sense to try and get more electrons into each capacitor, not fewer? After 2025, could they try for half an electron? Explain the fallacy of this argument.

Plastic-film capacitors These are constructed as shown in Figure 23.8, although most often the metal foil plates and insulating dielectric are rolled into a cylinder. The dielectric for these capacitors is usually mylar or polyester. The popular greencap capacitor uses a metallised polyester film. They have good temperature stability and are used commonly in audio, radio, computer and general electronics. Polystyrene dielectric or styroseal capacitors have a very low leakage rate but they are quite expensive, while mains (240 V AC) rated capacitors use a polycarbonate dielectric that can repair itself. Plastic film capacitors in general are non-polarised, which means they can be connected with either wire to the positive.

Ceramic capacitors These are most commonly constructed in the form of a flat circular disc. Silver is vaporised onto both sides of a ceramic material that produces a large capacitance for a small size. This type of capacitor has a very high ability to withstand high voltages without breaking down. Their excellent temperature stability and low inductance make them useful in digital circuits and radio–TV tuning circuits.

Electrolytic capacitors An aluminium oxide electrolyte forms the dielectric on the surface of the plates. This type of capacitor is polarised, meaning that it must be connected into circuit the right way round as indicated on the casing, one end to positive and the other to negative. Reverse polarisation allows DC conduction, with possible heating up and explosive breakdown. The electrolytics are usually high capacitance (1 –10 000 μF) but have rather high leakage rates and are thus used in low frequency applications, power supplies and in audio amplifiers. If the electrolyte dielectric is tantalum dioxide, the capacitor is known as a **tag tantalum**. This type has very low leakage and high stability but is expensive.

Variable capacitors These are used in conjunction with inductors in radio–TV tuning circuits. Usually they are constructed with two sets of metal vanes that can be rotated against each other on a common shaft. The degree of plate overlap determines the capacitance. The dielectric is often air, or mylar plastic film. A small version of the variable capacitor gang is usually known as a **trimmer**. (See the Photo 23.5 and Chapter 31).

Combining capacitors

The pigtails or connecting leads attached to capacitor plates are usually arranged to exit either from both ends (axial type) or from the same end (PCB mount type). This allows easy multiple connections in parallel or series to produce variations in capacitance if required. If two capacitors are connected in parallel, the area of the plates is effectively increased. This allows more charge per unit voltage, hence a linear increase in effective capacitance. If two capacitors are connected in series, the effective thickness of the dielectric is increased, resulting in a lower charge storage per unit voltage, that is, an effective capacitance decrease.

- Parallel connection: $C_{tot} = C_1 + C_2$
- Series connection: $\dfrac{1}{C_{tot}} = \dfrac{1}{C_1} + \dfrac{1}{C_2}$

Note that these combination rules for capacitors are opposite in behaviour to resistors. (See Figure 23.9.)

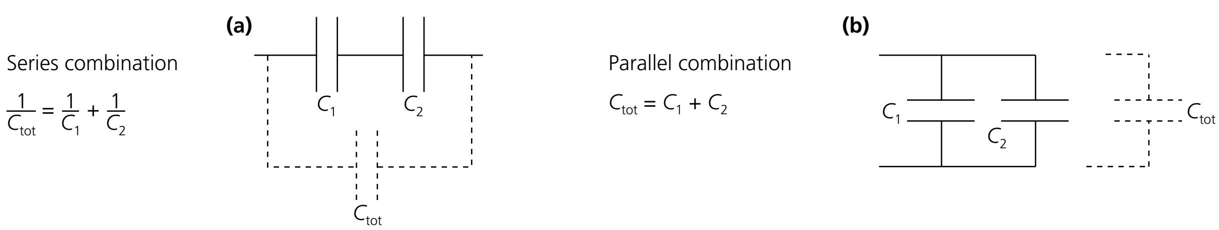


Figure 23.9 Combining capacitors: series combination **(a)**; parallel combination **(b)**.

The RC timing constant circuit

If you reconsidered the charging circuit of Figure 23.8**(d)** and added a series resistor, then the time it takes for the capacitor to become fully charged would be increased. The rate at which a capacitor charges (the voltage across it approaches the supply voltage) is determined

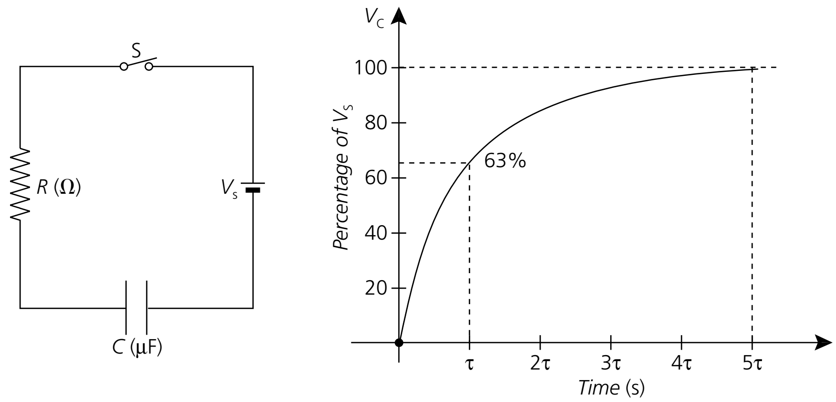


Figure 23.10
Charge curve for a capacitor.

by the size of the series resistance in the circuit, as this resistance controls the rate of flow of charge to the plates. The charging curve of a capacitor is shown in Figure 23.10. It is an exponential curve and the length of time it takes for the voltage across the capacitor to reach 63% of the supply voltage, V_S, is called one **time constant**. This is given the symbol τ, tau, which is the Greek letter *t*. Notice that the charging exponential curve shows that in theory a capacitor never really fully charges, but mathematical calculations show that after 5 τ the capacitor voltage V_C = 99% of V_S. Mathematically, one time constant period, measured in seconds, in this type of **RC circuit** is given by

$$\tau = R \times C$$

NOVEL CHALLENGE

It is sometimes said that the discharge of a capacitor is mathematically similar to the emptying of a bucket of water with a hole in the bottom. *Explain why the two discharge curves are the same.*

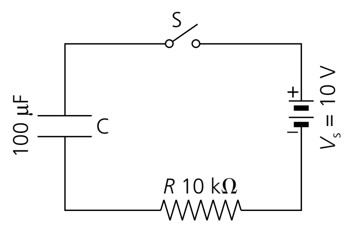
Symmetry of shape between the discharge curve and the charge curve allows us to predict that when a capacitor discharges through a resistor, its voltage will drop by 63% of its initial value also in one time constant period ($\tau = RC$).

The timing circuit that is produced by an RC series connection has many uses. To make a **timer** we need a circuit that triggers when a voltage reaches a particular value as determined by a charging capacitor. RC networks can also be used to modify a waveform shape. Usually RC networks are used as **filters**, such as bass and treble or loudness controls in a stereo amplifier. These RC filter circuits alter the frequency response and waveform shapes of the electric signals at their inputs.

Example

Consider the circuit of Figure 23.11, showing an RC timing circuit with a supply voltage of 10 volts DC. Calculate the time constant, τ, for the circuit and the time it takes for C to become fully charged after switch S is closed.

Figure 23.11



Solution

- If V_S = 10 V, C = 100 μF, R = 10 kΩ,
- then $\tau = RC = 1 \times 10^4 \times 1 \times 10^{-4} = 1.0$ s.

This will represent a voltage V_C = 63% V_S = 6.3 V.
The capacitor will be fully charged, V_C = 10 V, after 5 τ = 5.0 s.

EI Activity 23.2 CAPACITOR COLOURS

A capacitor purchased from an electronics store has the following markings on it: 473K–1000. This is known as the E.I.A. marking code and allows the capacitance, tolerance and voltage rating to be read.

1. Research the E.I.A. capacitor marking code to determine what these markings represent.
2. Some capacitors, such as TAG tantalums and polycarbonates, have a colour code system. Research the meaning and design of this colour code on these capacitors.

Inductors and DC

An inductor is another component for storing energy in an electronic circuit. Energy is stored in the form of a magnetic field. **Inductors** are simply coils of fine wire wound on a rigid former and vary in size from very large air-cored devices to handle large currents, to small ferrite (compressed iron dust)-cored devices often called 'choke coils' because they are used to severely reduce certain frequencies in **AC circuits**. (Refer to Photo 23.5.) Two inductors wound on the same laminated core assembly are called transformers and are used to change the magnitudes of AC voltages; see Chapter 26 for more details. Variable inductors can be easily manufactured using a ferrite rod that can be moved in and out of the inductor core (Figure 23.12). In fluorescent lights the inductor is often called the 'ballast'.

Figure 23.12
Inductor coils.

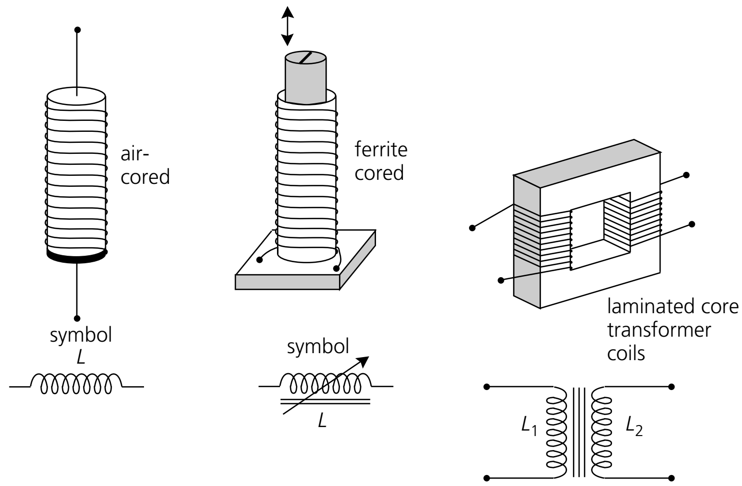
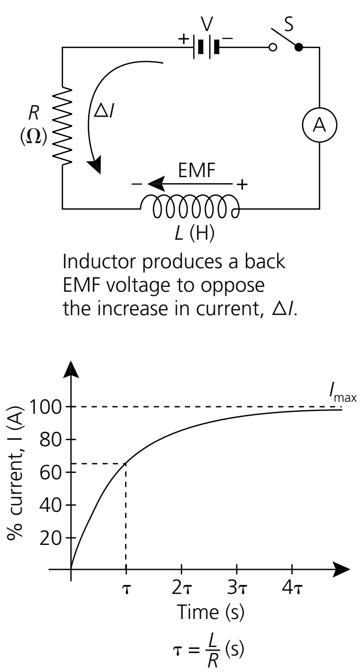


Figure 23.13
Inductor time constant.



Several principles are involved in explaining the operation of an inductor. Basically, the inductor follows the theory of an **electromagnet** or **solenoid**, as will be discussed in Chapter 25. It is important here to realise that:

- a constant direct current (DC) flowing through an inductor coil produces a constant magnetic field surrounding the inductor, which is similar in shape to a bar magnet field
- varying DC flowing through the inductor will cause the magnetic field to vary
- the varying magnetic field that cuts the turns of the inductor will induce a voltage or EMF across the inductor. This induced EMF always opposes the flow of a changing electric current through the inductor. Thus, an inductor has greatest influence on continuously varying AC currents and theoretically has no effect at all on DC.

If an inductor is connected into a DC circuit with a resistance, it will take longer for the current flow to reach its maximum value (as determined by Ohm's law) than if the inductor was not present (Figure 23.13). This property of an inductor is called its **self-inductance**, *L*, and is measured in a unit called a **henry** (H). An inductor has a self-inductance of 1 henry if an EMF of 1 volt is induced across its ends for a change of current of 1 amp per second.

$$V = -L\,\frac{\Delta I}{\Delta t}$$

Typical inductors in electronics have values of microhenrys (μH) or millihenrys (mH).

If the graph of Figure 23.13 is considered, the increase in current through the inductor produces an effective inductive time constant, τ, where $\tau = \frac{L}{R}$. Again, one time constant period is the time it takes for the current through the inductor to reach 63% of its final ohmic value.

All inductors store energy in their associated magnetic fields. When the DC current I reaches a constant value, the energy being stored in the coil will remain constant and is determined by the formula:

$$W = \frac{1}{2} I^2 L \text{ (joules)}$$

where L is the self-inductance in henrys.

Inductors connected in series in a circuit effectively increase the total self-inductance:

$$L_{tot} = L_1 + L_2 + L_3 + \ldots \textit{ in series}$$

Inductors connected in parallel effectively decrease the self-inductance:

$$\frac{1}{L_{tot}} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots \textit{ in parallel}$$

Reed switches and relays

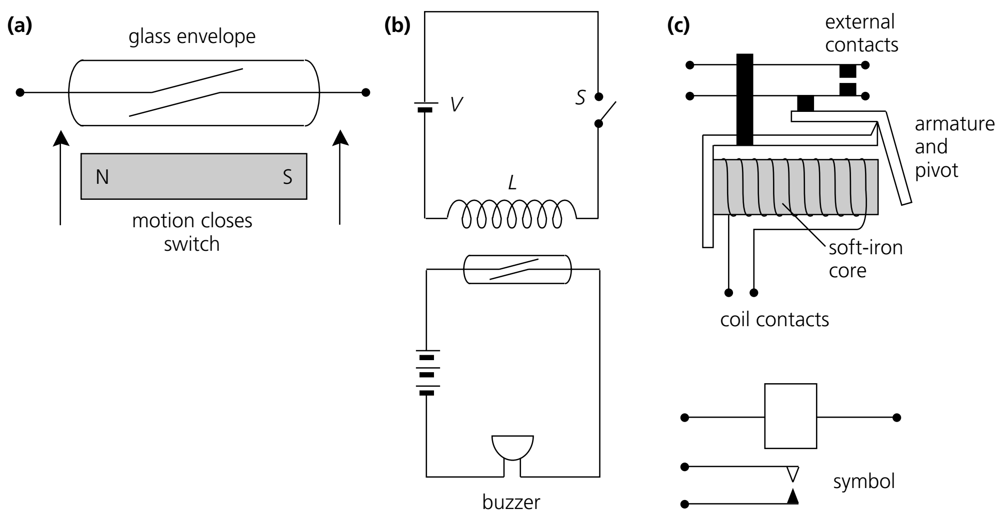


Figure 23.14
Reed switches and relays: simple reed switch **(a)**; reed relay **(b)**; electromagnetic relay **(c)**.

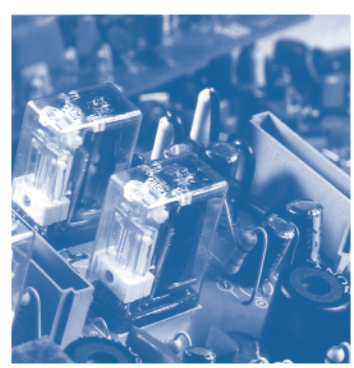
A simple **reed switch** consists of two pieces of metal inside a glass housing, which will contact together when an external magnet is brought into close proximity. This closed reed switch can then be used to carry an electric current and turn on a light bulb, for example (Figure 23.14**(a)**).

A **reed relay** replaces the permanent magnet in the reed switch assembly with an inductor coil carrying an electric current. The magnetic field generated by the inductor again closes the reed switch. The reed relay is most useful where one circuit is needed to independently control another. In Figure 23.14**(b)**, a small inductor coil current can control the larger DC current required to operate the buzzer alarm.

An **electromagnetic relay** takes the concept of a reed relay further. It has an L-shaped armature of soft iron, which is attracted to a much stronger iron-cored inductor coil energised by a small DC voltage (Figure 23.14**(c)** and Photo 23.6). The L-shaped armature is arranged generally to open or close sets of contacts. In this way one relay device may trigger several sets of external circuits from one coil. When the coil current stops, the relay latches out and breaks the external contacts.

Electromagnetic relays, switches and inductors are very common in household devices, such as washing machines and electric circuit-breakers. These applications of electromagnetism will be further discussed in Chapters 25 and 26.

Photo 23.6
Circuit board relays.



Questions

5 What is capacitance? If a capacitor is marked 0.01 μF, 240 V, what does this mean?
6 What are the common types of capacitors used in electronics?
7 Determine the time constant for a series RC circuit where $V = 12$ V, $R = 120$ kΩ, $C = 100$ pF.
8 How does an inductor store energy and what determines the amount of energy stored in any particular inductor?
9 An inductor coil has a resistance of 5.0 Ω and an inductance of 120 mH. If a voltage of 12 V is applied, what time will it take for the coil current to reach its maximum value and what will this current maximum be?

SEMICONDUCTOR DEVICES 23.5

Conduction and semiconduction

Metals are crystalline materials, with a structure made up of atoms in a regular lattice pattern. The atomic nuclei and inner electrons of the atoms are fixed in position in the lattice, while the outer electrons are virtually free to drift, at normal temperatures under thermal equilibrium, in an electron sea. If an electric field is applied across the metal, these conduction electrons drift with a velocity of about 1×10^{-4} m s^{-1} and provide excellent conductivity. Copper, silver and gold are among the best electrical conductors. Electrical insulators such as glass or plastic do not possess an electron sea and, hence, no free conduction electrons. In order to break the bonds between electrons and nuclei, extremely high voltages are necessary. They do not normally conduct electricity.

Semiconductor materials lie between these two conduction extremes and are made from elements of group IV on the periodic table. Some common **semiconductor** crystalline materials are silicon, germanium, gallium arsenide and cadmium sulfide. It is from this class of materials that devices in modern solid state electronics, such as diodes, transistors and integrated circuits, are constructed. Let's look closer at their conduction properties and how to improve them, using pure silicon as the primary example (Figure 23.15).

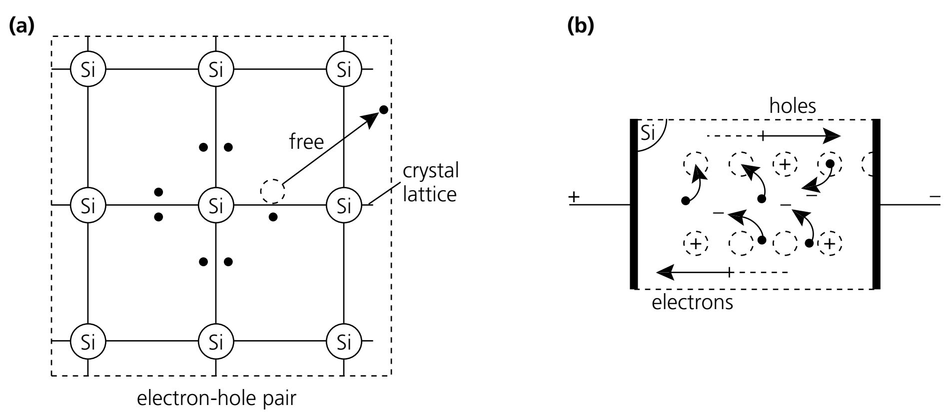


Figure 23.15 Intrinsic semiconduction in silicon: semiconductor lattice **(a)**; electron charge carriers **(b)**.

Figure 23.15 represents, in 2D, a small region of silicon crystal with the atoms tetrahedrally bonded to each other, with four pairs of electrons sharing covalent bonds. At temperatures above absolute zero (when semiconductors are in fact good insulators) some electrons break away from the nuclei and move into a conduction band where they are able to move easily through the crystal under normal thermal excitation.

In silicon, the thermal energy necessary for a conduction electron to break free is about one electron-volt (1 eV). Notice that this leaves behind a fixed hole in the valence bond of the crystalline lattice into which a free electron may drop. Thus, **electron-hole pairs** are produced randomly at normal ambient temperatures. As the temperature rises, more energy is available and so more electron-hole pairs are produced. Electron-hole pairs are being created and destroyed constantly in the semiconductor lattice as electrons fall back into holes, but an equilibrium is established. It is these electron-hole pairs that are able to conduct electricity through the semiconductor crystal when an external electric field is applied. Notice that semiconductors will therefore increase their conductivity at higher temperatures, which is an effect opposite to that of normal metallic conductors. A disastrous thermal breakdown in semiconductor electronic components can occur if they are allowed to overheat.

Electrical conduction through a semiconductor crystal via negative charge carriers (electrons) and positive charge carriers (holes) is called **intrinsic semiconduction** and is usually not very practical for electric circuits and devices. The basic conduction of semiconductors can be improved by a chemical process calling **doping**, whereby small proportions of impurity atoms are added to the crystal. This is done during a high temperature molten phase so that when the mixture recrystallises, the dopant impurity atoms will take the place of normal silicon atoms in the lattice pattern. Typical doping ratios are of the order of 1 in 10^8 atoms. Doping can also be produced by neutron bombardment (NTDS). Refer to Chapter 28.

N-type silicon

The doping addition of pentavalent atoms from group V of the periodic table, such as phosphorus, arsenic or antimony, produces extra electrons available for conduction because these atoms contain five valence electrons, one more than silicon. Since most of the current carriers are electrons, they are referred to as the **majority charge carriers** in this form of **N-type silicon**. Of course, there are still available holes and these are called the **minority charge carriers** in the conduction process through this type of silicon.

P-type silicon

The doping addition of small amounts of trivalent group III elements, such as aluminium, boron or gallium, produces extra holes available for conduction as these atoms have only three valence electrons, one less than silicon. Within the lattice of **P-type silicon**, majority charge carriers are thus holes, while the minority charge carriers are electrons.

The added versatility and improved electrical conduction of doped semiconductor crystal is much more useful and under certain circumstances approaches the very good conductivity of normal metals. The use of doped semiconductor properties is referred to as **extrinsic semiconduction** (Figure 23.16).

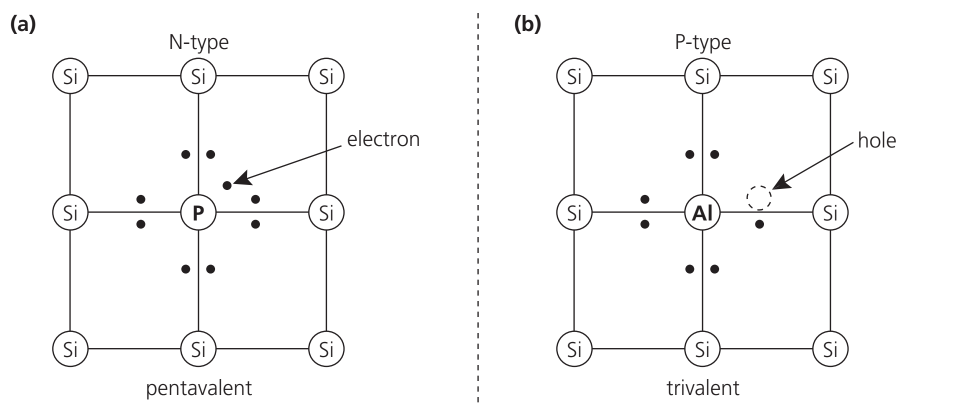


Figure 23.16
Impurity doping: N-type silicon **(a)**; P-type silicon **(b)**.

PN junction and biasing

Semiconductors, via either intrinsic or extrinsic conduction, allow current to flow in either direction through the crystal. When a piece of N-type material is fused with a piece of P-type material in such a way as to provide a continuous crystalline structure, then a device called a

PN junction diode is formed. The PN diode has the very useful characteristic of allowing direct current (DC) to flow in only one direction at a time. It becomes a **solid state rectifier** (Figure 23.17). When a PN diode junction is formed, electrons from the N-type side of the junction diffuse across into the P-type side of the junction and combine with holes. This effect produces a small potential difference, which eventually stops any further diffusion of electrons, resulting in a permanent potential barrier at the site of the PN junction. The zone of the crystal across which the potential barrier forms is called a **depletion layer**. At normal ambient temperatures in a silicon-based PN junction diode this depletion layer represents an effective EMF of about 0.6–0.7 V. This value in any particular diode is dependent on the exact doping conditions in the crystal and will decrease with lower temperatures.

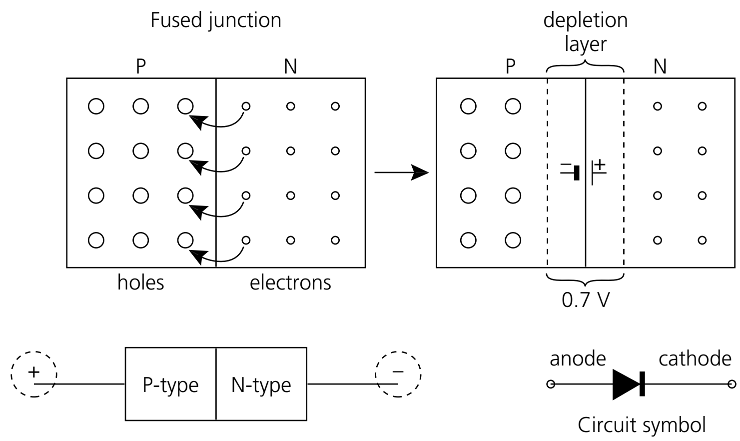


Figure 23.17
A PN junction diode (silicon).

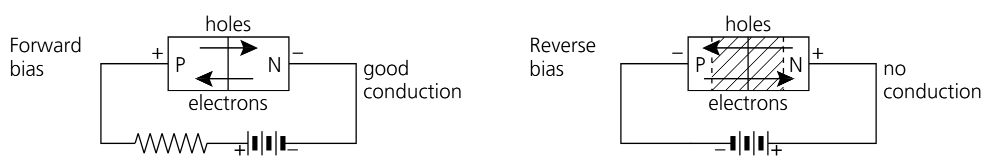


Figure 23.18
Bias conditions for a PN diode.

Consider now what occurs when this PN diode is connected in series with a battery whose EMF is greater than 0.7 V (Figure 23.18). The connecting of a fixed voltage to a semiconductor electronic device in order to set its operating conditions is called setting the **bias**. With **forward biasing**, that is, P-type positive and N-type negative, electrons are forced from the N region into the P region, while holes effectively move in the opposite direction. The depletion layer is effectively eliminated and the diode conducts easily, requiring a current-limiting resistor to avoid rapid overheating. With **reverse biasing**, that is, P-type negative and N-type positive, electrons are attracted away from the junction in the N region, and holes effectively move toward the positive end of the crystal. The depletion layer is effectively widened and the potential barrier becomes equal to the bias voltage. Conductivity is stopped through the crystal and no current at all will flow. It should be noted that if a very high external reverse bias voltage is connected to the PN diode, it may very well break down due to the crystalline structure not being large enough to prevent discharge current flow through the diode. This is very damaging to normal diodes and is called **avalanche breakdown**. Because of this, most diodes have a stated peak inverse voltage (PIV), with power diodes between 400 V and 1000 V and small signal diodes between about 70 and 100 V.

In summary, the conduction characteristics for a normal PN diode are represented graphically in Figure 23.19. A diode is a good conductor when forward biased above 0.7 V and a DC block when reverse biased. Notice that the graph of Figure 23.19 definitely shows that a diode is not an ohmic device in its I–V characteristics because it is not a linear relationship.

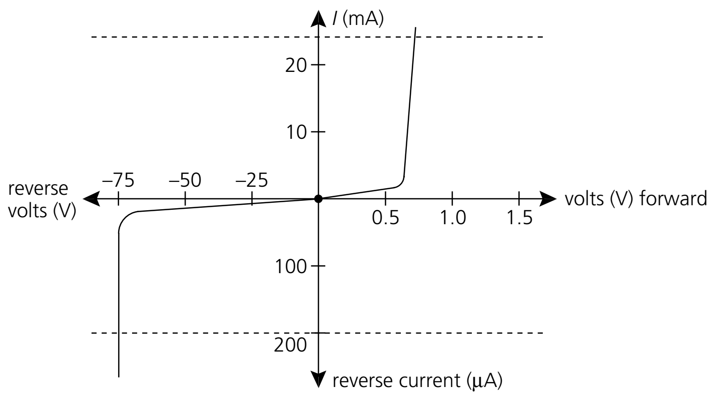


Figure 23.19
Current–voltage (I–V) characteristics of a diode.

Application devices

The PN junction is the basic element in a **silicon solar cell**. Light incident on a PN junction will produce an EMF between the sides of the junction (Figure 23.20). In practice this EMF is about 0.6 V. In an external circuit connected between the two sides of the junction and through a load resistor, R, current flows as long as light is incident on the junction. The current flow depends on the light intensity, as this controls the number of free electron-hole pairs created in the junction. Very pure silicon crystal is needed in solar cell design to minimise possible sites of electron-hole recombination. Low-cost silicon solar cells should make significant contributions to society's future energy production methods.

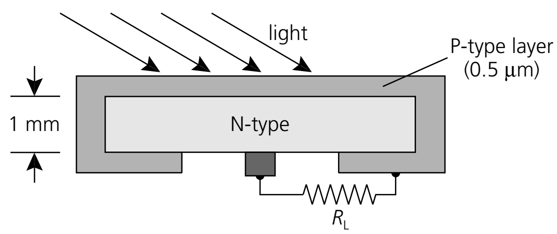
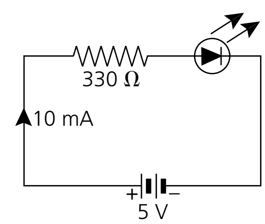


Figure 23.20
A silicon solar cell.

Another common usage of diode semiconductor technology is in a **light emitting diode**, or LED. These are unidirectional current-carrying devices based on the light emitting properties of a PN junction. Electrons give up energy as they recombine with holes in the semiconductor crystal lattice, and in certain types of doped semiconductors this energy is released as light of characteristic wavelengths (Figure 23.21). If the semiconductor is gallium arsenide phosphide (GaAsP), then the junction emits red light, but if the material gallium phosphide (GaP) is used then yellow or green light is produced. A blue emitting LED is based on silicon carbide. A typical red LED requires about 10 mA at a forward voltage bias of 2 V and has a very low PIV of only 5 V or so.

Figure 23.21
A light emitting diode LED.



A diode can also be used to detect light. The reverse current of a PN junction can be increased if the junction is open to incident light photons. A **photo diode** is constructed so that its junction is illuminated from a lens assembly attached to the top of the diode body. Phototransistors are an improvement on the basic design of a photo diode and will be further discussed in the next chapter.

A very useful diode type is actually designed to work in continuous reverse bias mode and is called the **zener diode** (Figure 23.22). These devices are often used as voltage reference sources in power supply and rectification circuits. If a zener diode is forward biased it

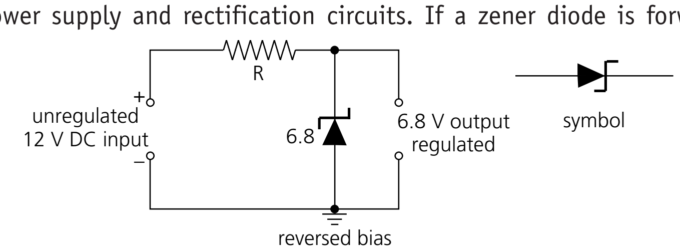


Figure 23.22
A zener diode regulator.

conducts as normal, but it will also conduct when it is connected in reverse bias mode. Zener diodes are designed to conduct in reverse bias over a range from about 2 volts up to several hundred volts. When a zener diode conducts in reverse bias, the voltage drop across the device remains almost constant even when there are wide variations in the current flowing through it. It is this property that makes it useful as a voltage regulator. Note that even a zener diode has a maximum reverse current that it can handle, this being dependent on its power rating.

Finally, note that two further transducers are often used in electronic circuits:

- The **thermistor**, which is a resistor whose resistance decreases considerably with temperature increase and thus can be used to indicate temperature changes electrically.
- The **light dependent resistor** (LDR), whose resistance decreases considerably when it is illuminated by light and thus can be used to indicate light level changes electrically. Sometimes this device is called a photocell.

NEI

Activity 23.3 DIODE RESEARCH

A multitude of semiconductor devices related to the simple PN diode are used in electronics. It should be possible for you to find research material from the library on the following devices and present a short verbal report to your class colleagues on their physical and electronic characteristics:

1. High speed switching or Schottky diodes.
2. Varactors or varicap diodes.
3. Solid state radiation detectors.
4. Laser diodes as used in CD players.
5. Infrared (GaS) diode emitters and detectors.
6. Humidity sensors as used in video cassette recorders.

AC RECTIFICATION 23.6

Using the properties of components so far discussed in this chapter, it is now possible to construct an operating circuit that efficiently converts AC voltages to DC voltages. This process is called **rectification**. The derivation of this word is from the Latin *rectus*, meaning 'straight', which describes exactly what the process does to an AC waveform; it straightens it out. The circuit is one of the commonest blocks in any electronic device, especially if the device needs to be operated from the 240 V AC mains. In most **power supply** units a transformer involving inductor coils steps down the AC voltages from mains 240 V_{RMS} to a much lower value, say, 20 V_{RMS}. You may need to refer to Chapter 26 for the underlying physics of these devices. A rectifying circuit is then used to convert the low voltage AC to a DC voltage. Let's see how this is done.

Half-wave rectifiers

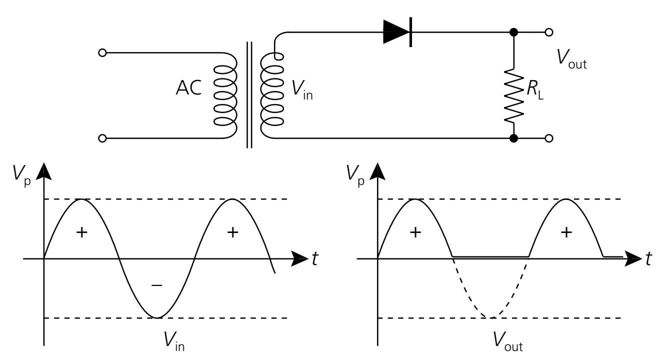


Figure 23.23
Half-wave rectifier circuit.

The diode in the circuit (Figure 23.23) will only conduct when the input sine wave is positive and above 0.7 V in amplitude. It will not conduct when the input sine wave is negative because the diode is in reverse bias. Hence, the voltage output waveform of the circuit is referred to as 'half-wave rectified'. If an oscilloscope (CRO) is used to observe this waveform across the load resistor as shown, its amplitude peak will be:

$$V_{out} = V_{in} - 0.7\text{ V}$$

Full-wave rectifiers

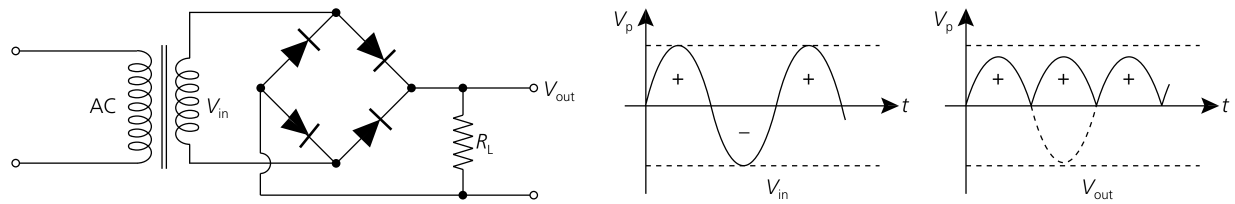


Figure 23.24
Full-wave rectifier circuit.

The four diodes in the circuit shown in Figure 23.24 are arranged in a bridge formation and are referred to as a full-wave rectifier. Full-wave rectification allows the output voltage to be pulsed every half-cycle of the input AC waveform, as the diodes combine in pairs to allow conduction through the load resistor during both positive and negative swings of the input sine wave. Often in power supply design, the four diodes are constructed in one package, with two AC terminals and two DC terminals. This package is called a diode bridge or a bridge rectifier. Because two diodes are involved in each forward conduction cycle, the value in peak amplitude of the output of a full-wave rectifier will be:

$$V_{out} = V_{in} - 1.4\text{ V}$$

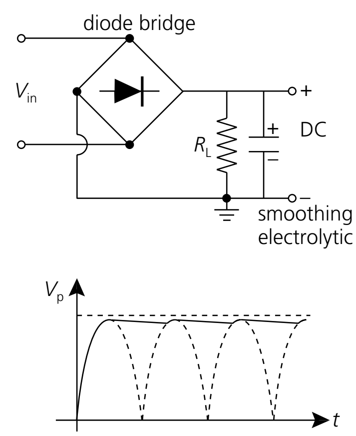
If a DC voltmeter were used at this time to measure the output voltage it would show the average voltage, V_{av}, over two half-cycles, where:

$$V_{av} = \frac{2V_{out}}{\pi}$$

Capacitor smoothing

The output waveforms of both half-wave and full-wave rectifier circuits still contain what is called **AC ripple** or a variation in amplitude with time, and this requires filtering or smoothing out with a large electrolytic capacitor, as shown in Figure 23.25. Recall that the time constant of a discharging capacitor will prevent it from losing charge quickly. This has the effect of holding the voltage output from the rectifier at a high value from one rectified pulse to another. If an oscilloscope is used to observe the output waveform with a large value (400–1000 μF) capacitor in place, then very little remaining AC ripple will be observed, especially if the current drawn from the load resistor is small. The output of these circuits is now **smoothed DC** because it remains constant over time.

Figure 23.25
Effect of a smoothing capacitor.



Voltage regulators

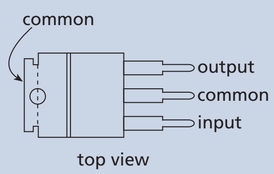
Even though the oscilloscope output waveform of the circuits above looks like smooth DC, we still do not have a fully **regulated DC** power supply. The above circuits are not yet satisfactory for electronic voltage supplies for the following reasons:

- If the load current increases, the voltage available from the transformer will decrease.
- Temperature changes may cause the diodes in the bridge to vary the output voltage.

Ch23-Walding 4th 25/8/04 12:30 PM Page 521

VOLTAGE REGULATOR CHIPS

The three terminal voltage regulator chips in the 78XX and 79XX family make power supply construction simple. These chips only require a filtered DC input voltage. They have internal current-limiting and thermal-shutdown under short-circuiting conditions. A common example is the 7812 positive 12 volt regulator chip in the TO-220 package.



- The mains voltage itself supplying the input transformer may vary, depending on consumer demand and time of day or night.

What is needed is a fully regulated DC supply whose output voltage will remain constant irrespective of these types of changes. The simplest alteration involves the use of a zener diode, which will clamp the output voltage to a particular fixed value, such as 6.8 V. Integrated circuit voltage regulators are a more common device, and will be discussed in Chapter 31.

Questions

10 What is the difference between intrinsic and extrinsic semiconduction?

11 Sketch a circuit that could be used to full-wave rectify an AC voltage of 12 V_P. What would be the DC output voltage of the circuit and how would the AC ripple voltage remaining be reduced?

12 Design a simple application circuit that might use an LDR as part of a voltage divider network controlling a small DC buzzer. The buzzer should go off when ambient light levels fall to a certain point.

Practice questions

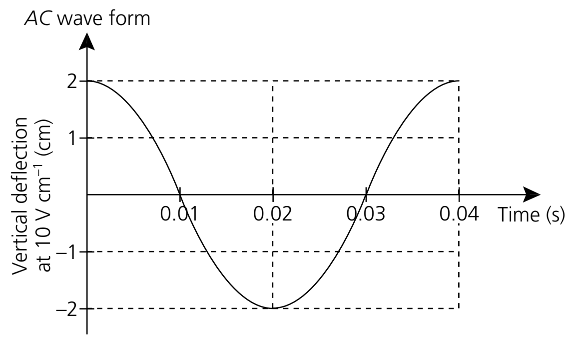
NOTE: Any questions in this set that involve silicon diode conduction will assume 0.7 V as the forward voltage drop.

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***13** Consider Figure 23.26, showing an AC waveform as displayed on an oscilloscope. Determine **(a)** peak voltage V_P; **(b)** peak to peak voltage V_{PP}; **(c)** average voltage V_{av}; **(d)** RMS voltage V_{RMS}; **(e)** frequency of the AC wave.

Figure 23.26
For question 13.



***14** If the alternating current as a function of time in an AC heating coil is given by the expression $I = 5.6 \sin 120\,\pi t$, what is **(a)** the peak current (I_P); **(b)** the RMS current; **(c)** the AC frequency?

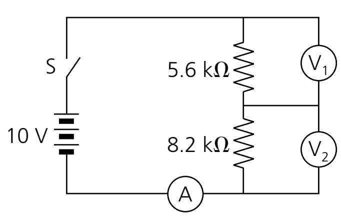
***15** What colour bands would you expect to find on these resistors, assuming a four band system of labelling?

(a) 120 Ω ± 5%
(b) 330 kΩ ± 5%
(c) 5.6 MΩ ± 5%
(d) 2.2 kΩ ± 5%
(e) 1.5 kΩ ± 10%
(f) 470 kΩ ± 10%
(g) 1.8 MΩ ± 10%
(h) 1.0 kΩ ± 10%

***16** Calculate the largest voltage that is possible across a 5.6 kΩ resistor if it is labelled as a 1 W type. What current is flowing through it at this voltage?

***17** The circuit of Figure 23.27 is set up. Determine the readings on the meters shown in the circuit when switch S is closed.

Figure 23.27
For question 17.



***18** Explain the following terms as applied to DC circuits combining resistance, capacitance and inductive components: **(a)** voltage divider; **(b)** self-inductance; **(c)** dielectric; **(d)** millihenry; **(e)** microfarad; **(f)** filter; **(g)** RC time constant; **(h)** electrolytic.

***19** Why should electrolytic capacitors always be connected into the circuit with great care when using DC power supplies?

***20** Determine the time constant for an RC series network where $R = 100\ \text{k}\Omega$ and $C = 470\ \mu\text{F}$. If the applied DC voltage in the network is 9 V, how long will it take for the capacitor to be fully charged and at this point what energy is stored in this component?

***21** Explain, using diagrams where necessary, the following terms applying to semiconductor action in electronics:

(a) N-type doping, and majority and minority charge carriers.
(b) Barrier potential in a silicon diode.
(c) Diode rectification of an AC waveform.
(d) Zener diode regulator.

****22** A silicon-based PN junction diode is forward biased and carrying a current of 4.5 A. What power is it dissipating?

****23** An AC wave of $V_{PP} = 15$ V is fed through an unfiltered full-wave rectifier circuit. What would be the peak output voltage and what reading would be indicated by a DC voltmeter at the rectified output if it in fact reads an average value?

***24** What effect would decreasing the output load resistance have on the ripple voltage of a capacitor-filtered full-wave bridge rectifier circuit?

***25** List some of the practical application devices in electronics that make use of PN junction semiconductor technology. Give a typical use for each of your applications.

****26** The circuit of Figure 23.28 is set up and includes a rectifying diode.

(a) Sketch the waveforms expected at the points labelled A and B in this circuit as displayed on an oscilloscope with vertical amplification set to 5 V cm^{-1}.
(b) Describe how these waveforms would alter if a 470 μF capacitor was connected in parallel with the 1.2 kΩ resistor.

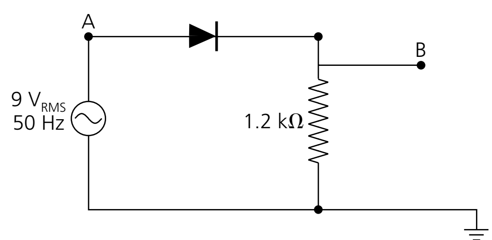


Figure 23.28
For question 26.

****27** A student uses the circuit of Figure 23.29 to obtain the data table below. Plot an I–V curve from the data.

I (mA)	0	0	0.1	4	16	40
V (V)	0	0.2	0.4	0.6	0.7	0.8

(a) What type of semiconductor device is the circuit element X?
(b) What is the switch-on voltage for this device?
(c) What is the resistance of the device at a voltage of 1.0 V?
(d) What current flows if the device is reverse biased by a voltage of 0.8 V?

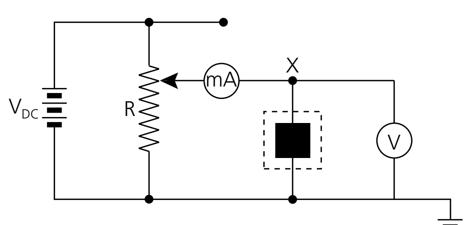
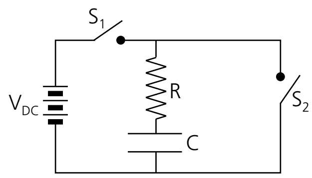


Figure 23.29
For question 27.

****28** Draw a circuit to rectify, smooth and regulate the output voltage from a step-down transformer. Incorporate in your circuit diagram a bridge rectifier, a smoothing capacitor, a zener diode and stabilising load resistance connected across the output. Sketch the output voltage waveform expected from your circuit.

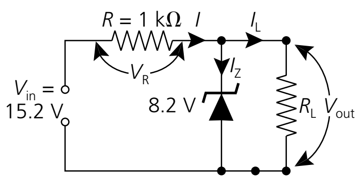
Extension — complex, challenging and novel

Figure 23.30
For question 29.



*****29** Consider the timing circuit of Figure 23.30. Given R = 2.0 kΩ and C = 10 μF,
- **(a)** calculate the time constant of the circuit if switch S_1 is closed;
- **(b)** calculate what happens if another similar capacitor is placed in series with C;
- **(c)** if S_1 is opened and S_2 is closed, describe the current flow in this circuit;
- **(d)** explain why switches S_1 and S_2 should not be closed at the same time.

Figure 23.31
For question 30.



*****30** Analyse the zener diode stabiliser circuit of Figure 23.31.
- **(a)** State a relationship between I, I_L and I_Z.
- **(b)** State a relationship between V_{in}, V_R and V_Z.
- **(c)** Derive an expression for R in terms of V_{in}, V_Z, I_Z and I_L.
- **(d)** Calculate the maximum power dissipated by the 8.2 V zener diode.
- **(e)** What is the load current, I_L, when the zener current I_Z = 3 mA?
- **(f)** What value of R is needed if I_L is to be 10 mA with I_Z = 3 mA?

*****31** Capacitors of 5.0 pF and 20.0 pF respectively are charged so that the respective potential differences between their plates is 200 V and 300 V. They are then connected in parallel maintaining correct charge polarity. Calculate: **(a)** the new charge on each capacitor; **(b)** the common potential difference between plates; **(c)** the energy dissipated during charge rearrangement.

*****32** Light-dependent resistors have a resistance that changes in proportion to the amount of light falling on their windowed surface. The resistance of such a light-dependent resistor was measured over a range of light intensities and the following tabulated data values obtained:

Light intensity in candela (cd)	100	200	300	350	400	500	600	800
Resistance (MΩ)	540	480	410	380	350	290	230	100

Graph these data and develop a conversion formula from resistance to light intensity.

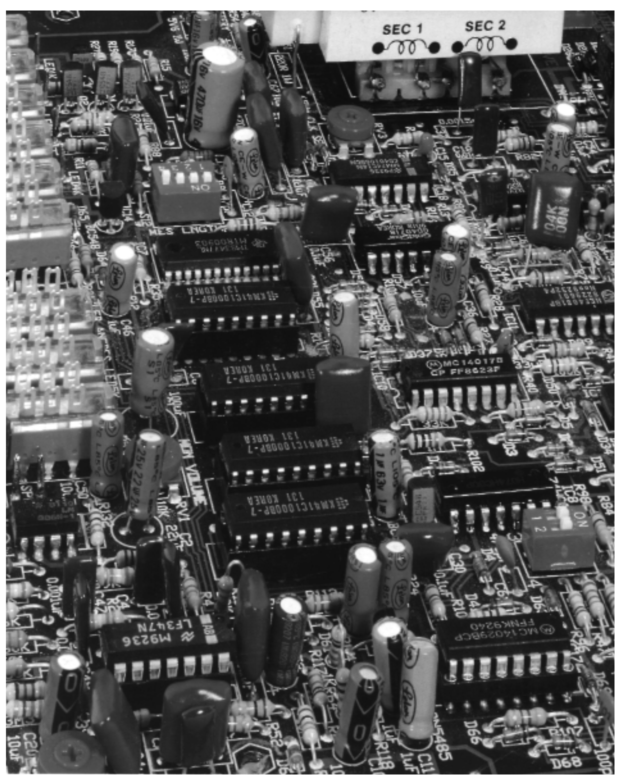


Photo 23.7
Complex electronic circuit system involving components discussed in Chapters 23 and 24.

CHAPTER 24

Electronic Systems

24.1 INTRODUCTION

Semiconductor devices are able to do much more than rectify AC voltages. When semiconductors replaced the older vacuum-tube valve devices in electronic applications, the circuits were termed 'transistorised' or 'miniaturised' since the transistors were so robust and small compared with the delicate and bulky thermionic vacuum-tube valves that had been the mainstay of electronics until then. It has been said that if the car industry progressed at the same rate as the electronics industry, a Rolls Royce would be the size of a matchbox, cost fifty cents and get a million kilometres per litre!

In this chapter we will look at the development of electronic systems involving input–output transducers such as microphones, loudspeakers and motors as well as transistors and **integrated circuits** (ICs). The resistance of semiconductor material is dependent on the level of impurity doping. The capacitance of a reverse bias PN diode is altered by the voltage applied to the base–emitter junction of a transistor. These properties make it possible to combine arrays of resistors, capacitors and transistors onto a single piece (or **chip**) of silicon, which ultimately becomes an integrated circuit. These ICs can be mass-produced to accomplish any desired electronic function. Today, there exist hundreds of integrated circuit families. Very large scale integration (VLSI) circuits and very high speed integration (VHSI) circuits in recent years have seen tremendous improvements in device reliability, performance speed and lowering of cost within electronic systems such as audio, video, telephone and computer technology. There is hardly a domestic or industrial machine existing today that is not, in part at least, an electronic system.

In 1969 the American corporation Intel developed a range of medium-sized ICs for use in hand-held calculators. An Intel engineer, Edward Hoff, realised that it was simpler to design a single-purpose IC for all calculators and to make this chip 'programmable' so that it could be controlled by a unique external chip containing a specific set of instructions. This general-purpose chip was called a **microprocessor**, or single chip central processing unit (CPU). The microprocessor has become the heart of many modern electronic systems. The chip can perform a vast range of functions depending on the external instructions given to it. The modern personal computer is nothing more than a microprocessor joined to an array of external chips such as memory and the basic input–output system as well as keyboard input and video output circuitry.

24.2 INPUT–OUTPUT TRANSDUCERS

To produce the control functions carried out by circuits in electronics, **electrical signals** (the actual voltages and currents), usually, are manipulated by components in various ways. Electrical information is obtained as voltages and currents from the input to a particular circuit or circuit system. For example, a radio receives small electrical voltages (a radio

Ch24-Walding 4th 25/8/04 12:32 PM Page 525

signal) from the input antenna, or a public address system uses an input signal produced by a microphone. The circuit performs some function such as amplification and this new electrical information is available to the electronic system output. This output may be a loud-speaker or a visual display such as an LED array or a video screen. An electronic system allows us to control electrical information and usually allows several stages of electrical energy conversions to take place. Thus, an electronic system consists of a defined set of blocks of circuitry, as shown in Figure 24.1. This is called a schematic or an electronic **block diagram**.

Figure 24.1
Block diagram of an electronic system.

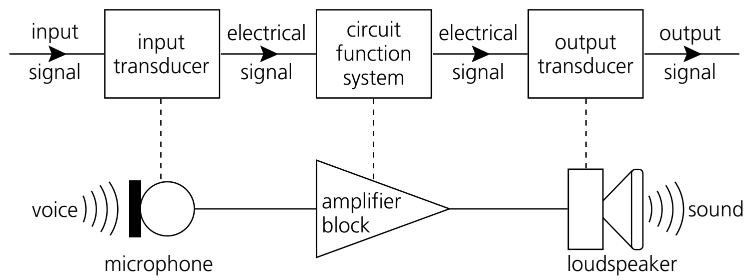
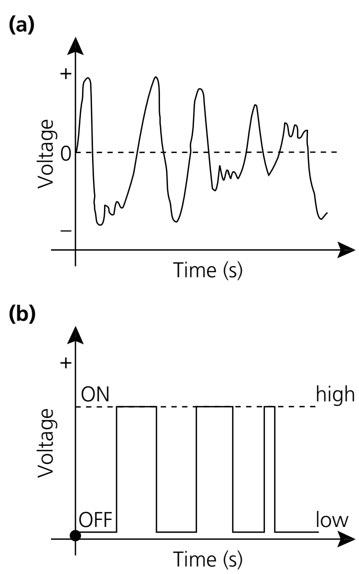


Table 24.1 TRANSDUCERS IN ELECTRONICS

INPUT TRANSDUCERS	OUTPUT TRANSDUCERS
Microphone	loudspeaker
Antenna	light bulb
Thermocouple	LED and display
Photocell	relay coil
LDR	electric motor
Thermistor	cathode ray tube
Laser diode	audio–video heads

Transducers are devices that convert energy from one form to another in electronic systems, usually involving electrical energy. Table 24.1 lists commonly used input and output transducers in electronics. Occasionally an input transducer can be the same as an output transducer. For example, in some simple intercom systems the loudspeaker can be both the input microphone as well as the output speaker, although, of course, not at the same time. For instance, the two functions may be controlled by a push-to-talk switch.

Figure 24.2
Electrical signals: analog **(a)**; digital **(b)**.



Analog to digital conversion (ADC)

Electronics is concerned with the control of electrical signals that carry information to devices that interpret the signals and perform a particular function. Block diagrams are often used to get an overall picture of the various functions performed within the system as a result of the electrical signals being interpreted. Two types of electrical signals are commonly found in modern electronic systems. These are **analog** or continuously varying electrical voltage signals, and **digital** or electrical voltage signals that are either ON or OFF. If a digital signal is ON it means that the voltage level is high or equal to the circuit supply voltage whereas if a digital signal is OFF then the voltage level is low or at zero volts.

Figure 24.2 illustrates the nature of these two different types of signals graphically.

Analog signals are widely used in audio, video and television systems, while digital signals are used in computers and microprocessor- or microcontroller-based consumer devices. Digital electronics is a well-established field, with many older analog devices now phased out and replaced by digital processing; examples include the digital DVD and audio CD revolution, and mobile phone technology. Underlying these processes is a technique called 'analog to digital conversion' or **ADC**. The reverse process, digital to analog conversion or DAC, is equally important.

The input to an ADC consists of a voltage that varies among a theoretically infinite number of values. Examples are sine waves, the waveforms representing human speech, and the signals from a conventional television camera. The output of the ADC has defined levels or states. The number of states is almost always a power of two — that is, 2, 4, 8, 16, etc. The simplest digital signals have only two states and are called **binary**. All whole numbers can be represented in binary form as strings of ones and zeros.

Information is stored in a computer as groups of bits. A **bit** stands for a binary digit, 0 or 1. The only practical way of representing these two states in an electronic circuit such as a computer is to use two-state logic, or ON and OFF. 'OFF' represents logic 0, and 'ON' represents logic 1. In electrical terms, for most digital logic circuits, 0 volts represents logic 0 and 5 volts represents logic 1. Due to the ever-decreasing power consumption and increasing speed of modern digital circuits, the logic 1 voltage is decreasing to around 3 volts. In general, microprocessors use bits in groups of eight, which are called **bytes**. Groups of four bits are also used, and these are called **nibbles**. There are two nibbles in a byte.

The maximum number that a single byte can hold equals 255, and there are 256 different combinations of binary numbers, including zero, that can be represented. In the binary system, each binary digit or bit represents a power of 2. In the decimal number system, each digit represents a power of 10.

Example. 0 0 0 0 1 1 0 1 in binary (from right to left) represents;

$1 \times 2^0 = 1 \times 1 = 1$

$0 \times 2^1 = 0 \times 2 = 0$

$1 \times 2^2 = 1 \times 4 = 4$

$1 \times 2^3 = 1 \times 8 = 8$

Add these up to get a decimal equivalent $1 + 0 + 4 + 8 = 13$.

In the hexadecimal system, which is used in coding or programming computers, each digit represents a power of 16. There are no numbers past the digit 9, so the letters A to F are used to represent the digits 10 to 15.

Example: 0 1 2 3 4 5 6 7 8 9 A B C D E F These hex digits fit very nicely into a nibble.

0 = 0000 = 0 decimal

1 = 0001 = 1 decimal

2 = 0010 = 2 decimal, etc.

down to the hexadecimal value

F = 1111 = 15 decimal.

As two nibbles fit into a byte, there are also two hexadecimal numbers that fit into a byte. It becomes easy to break down large binary numbers into more something more manageable by splitting them into nibbles and into hexadecimal numbers.

Let's take a quick look at the process of ADC. Refer back to the section of Chapter 16 on modern sound technology (Section 16.10) for audio digital devices that make use of ADC.

In analog technology a simple waveform is recorded and used in its original form. For example, in a cassette tape recorder the varying voltage wave output from a microphone is applied directly through the magnetic recording head onto the magnetic tape. (Refer to Section 25.4.) The pickup head takes the analog signal back off the tape and sends it to the amplifier and speakers. Typical audio frequencies range from about 20 Hz up to about 20 kHz, which represents the music fidelity.

In digital technology a process of sampling the analog wave at a fixed interval is used. The amplitude of the sampled wave section produces a voltage that is converted into a number that is stored in the digital device such as a music or data CD. The sampling rate used in normal music CD recorders is 44.1 kHz or 44 100 numbers per second of music, while that used in DVD audio discs is 192 kHz. When an ADC sampling recording is made, engineers have control over two factors (Refer to Figure 24.3.):

- the **sampling rate** — how many samples are taken per unit time
- the **sampling precision** — the number of different gradations in amplitude that are used when sampling.

NOVEL CHALLENGE

'It is said that there are 10 types of people in the world: those who understand binary and those who don't.' Explain.

Ch24-Walding 4th 25/8/04 12:33 PM Page 527

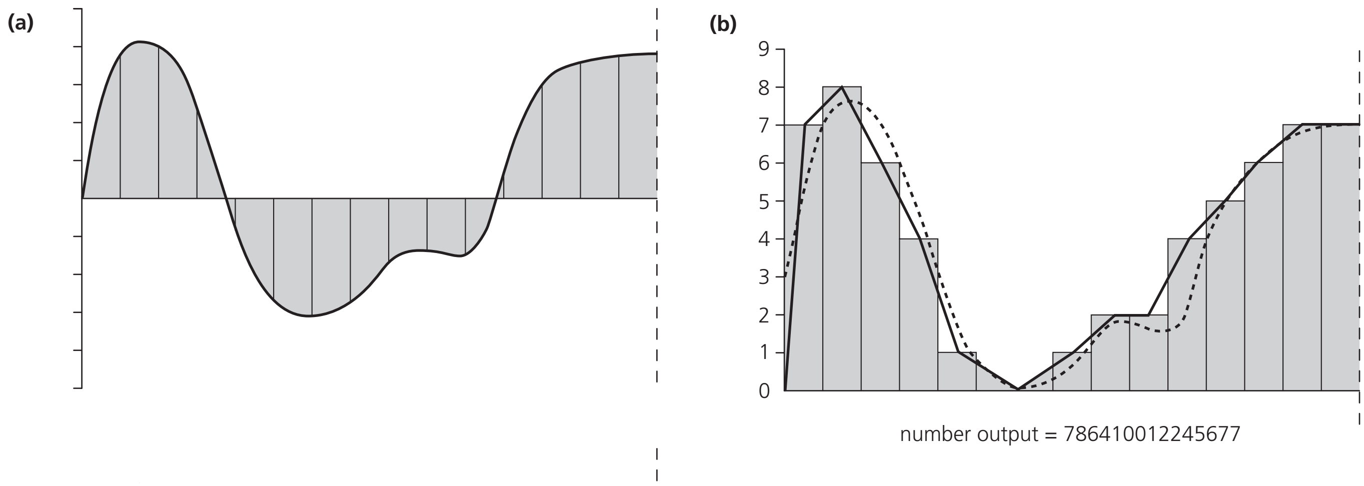


Figure 24.3
Analog to digital conversion (ADC):
(a) original analog waveform;
(b) sampled digital waveform.

You can see that in this diagram the horizontal time base might represent blocks at every one-thousandth of a second with an amplitude precision of 10 units. The shaded rectangles represent samples, with the ADC looking at the wave and assigning a closest decimal number value of between 1 and 9. These decimal numbers are converted to binary form. The ADC thus outputs a string of numbers in succession (a digital word) and produces a digital waveform with an obvious sampling error compared to the original analog waveform. However, as the sampling rate and precision increases (the number of shaded rectangles increases dramatically), the difference between the digital and analog waveforms reduces to nothing. In fact digital waveforms at high sampling rates are at a much higher fidelity.

The actual binary output from the ADC chip can be produced by a variety of electronic methods. The successive-approximation ADC is one of the most commonly used designs. This requires only a single comparator but will be only as good as the DAC used in the circuit.

Figure 24.4
ADC chip block diagram.

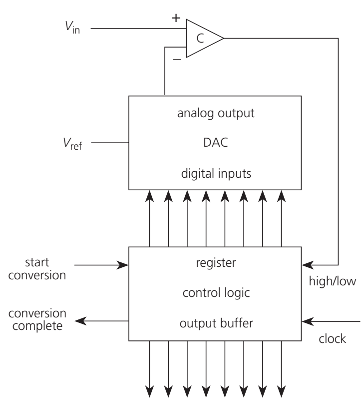


Figure 24.4 is the block diagram of an 8-bit successive-approximation ADC. The analog output of a high-speed DAC is compared against the analog input signal. The digital result of the comparison is used to control the contents of a digital buffer that both drives the DAC and provides the digital output word. As examples in 8-bit binary the decimal 7 = 00000111, 5 = 0000101, 2 = 0000010. You might like to check some websites that show you how to convert between decimal and binary. Note that modern digital ADC and DAC chips can have resolutions up to 18 bits and sampling rates up to 1.5 GHz.

High sampling rates and precisions produce large amounts of digital data in the form of number strings. On a CD, digital numbers are stored by the ADC as word bytes and it takes 2 bytes to represent 65 536 amplitude precision gradations or sample. On a normal stereo (2-channel) music CD which may hold up to 74 minutes of music data there will be about 780 megabytes of digital number storage. You should test this by doing the calculations — 2 bytes per sample at 44.1 kHz for 74 minutes!

The surface of a CD contains one long spiral track of data, which may be up to 5.0 km long. This track is made up of flat laser-light-reflective areas representing digital binary 1, as well as non-reflective bumps representing digital binary 0. This pattern of 1s and 0s is read by the scanning laser head and back-converted by the DAC, first into the digital waveform representing the original precision graduations at the sampling rate, and finally into the analog waveform.

One of the best advantages of digital recording is that the quality does not deteriorate over time like magnetic analog recordings. As long as the laser head can read the number string the information can be decoded. Today error correction techniques as well as number group compression techniques have allowed a tremendous amount of material to be recorded onto a disc. This is evident in multi-layer DVDs which can contain many hours of compressed audio and video information using MPEG technology. Many channels of digital audio are possible to give a finely tuned sound field from multiple speaker systems. Again refer to Section 16.10.

Let us now look at the basic electronic devices that underlie both analog and digital electronic processing systems. After all, to begin to understand complex systems, we need to be able to break them down into their component parts.

Ch24-Walding 4th 25/8/04 12:33 PM Page 528

24.3 TRANSISTORS

The transistor is one of the most useful devices in electronics. The transistor, just like diodes, can change the form of electrical signals it receives and is therefore called an **active device**. The first transistor was developed by William Shockley, John Bardeen and Walter Brattain at the American Bell laboratories in New Jersey in 1947. This solid state device 'transferred' a current across a high-resistance material, so they called it a 'transfer-resistor' or transistor for short. This first point-contact transistor device had several limitations, including noisy amplification, low power handling capability and limited applicability because of its delicacy. William Shockley had also conceived the idea of the junction transistor, which was free of many of these defects and limitations. Today, most transistors are made of the junction type. Although they are being replaced in circuitry blocks by integrated circuits, the transistor is still essential in many applications and is manufactured in a wide variety of shapes and sizes.

Photo 24.1
Various transistors.

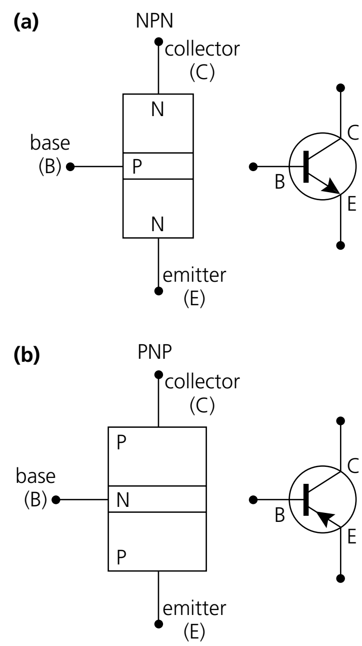
Transistor structure

A **bipolar junction transistor** consists of two PN junctions joined together, as shown in Figure 24.5. This is a bit like two diodes placed back to back. You might like to refer back to Chapter 23 for revision of semiconductor diodes. Two basic types are possible and illustrated as PNP or NPN, together with their symbols showing the direction of conventional current flow through the device. The transistor is called a bipolar device because current flows through the transistor using two different modes. In N-type silicon, current flows as mostly electrons, while in P-type silicon, current flows as holes. The three layers of the transistor are:

- **collector** (C), which in the case of an NPN-type is constructed as a lightly doped N-type silicon layer
- **base** (B), which is a very thin layer of lightly doped P-type silicon
- **emitter** (E), which is a heavily doped N-type silicon layer.

Notice the difference in the thickness of these layers from those of standard diodes. The transistor device is called bipolar because both majority and minority charge carriers are used during conduction. In general, the analysis of an NPN transistor in terms of current flow and bias voltage conditions can be reversed for a similar analysis of a PNP transistor. It is common in analysing the operation of a transistor to discuss a flow of electron current through the device; however, this is always precisely opposite to normal conventional current flow, which is used in most diagrams. Figure 24.6 shows the typical structure of an actual transistor constructed using a crystal of silicon on which layers are formed in a process called the planar epitaxial technique. The centre layer of a transistor, called the base, is always thinner than the outer emitter and collector layers. This is essential for the correct operation of the transistor and is the reason why two simple PN diodes cannot be used back to back to produce the same effect as a transistor in circuits.

Figure 24.5
Transistors: NPN **(a)**; PNP **(b)**.



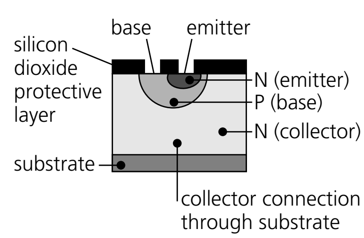
In any practical transistor package, the very small piece of semiconductor is usually protected in an epoxy plastic housing and three connecting wire 'pigtails' or pins are attached to the three terminals, C, B and E. Manufacturers provide transistor pin-out diagrams to enable electrical connections to be correctly made to different types of transistor packages such as the common TO-92 pack, TO-220 pack and the TO-3 pack. It is important that correct pin connections are made in transistor circuits, as damage can easily occur if the power supply connections are reversed, for instance. You can refer to various electronics component catalogues for examples of these pin-out diagrams.

Transistors are used in electronic circuits in three ways, basically:

- as direct current (DC) amplifiers
- as fast acting electronic switches
- as AC voltage amplifiers.

We will now examine the first two of these ways, together with some practical circuit applications; the AC voltage amplifier will be left to Chapter 31.

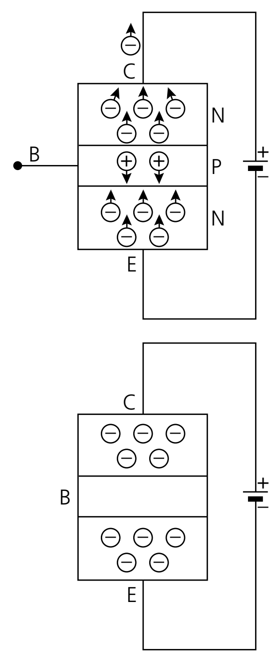
Figure 24.6
Transistor construction in cross-section.



Ch24-Walding 4th 25/8/04 12:33 PM Page 529

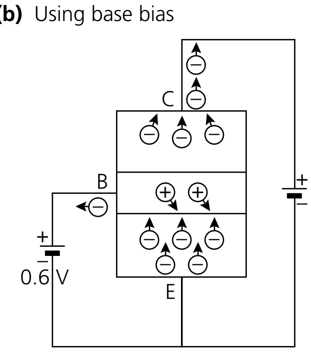
Figure 24.7
Transistor current amplifier: at switch-on and equilibrium **(a)**; using base bias **(b)**.

(a) At switch-on and equilibrium



no current—
transistor 'OFF'

(b) Using base bias



transistor 'ON'—
base current

Direct current (DC) amplifiers

Let us consider what happens when an NPN transistor is connected into a circuit, such as in Figure 24.7**(a)**. When the voltage is applied free electrons in the N regions will tend to move from emitter to collector and free holes in the P region will move towards the emitter. These few holes moving against the electrons will be filled in by electrons in a short time and will cease to exist. This produces an effective potential barrier at the base layer and will stop any further current flow. In order to allow electron current to continue to flow from emitter to collector, the base layer requires a positive voltage. (See Figure 24.7**(b)**.) When this positive base voltage rises above 0.6 V, majority charge carriers begin to cross the emitter–base junction. Because the base is only lightly doped while the emitter is heavily doped, there will be many more electrons coming from the emitter than there will be holes coming from the base. Also, because the base region is very thin and lightly doped, most of the electrons avoid falling into holes in the base region. A continuous flow of electrons through the device is established by maintaining a positive voltage of 0.6 V at the base. Thus, electrons are 'emitted' from the emitter layer and 'collected' by the collector layer. Typically, only about 1% of the emitted electrons will fall into holes generated by the positive base and constitute the base current, I_B, while the remaining 99% of electrons pass from the emitter to the collector to form the collector current, I_C. Thus, the collector current is 99 times the base current and the transistor has produced direct **current amplification**.

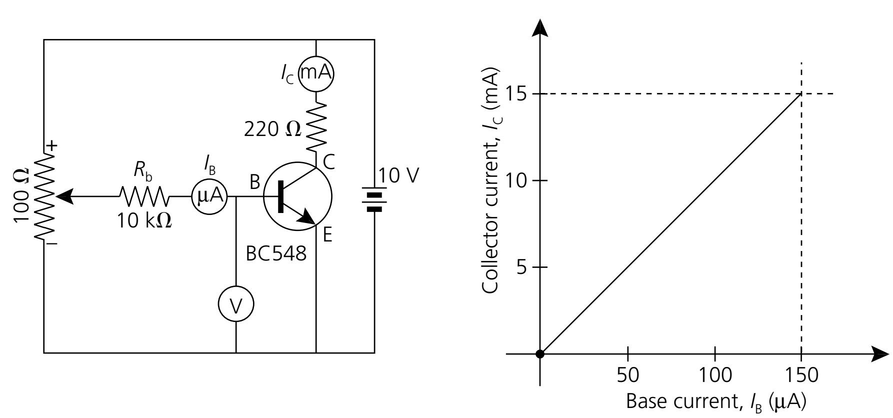
The ratio of the collector current to the base current is known as the current amplification factor or **current gain** of the transistor and is denoted by the symbol β (Greek letter beta) or h_{fe}. It varies for different transistors but usually lies in the range between 20 and 800.

$$\beta = h_{fe} = \frac{I_C}{I_B}$$

Since small variations in base current are controlling larger variations in collector current, the transistor circuit involving base and emitter is often called the '**control circuit**' and the collector–emitter circuit is called the '**working circuit**'. If the base–emitter junction is forward biased, current will pass through both control and working circuits as long as the collector is positive with respect to the emitter. Small changes in the base current can produce large changes in the working current (NPN current amplification).

Figure 24.8 illustrates an actual test circuit that may be used to show current amplification by measuring I_B and I_C. Note that a voltage divider potentiometer is used to control base–emitter forward bias. The graph of output data I_C versus I_B can be used to derive the current gain, β, of the device being tested. Note also that a collector resistor is used in the circuit, which limits the maximum size of the collector current in order not to overheat and destroy the transistor.

Figure 24.8
Testing current gain, β.



Ch24-Walding 4th 25/8/04 12:33 PM Page 530

A special type of transistor designed for high power levels and high current gain is called the **darlington**. It is essentially two bipolar transistors back to back on the same piece of silicon. The darlington pair has a switch-on voltage of about double that of a normal transistor, a DC gain value of about 1000, and if necessary can be made to handle up to about 5 A of current.

Example

Consider the circuit and graph of Figure 24.8.

(a) What voltage should register on the base–emitter voltmeter when a small collector current I_C begins to be measured?

(b) Use the graph of data supplied to calculate the current gain of the BC548 transistor in this circuit.

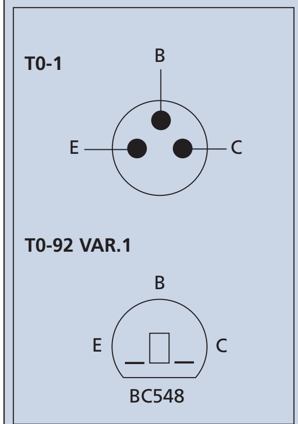
Solution

(a) Milliammeter, I_C, will not show any working current until V_{BE} reads at least 0.6 V as this is the switch-on voltage for the transistor.

(b) $$\beta = \frac{I_C}{I_B} = \frac{15\ \text{mA}}{150\ \mu\text{A}} = 100$$

The current gain has a value of 100.

TRANSISTOR OUTLINES BOTTOM VIEW



Transistor switches

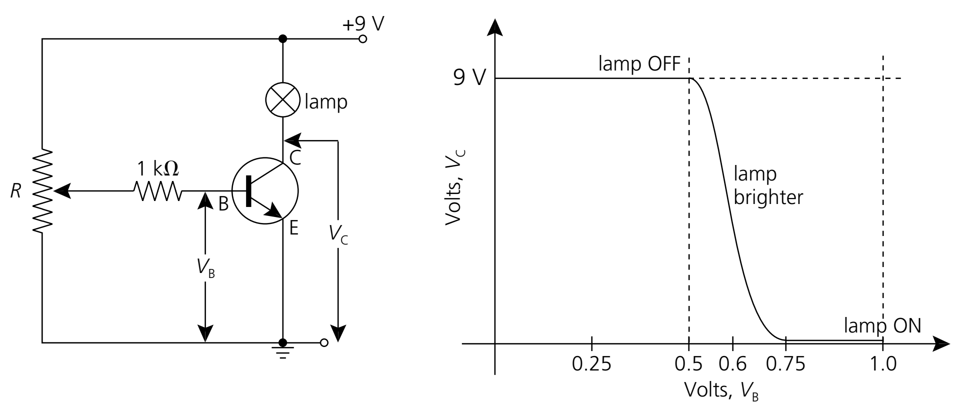


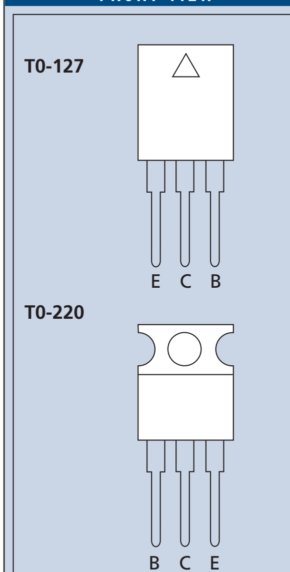
Figure 24.9
Transistor switch action.

The switching action of a transistor is produced by varying the voltage applied to the base so that the transistor is either turned ON (large collector–emitter current flowing) or turned OFF (no collector–emitter current flowing). Consider the circuit of Figure 24.9, showing a lamp in the working circuit. A potentiometer voltage divider is used to control and vary the voltage applied to the transistor's base. V_B and V_C are the voltages as measured at the base and collector with respect to the earth. The graph shows how the lamp's brightness is controlled as the potentiometer is varied. The analysis is as follows.

As the base voltage rises from zero, the lamp remains off, because the collector voltage is at 9 V so there is no potential difference across the lamp. The transistor is said to be turned OFF. It acts like an open switch (collector–emitter path). When the base voltage reaches 0.6 V (for a silicon transistor), the base current turns the transistor ON and it starts conducting. It takes only a small rise in base voltage, to about 0.75 V, until the transistor is fully ON or fully conducting. The transistor has a very low collector–emitter path resistance and it acts like a closed switch. The transistor's collector is now at a very low voltage, hence the full 9 V is applied across the lamp. It is turned ON and produces full brightness. Lowering the base voltage to below 0.6 V will turn the lamp off again.

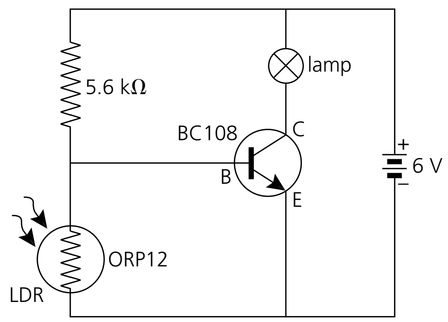
Let's look at three more practical circuits making use of switching action.

TRANSISTOR OUTLINES FRONT VIEW



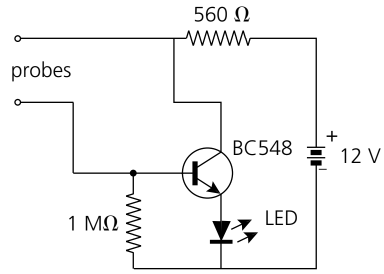
- Figure 24.10 shows an LDR or **light dependent resistor** controlling the base voltage. In bright light conditions the LDR has a low resistance, which means that the base voltage is held low, the transistor is OFF, its collector voltage is high and the lamp is turned off. In dark conditions the LDR has a high resistance, which means that the base voltage is held high, the transistor is ON and conducts, causing the collector voltage to go low and the large potential difference across the lamp causes it to switch on.

Figure 24.10
LDR switching circuit — automatic light switch.



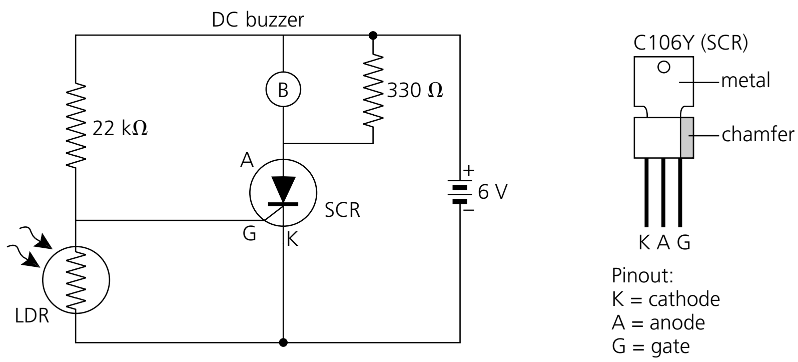
- Figure 24.11 shows a transistor switching an LED as a result of two circuit points being connected together with a resistance such as human skin. If a small base current is produced by holding the probes with opposite hands, the transistor is correctly biased and it switches on. This allows a collector–emitter current to flow through the LED and it illuminates. The LED will turn off again if the touch probes are released, because of the open circuit or very high resistance between collector and base.

Figure 24.11
Transistor touch switch.



- Figure 24.12 is a light-controlled high current switch, which might be used as a burglar alarm. The circuit contains both an LDR and a **silicon controlled rectifier** (SCR) device or **thyristor**, which differs from a transistor in that it not only can carry larger currents but once switched ON cannot be turned OFF by removing the voltage at the base terminal or gate, G. This circuit would operate as an alarm by ensuring that a bright light beam is shone onto the LDR, which would provide a low voltage at the SCR gate to keep it turned OFF. This keeps the buzzer turned OFF because the SCR anode is held high. If a burglar breaks the light beam, the LDR now has a high resistance, causing the gate to trigger, or turn ON, the SCR, and the buzzer would sound. The action of the SCR will keep the buzzer turned ON until such time as the power supply is disconnected.

Figure 24.12
Light controlled high current switch (SCR).



Activity 24.1 CHANGING THE CIRCUITS

Consider how to modify the circuits already presented to do slightly different jobs. For instance, how would you modify the following circuits?

1 Modify the circuit of Figure 24.10 to trigger when a light is turned ON rather than OFF, as, for example, in a circuit to automatically turn on the garage lights when the car is driven in at night.

2 Modify the circuit of Figure 24.12 to act as a fire alarm buzzer, using a thermistor sensor.

Before leaving the various transistor circuits to look at other semiconductor devices, let's consider the special type of transistor called a **field effect transistor** or FET. Remember that in a junction transistor the base current is needed to remove excess electrons in the base region to allow a larger working current to flow from collector to emitter. In some circuits even this small base current can provide difficulties. Consider Figure 24.13. In the FET device a current flowing from the 'drain' (collector) through the channel to the source (emitter) is controlled by an electric field produced by charges present on the gate (base). Only an extremely small electric field strength is necessary, provided by a very small gate current and this produces an extremely high effective input resistance for the device. FETs are used in amplifier circuits where large voltages are controlled by extremely small voltages at the gate. Such a situation might arise when trying to amplify the very weak electrical signals produced by the human body during muscle or nerve activity in medical diagnostic equipment such as an electrocardiograph.

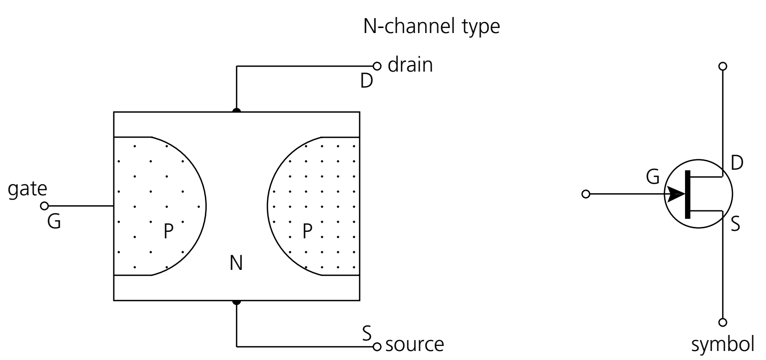


Figure 24.13
Field effect transistor (FET) — N-channel type.

Activity 24.2 SPECIAL TRANSISTORS

Transistor design has come a long way since the first bipolar junction type. Some further types of transistor you might like to research and present a report on are listed below. Find a common application circuit for each type:

1 Junction field effect transistor JFET.
2 MOSFET.
3 Power MOSFET.
4 Phototransistor MEL-12.

Questions

1 What is the difference between electronic analog and digital voltage signals? Give an example of electronic systems that interpret voltages of both types.

2 Explain what is meant by each of the following terms: **(a)** P-type semiconductor; **(b)** NPN transistor emitter; **(c)** NPN transistor base; **(d)** PNP transistor collector; **(e)** current value I_B.

3 An operating transistor has the following parameters: $\beta = 200$, $I_B = 15\ \mu A$. What are the values of I_C and I_E?

4 Figure 24.14 shows the schematic diagram for an NPN transistor in normal operation.

(a) Redraw this diagram showing a normal symbol for the transistor. Label E, B and C.

(b) Add the voltages X and Y, showing correct polarity.

(c) Show how the voltages X and Y can be obtained in practice from a single supply voltage V_{CC}.

Figure 24.14
For question 4.

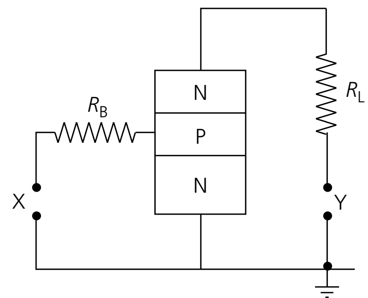
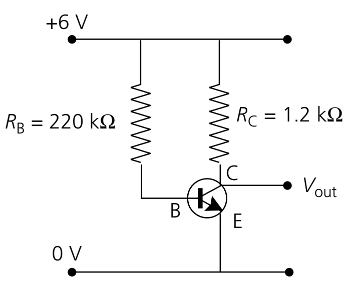


Figure 24.15
For question 6.

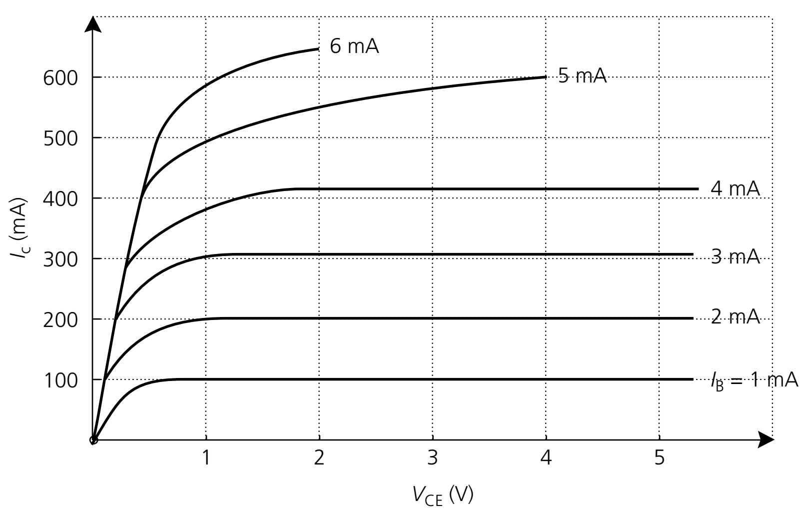


5 Explain the operating conditions that are necessary to make a transistor function as a switch.

6 In the circuit of Figure 24.15, the current gain of the transistor is 150. Calculate the value of the base current and the output DC voltage at the collector. Use the value $V_{BE} = 0.7$ V.

7 Figure 24.16 shows graphically the collector characteristics I_C/V_{CE} for a particular transistor. Calculate the transistor's current gain, β, when the base current is 2.5 mA and the collector–emitter voltage is 3.0 V.

Figure 24.16
For question 7.



Collector characteristics

24.4 INTEGRATED CIRCUITS

During the 1960s and 1970s advancements in semiconductor technology made it possible to combine larger numbers of active devices like transistors and diodes, as well as other passive components, onto a silicon chip. It has become possible to produce complete functional system circuits, involving both analog and digital processes, within a single integrated circuit package. The original integrated circuit concept was presented to engineers at an Institute of Radio Engineers symposium in Washington DC on 5 May 1952 in a paper by G. W. Drummer entitled 'Electronic components in Great Britain'. In this paper he stated:

> At this stage, I would like to take a peep into the future. With the advent of the transistor and the work into semiconductors generally, it seems now possible to envisage electronic equipment in a solid block with no connecting wires. The block may consist of layers of insulating, conducting, rectifying and amplifying materials, the electrical functions being connected directly by cutting out areas of the various layers.

The first microelectronics integrated circuit patent was filed on 6 February 1959 to a J. S. Kilby as US Patent No. 3,138,743, with the following statement:

> It is therefore, a principal object of this invention to provide a novel miniaturised electronic circuit fabricated from the body of semiconductor material containing a diffused P–N junction wherein all components of the electronic circuit are completely integrated into the body of semiconductor material.

Such integrated circuit (IC) packages represent the miniaturisation of electronics with consequent increase in speed and reliability of operation, as well as an overall massive reduction in costs of manufacture. An integrated circuit package may contain the equivalent of thousands of transistors and the most common method of construction is similar to the silicon planar method described for transistors. The IC fabrication technique is very complex and expensive as it involves etching layers of semiconductor, multiple photographic exposure through light-resistive masks and numerous steps of subsequent metallic vacuum deposition to form the transistor blocks and conducting pathways. It is important to realise that integrated circuits have been the single most important factor in the development of modern electronic systems.

Integrated circuits are mass-produced with variations in the type of basic transistor building elements within the chip design itself. These types are commonly referred to as **integrated circuit series**, such as the 74 series TTL chips and the 74LS series low power Schottky chips, both requiring power supply voltages of 5 V. More versatile are the 74C series CMOS chips (3–15 V), the 74HC series high speed CMOS chips (2–6 V) and the 4000 series CMOS chips (3–15 V). All these CMOS chips involve metal oxide semiconductor material and use field effect transistor (FET) construction. A disadvantage of all CMOS type ICs is that they are easily damaged by stray electrostatic charge and therefore must usually be handled taking special precautions.

The only component that cannot be integrated onto a silicon chip IC package is the electromagnet coil or inductor. Inductors need to be added externally to the chip in circuit construction, although it is possible to simulate their behaviour with a combination of passive capacitors and other active devices.

Integrated circuits generally fall into two categories, analog IC or **linear IC** devices, and **digital IC** devices. Analog ICs are used for processing analog voltage signals by such operations as amplifying, adding, multiplying, filtering and other more complex functions. The input and output voltages of these devices can have any value between the power supply voltages driving the device and there is a direct or linear functional relationship between the input and the output. Digital ICs are used to process **binary signals** where the input and output voltages of the device exist at only two states, called ON and OFF, or HIGH and LOW. Binary voltage states are often represented mathematically as a 1 or a 0: 1 = HIGH (ON); 0 = LOW (OFF).

PHYSICS UPDATE

Ever since 1999 the microchip giant IBM has continually perfected new forms of silicon–germanium chip processing technology. This technology involves the use of a combination of germanium atoms embedded in a silicon substrate, allowing much faster current conduction through the crystal and much smaller chip size. The heart of IBM's SiGe technology is a heterojunction bipolar transistor (HBT) doped with germanium to increase the electron transfer.

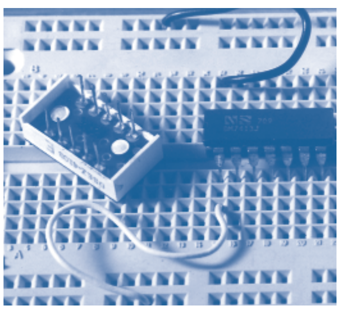
Today's 2 GHz microprocessors could be boosted to 50 GHz or more using the latest silicon–germanium technology. With over 1000 microelectronic patents in 2002 alone, IBM offers a range of industry-standard CMOS, RFCMOS, and silicon–germanium Bi-CMOS process technologies ranging in transistor size from 0.5 μm to 90 nm. Uses for the latest SiGe devices include wireless Bluetooth component chips, wireless LAN and global positioning systems chips, multi-action mobile telephone systems, and optical networking components.

The transistor operation underlying these digital ICs is a rapid switching action as described in the previous section. Digital IC circuits usually require fewer external discrete components than linear IC circuits and are tending in recent times to replace them in a lot of electronic systems. These digital IC application circuits will be discussed in more detail in Chapter 31.

Once the integrated circuit is fabricated, it can be encapsulated into a protective cover and connecting pins added so that it may be connected into a circuit with appropriate external components such as resistors, capacitors, diodes and transistors. A very common IC package is the **dual in line** (DIL) type. It was first fostered by Bryant Rogers as a dual in-line package (DIP) while at Fairchild Semiconductor in 1964. This type uses an epoxy plastic case with the IC chip embedded in it and two sets of parallel pins down each long side of the IC. Usually DIL packages contain 8, 14 or 16 pins. Manufacturers provide detailed drawings (schematics) and IC pin diagrams that show, most importantly, the power supply pins and other input–output pins. The pins on an IC are usually numbered consecutively anticlockwise, starting with pin 1 at the top left-hand corner when viewing from above the top of the IC. Photo 24.2 illustrates the IC package and its pins lying ready to be inserted into a circuit building protoboard.

Linear devices and application circuits

Photo 24.2
An IC ready to be inserted into a protoboard.



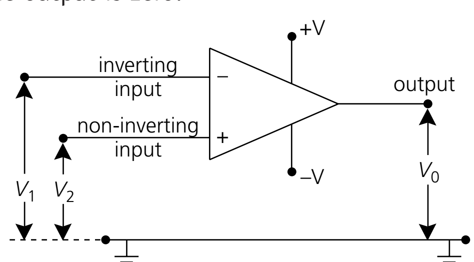
Linear integrated circuits include the voltage regulator IC family, which may form the basis of both positive and negative power supply designs. These are discussed in more detail in Chapter 31. Two other important types of linear integrated circuits are the **operational amplifiers** (Op-Amps), which are used in audio, video and medical electronics applications, and the electronic **timers** used in timing and digital processing circuits. The simplest of these circuits is discussed now, with further examples given in Chapter 31.

Operational amplifier integrated circuits

An Op-Amp is a very high gain (amplification factor) voltage amplifier designed to amplify signals over a wide frequency range. The AC signal frequency range over which an amplifier produces equivalent amplification is called its **bandwidth**. Op-Amps have two input terminals: the inverting input labelled – and the non-inverting input labelled +. The device amplifies the difference in voltage between these two inputs even if one of the inputs is earthed. They usually operate from a dual polarity power supply, which means that the voltages needed to operate the integrated circuit are, for example, +9 V and 0 V, as well as –9 V and 0 V (Figure 24.17). In a lot of circuit diagrams the power supply connections to the integrated circuits are not shown. Note that in this circuit:

- if V_2 is zero, then $V_o = -A_vV_1$, where A_v = voltage gain
- if V_1 is zero, then $V_o = A_vV_2$
- if V_1 is larger than V_2, then the output is negative
- if $V_1 = V_2$, then the output is zero.

Figure 24.17
Operational amplifier.



The voltage gain of the Op-Amp is dependent on the frequency of the signal input. The amplifier gain decreases very quickly as the input frequency rises. An ideal Op-Amp has an infinite voltage gain and an infinite input impedance as well as zero output impedance. The concept of circuit impedance is effectively an AC or frequency dependent resistance.

Remember that these integrated circuits work with AC voltage signals and their input and output resistance will vary with the frequency of the signal they are handling. You will come across this concept again in Chapter 31 when dealing with AC behaviour of components and devices.

If the Op-Amp was operated like this in practice, this very high voltage gain, which is referred to as the amplifier's open-loop gain, would be in the order of 10^5–10^6 times. As well, the circuit input impedance would be approximately 10–100 MΩ. This would make the amplifier circuit difficult to control. For this reason, these amplifiers are usually operated with a **feedback resistor**, R_F, from output to input. This has the effect of reducing voltage gain and input impedance to a manageable level (Figure 24.18). In this circuit:

$$\text{input impedance, } Z_{in} = R_1$$

and

$$\text{voltage gain, } A_V = \frac{R_F}{R_1}$$

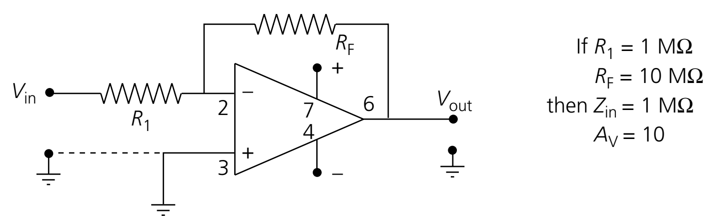


Figure 24.18
Basic inverting amplifier with feedback.

Operational amplifiers are used to produce circuits whose functions may include waveform generators, filters, amplifiers, adders or mixers, integrators and differentiators. One of the most common Op-Amp devices is the **741 Op-Amp**, which is manufactured either singly as an 8 pin DIL package or as a 14 pin DIL Quad Op-Amp package. It has a power dissipation of 310 mW and a temperature range of 0–70°C. A typical application circuit for this device is the inverting amplifier.

Using an Op-Amp as an inverting amplifier Refer to Figure 24.19. The most common application of the Op-Amp is the inverting amplifier configuration that produces an amplified output 180° out of phase with the input. This means that when the input signal is a maximum amplitude the output signal is a minimum amplitude and vice versa. The closed loop voltage gain for the circuit is given by:

$$A_{Cl} = \frac{V_0}{V_{in}} = \frac{-R_F}{R} = \frac{R_2}{R_1}$$

and the effective input impedance of the circuit is simply R_1, as shown in Table 24.2.

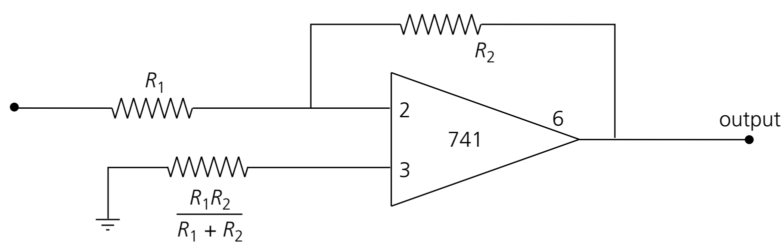


Figure 24.19
Inverting amplifier.

Ch24-Walding 4th 25/8/04 12:33 PM Page 537

CHIP PINOUT DIAGRAMS

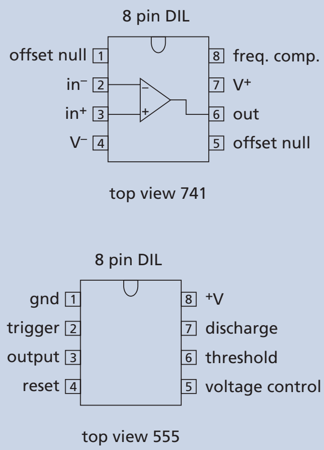


Table 24.2

GAIN, A_V	R_1	R_2	BANDWIDTH	R_{in}
1	10 kΩ	10 kΩ	1 MHz	10 kΩ
10	1 kΩ	10 kΩ	100 kHz	1 kΩ
100	1 kΩ	100 kΩ	10 kHz	1 kΩ
1000	100 kΩ	1000 kΩ	1 kHz	100 Ω

Example

In the circuit of Figure 24.19, the operational amplifier circuit is constructed using R_1 = 15 kΩ and R_2 = 500 kΩ. Calculate:

(a) the closed loop voltage gain, A_V;
(b) the input impedance of the circuit;
(c) the optimum value of the input pin 3 resistance to ground.

Solution

(a) Voltage gain $A_V = \dfrac{R_2}{R_1} = \dfrac{500\ \text{k}\Omega}{15\ \text{k}\Omega} = 33.3$.

(b) Input resistance is simply value of R_1 = 15 kΩ.

(c) Optimum value of resistance $= \dfrac{R_1 \times R_2}{R_1 + R_2} = \dfrac{15\ \text{k}\Omega \times 500\ \text{k}\Omega}{515\ \text{k}\Omega}$ = 14.6 kΩ, or its nearest preferred value of 15 kΩ.

Other applications of the Op-Amp include circuits that act as waveform generators, adders and comparators, and simple integrators. These are further discussed in Chapter 31.

Timer integrated circuits

The **555 timer**, produced by several manufacturers, is specifically designed for precision timing circuits. It can also be used in digital multivibrator modes and as a Schmitt trigger. The Schmitt trigger is a circuit that has the function of converting an analog frequency signal to a digital frequency signal or restoring an electrically noisy digital signal to a very stable and well formed noiseless digital signal. The 555 timer is also made in single or Quad DIL packages. It can operate with supply voltages from 4.5 V to 16 V and can directly drive loads such as relays, LEDs, low power amplifiers and high impedance speakers. Accurate timing periods variable from a few microseconds to several hundred seconds can be produced by using a square wave output controlled by external R–C networks. Timing periods can be started by a trigger signal and stopped using a reset signal. When used as a signal frequency generator (astable mode) the device output can be varied simply using external capacitors and resistors.

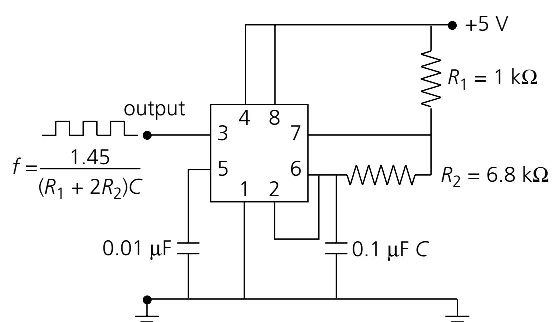


Figure 24.20
555 square wave generator.

The circuit of Figure 24.20 is a simple square wave generator circuit, which could be used as a circuit providing stable timing pulses. This circuit is often called a **square wave clock**. The 555 timer is being used in its astable or free running multivibrator mode because it will continue to produce a series of wave pulses as long as the power supply voltage is connected to the chip. With the values shown in the figure, the frequency of the output is about 1.0 kHz. However, the output frequency is adjustable with either R_1, R_2 or C and is given by:

$$f = \frac{1.45}{(R_1 + 2R_2)\cdot C}$$

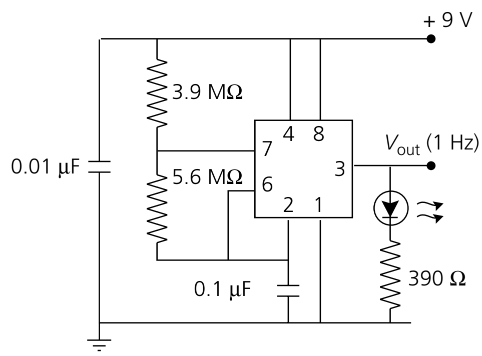


Figure 24.21
555 second interval clock.

The circuit of Figure 24.21 is a similar clock using the 555 timer, but its output frequency is only 1.0 Hz and thus it acts as a second counter. Again, the circuit is a free running or **astable multivibrator**.

Questions

8 List the advantages of modern integrated circuit chips. Do they have any disadvantages?

9 What is the circuit symbol for Op-Amp? Sketch the full circuit diagram for an inverting amplifier whose voltage gain is 200, if an input resistor of 10 kΩ is used. What is the circuit's effective input impedance?

10 Explain the difference between linear and digital ICs. Give two examples of linear ICs.

11 Draw a diagram representing the output waveform of a 555 timer IC chip connected into its astable mode of operation. How would you describe the waveform?

12 Sketch the input and output waveforms on a common set of axes for an Op-Amp of gain 10 and input signal of 1 kHz. Consider an inverting mode of operation.

13 What is meant by the digital term binary levels? How do these binary levels correspond to voltage levels in a digital IC circuit?

14 What is the difference in circuit usage between TTL, 74LS and CMOS digital IC chips?

15 What is the output frequency of a 555 timer IC circuit if the chip is connected as in Figure 24.20 and components R_1 = 15 kΩ, R_2 = 28 kΩ, and C = 0.1 μF?

EI Activity 24.3 LET'S BUILD CIRCUITS

If you have access to circuit building boards especially designed for ICs, such as the SK40 protoboards, then your teacher may supply you with some integrated circuits as described in this chapter so that you can actually build some of the circuits. The simplest circuit to build and get working is the 741 Op-Amp in its inverting mode. Remember to always connect power supply voltages to the correct pins of the IC and connect the power supply last of all.

MICROCONTROLLERS AND ROBOTICS 24.5

One of the most important types of integrated circuits today is the **microcontroller**. In simple terms a microcontroller is a specialist digital computer chip embedded inside a device or product. The best examples of modern consumer devices that contain these chips are vehicle engine controllers, TVs, VCRs and DVD players, digital cameras and mobile phones, refrigerators, cooking ranges and washing machines. In fact any modern device that requires some sort of input from its user is most likely to contain an embedded microcontroller chip somewhere in its internal circuitry. These IC chips contain a CPU (central processing unit) and are usually programmed to perform a small range of tasks, or may even be dedicated to perform only one prescribed task.

Typical microcontroller chips that you may encounter in your electronics course at school include the PIC family from 'Microchip', the AVR family from 'Atmel', the BASIC STAMP controllers by 'Parallax'(which actually contain a PIC chip and are made to allow programming with the BASIC language). You may also encounter the RCX Hitachi controller used in the Lego-Mindstorms robotics kits. Each of these devices will have one or more of the following features:

- a specific program set of tasks stored in ROM memory
- programs that can be changed through programmable CMOS Flash and EEPROM memory
- a dedicated set of input–output (I/O) structures that limit the need for external components
- very low power consumption, typically 50 mW
- are housed inside rugged multipin DIP packages that allow for a wide variety of physical operating conditions such as high temperature or acidic environments.

It is the dedicated I/O structure that is most the important feature of any microcontroller. In the case of the television receiver controller, input signals via infrared beams are received from the remote handset, and the microcontroller sends signals to its outputs which in turn control processes such as picture quality, channel selection and speaker amplifier volume. Outputs are also displayed on LCD panels which respond to inputs from an operator touchpad. This type of microcontroller operation is commonplace in the kitchen or laundry. The proper operation of a modern motor vehicle engine or its peripheral systems, such as ABS brakes or air-bag safety, could not be done without complex microcontroller operation.

The actual microprocessor used in microcontroller chips can vary widely, but most are based on similar processors that once formed the heart of personal computers, such as Z80, 80386, 80586 and Pentium processors. Let's now take a look at some of the types mentioned above.

AVR 8-Bit RISC chips

Atmel's **AVR** microcontrollers have a RISC core running single-cycle instructions and a well-defined I/O structure that limits the need for external components. Internal oscillators, timers, UART, SPI, pull-up resistors, pulse width modulation, ADC, analog comparator and watch-dog timers are some of the features you will find in AVR devices. AVR instructions are tuned to decrease the size of the program whether the code is written in C or Assembly. With on-chip in-system programmable Flash and EEPROM, the AVR is a perfect choice in order to optimise cost and get products marketed quickly. The Atmel® AVR is an 8-bit MCU with up to 128K of programmable Flash and EEPROM.

Table 24.3 ATMEL AVRs

AVR 8-BIT RISC MICROCONTROLLERS		MEMORY CONFIGURATIONS (BYTES)		
Processor	Package	Flash	EEPROM	RAM
tiny AVR	8-32 pin	1-2K	up to 128	up to 128
low power AVR	8-44 pin	1-8K	up to 512	up to 1K
megaAVR	32-64 pin	8-128K	up to 4K	up to 4K

Parallax BASIC Stamps

Named because of their size similarity to a postage stamp, these microcontrollers are made in two forms, BS-1 and BS-2. They usually come on a small development board that is powered from a 9 volt battery and can be connected to one of the ports on a PC so that it can be programmed easily in BASIC. They are most often used in prototyping circuit and program designs.

Table 24.4 BASIC STAMPS

	BS-1	BS-2
RAM	14 bytes	26 bytes
EEPROM	256 bytes	2000 bytes
Max program length	75 instructions	600 instructions
I/O pin number	8	16
Execution speed	2000 lines/s	4000 lines/s

The power of any microcontroller lies within the programming language used to drive the embedded program. In all cases a high-level language such as BASIC, PASCAL, C, a symbolic icon-based language such as Lego RoboLAB or even JAVA can be used. The instructions sequence (program) is then compiled into a form that the microcontroller will understand by further computer software. This also allows it to be downloaded as machine code directly to memory addresses in the microcontroller chip flash memory by way of the computer's parallel communications port, either by direct cable or by infrared beam. Several well-known compilers and microcontroller programmer software are freely available on the Internet for a range of chips. The following listing shows some of the commands in the instruction set that is available for the BASIC STAMP.

Standard BASIC commands:
for...next — normal looping statements
if...then — normal decision making
let — optional assignment
goto — go to a normal label in the program
gosub — go to a subroutine

I/O instructions:
high — set an I/O pin to its high value (1)
output — set the direction of an I/O pin to output
pot — read the value of a potentiometer on an I/O pin
pulsin — read the duration of a pulse coming from an input pin.
sound — send a sound of a certain frequency to an output pin.

Instructions specific to BASIC Stamp:
branch — read a branching table
eeprom — download a program to EEPROM
nap — sleep for a specific time
random — pick a random number
read — read a value from EEPROM

NOVEL CHALLENGE

There are many electronics stores, such as Dick Smith, Jaycar or Altonics, that market electronic construction kits using pre-programmed PIC chips. You may also be able to find experimenters' kits that actually allow programming of the chip by connection to your computer. Also **use the Internet** to find Australian companies marketing tutorial software that teaches you about PIC programming. Just like the BASIC Stamps, PICs can also be programmed in the BASIC language, with several free software programs able to do this, such as PICBasic-lite.

Microchip's PIC family

One of the most successful microcontroller chip families belongs to the PIC range — for example the commonly used PIC16F84A device. It is powerful (200 nanosecond instruction execution) yet easy to program (only 35 single-word instructions). CMOS Flash/EEPROM-based 8-bit microcontroller packs Microchip's powerful PIC architecture into an 18-pin package. The same device can be used for prototyping and production and the end application can be easily updated without removing the device from the end product via the ICSP. It is easily adapted for automotive, industrial, appliances, low power remote sensors, electronic locks and security applications.

Program memory: 1792 (bytes), 1024 (words).

Table 24.5 SPECIFICATION CHART

DATA RAM	SPEED MHz	I/O PORTS	TIMERS	BROWN OUT	ICSP
68	20	13	1+WDT	False	True

Additional features:	20 mA source and 25 mA sink per I/O, 64 bytes data EEPROM
Low voltage device option:	PIC16LF84A
Package options	18 PDIP, die-waffle, uncut wafer, 18 SOIC 300mil, 20 SSOP 208mil, wafer-frame

Robotics and the LEGO RCX brick microcontroller

In the early 1980s Seymour Papert, an early pioneer in artificial intelligence at the Massachusetts Institute of Technology, had the idea to include a computer inside a LEGO® block. This idea has been further refined into the 'LEGO Mindstorms invention system'. Photo 24.3 shows a typical LEGO RCX 1.0 brick with attached sensors and used as part of a robot chassis in classroom robotics study. The RCX (Robotic Control X) is an autonomous LEGO microcontroller based on the Hitachi H8 chip. Table 24.6 shows the specifications. The microcontroller has three input ports (1, 2 and 3) for sensors such as touch, light and rotation, and also has three output ports (A , B and C) able to drive motors, lamps etc.

Table 24.6 RCX 1.0 SPECIFICATIONS

SERIES	H8-3297
ROM size	16K internal
SRAM size	512 internal & 32K in brick
Execution speed	16 MHz @ 5 V
Timers	8-bit × 2 & 16-bit × 1
ADC	8-bit × 8
I/O pins	43
Input only pins	8
Serial port	1
10 mA outputs	10
Power supply	6 × 1.5 V alkaline cells or AC plug-pack

Photo 24.3
Robotics RCX brick.

In the school lab environment, students use the RCX brick as the heart of a course of study in robotics and control systems using PC **RoboLAB** programming software. Sensors are electrically connected to the input ports to take data from the environment, process the data (data-logging) and signal output devices such as motors to rotate. Mobile robots can be built from standard LEGO building pieces that allow the RCX brick and sensors to move about the environment and perform specific tasks or challenges. RoboLAB is the icon-based programming language that comes in a variety of complexity levels: pilot, inventor and investigator. This program is based on National Instruments' 'LabVIEW' virtual instrument software and it can become the basis of powerful experimental investigations. In fact, it was LabVIEW software that NASA used to monitor the Mars sojourner *Rover's* location in 1997.

Programs constructed in RoboLAB are downloaded to the brick using an infrared transmitting tower connected to the USB port of the PC. The RCX brick can also be directly controlled remotely from the computer. An on-chip 16K ROM contains a driver that is run when the brick is first powered up and this driver is extended by downloading firmware to the brick initially. The driver and firmware accept and execute commands from the downloaded student byte code programs which are stored in a 6K region of memory.

An even higher-level language that is available for the RCX microcontroller brick is the C-language variant called Not-Quite-C or **NQC**, written by Dave Baum. This language allows direct command line programming to the brick through an interface called the BricXCC or Brick X command centre.

Photo 24.4 shows a screen dump of a typical RoboLAB inventor-level program for the RCX brick allowing the robot to perform a series of dance steps under software control.

NOVEL CHALLENGE

If your school has access to the LEGO robotics system discussed above, you might try to investigate the robot construction and subsequent programming necessary to complete the following challenge task.

Design and build a robot that effectively senses and follows a dark or black line drawn in a circle of approximate diameter 2.5 m on the lab floor. On starting the program, the robot must follow the circular line until it comes into contact with a solid block, at which point it turns through 90 degrees, travels in a straight line for 1.0 m and then stops. Present your findings and demonstrate your final robot performing its task.

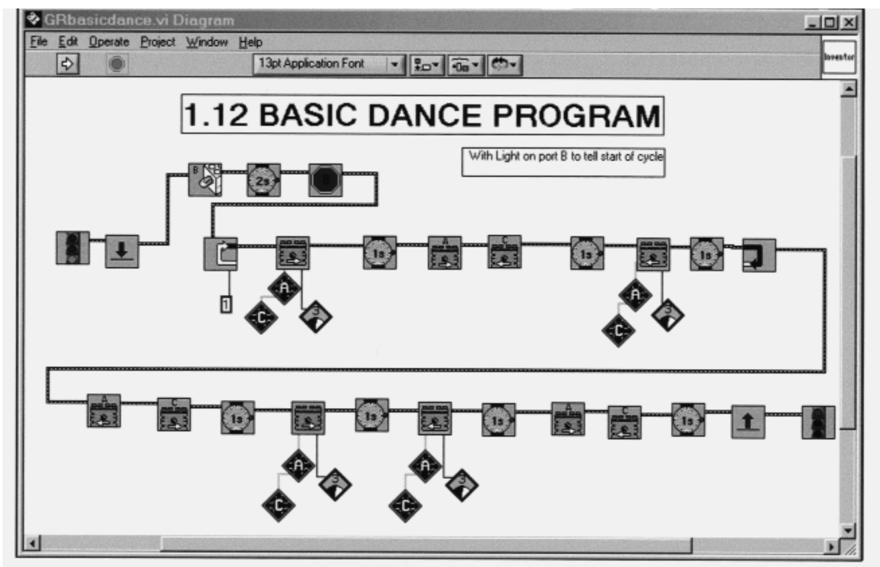


Photo 24.4
Sample RoboLAB program.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***16** An electronic circuit is to be used to convert light intensity changes to sound. What might be the input and output transducers? Briefly describe the functional blocks that might make up the circuit and illustrate these in a block diagram.

***17** Name the three terminals of a transistor and illustrate these on a symbol used to represent an NPN transistor.

***18** What is the difference between an Op-Amp IC and a timer IC? State a common example of each type of integrated circuit.

****19** Sketch the circuit for a non-inverting amplifier whose voltage gain is 500 and whose input impedance is 20 kΩ. Use a 741 Op-Amp chip and include all power supply connections.

****20** Figure 24.22 shows the graphical current transfer characteristics of a particular transistor. Estimate the current amplification factor, β, at I_B = 35 μA and at I_B = 60 μA.

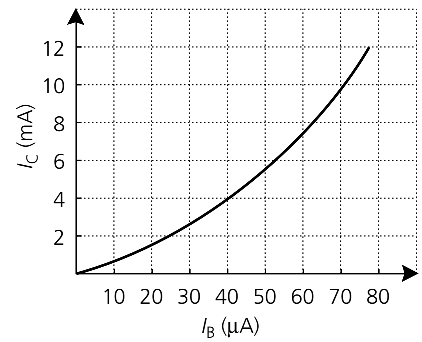


Figure 24.22
For question 20.

****21** A student obtains the following data table from an experiment to measure transistor current gain:

I_B (μA)	20	40	60	80	100	120
I_C (mA)	3.6	7.0	10.8	14.5	17.9	21.5

(a) Plot I_C versus I_B values correctly in order to produce a line of best fit and deduce the current gain, β.

(b) Sketch the circuit that the student may have used to obtain this type of data.

****22** The circuit of Figure 24.23 is used to investigate the voltage gain of a transistor. If the supply voltage used is V_{CC} = 18 V and the following table gives values of the collector current, I_C, and the base–emitter voltage, V_{BE}, use the fact that voltage gain $A_V = \dfrac{\Delta V_C}{\Delta V_{BE}}$ to calculate the gain for this circuit. Complete the table for the values of V_C voltage at the collector.

I_C (mA)	V_{BE} (V)	V_C (V)
10.0	0.62	
30.0	0.69	
50.0	0.76	

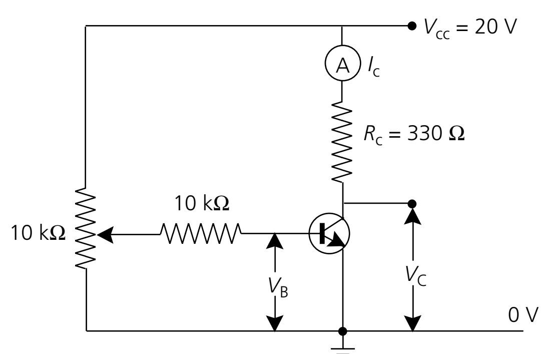


Figure 24.23
For question 22.

***23 Figure 24.24 shows an Op-Amp temperature sensing and heater circuit. The resistance of the thermistor decreases with an increase in temperature.
(a) In what mode is the Op-Amp operating in this circuit?
(b) As the temperature drops, explain what happens to the output of the Op-Amp.
(c) What will be the subsequent effect on the transistor collector working current?
(d) Why does the heater then switch on?
(e) What is the function of the set-temp potentiometer in the circuit?
(f) Redraw the circuit correctly, showing the power supply connections to the Op-Amp chip.

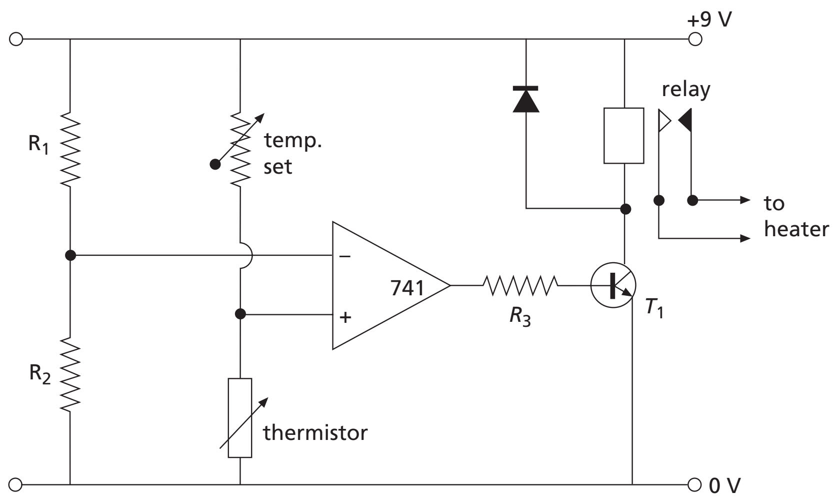


Figure 24.24
For question 23.

Extension — complex, challenging and novel

***24 In Figure 24.23 explain if the LED would be ON or OFF under the conditions of **(a)** full sunlight on the LDR; **(b)** darkness or no light on the LDR.

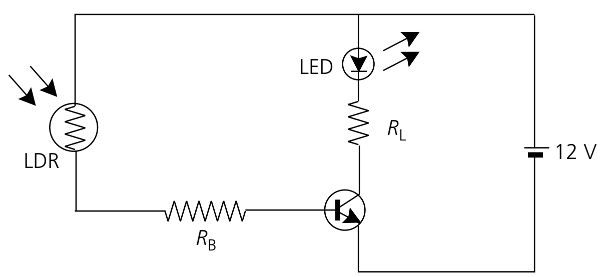
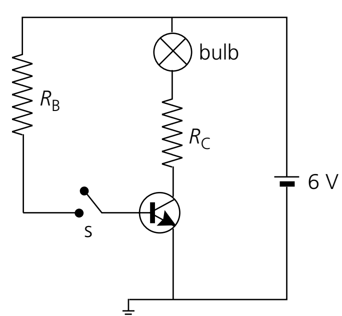


Figure 24.25
For question 24.

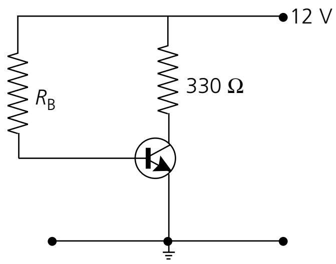
***25 In Figure 24.26 a transistor is to switch a torch bulb as shown. If the bulb is intended to operate at 3 V and dissipate 0.3 W, calculate:
(a) the collector current, I_C, at proper illumination;
(b) the collector resistor, R_C, value required;
(c) the base resistor, R_B, value if the transistor has $\beta = 100$.

Figure 24.26
For question 25.



*****26** In the circuit of Figure 24.27 the transistor has a current gain value of 500 and I_C = 20 mA. Calculate **(a)** the value of R_B; **(b)** the power dissipated in R_B and R_L; **(c)** voltage V_{CE}; **(d)** voltage V_{BE}.

Figure 24.27
For question 26.

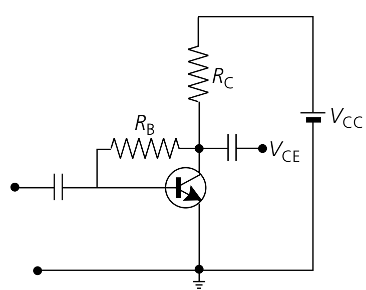


*****27** Figure 24.28 is referred to as a common emitter amplifier with collector feedback. Assuming a silicon transistor, and the values as listed: R_C = 10 kΩ, R_B = 100 kΩ, V_{CC} = 10 V and β = 120,

(a) show that $I_B = \dfrac{(V_{CC} - I_C \times R_C - V_{BE})}{R_B}$;

(b) calculate I_B, I_C and V_{CE} for this circuit.

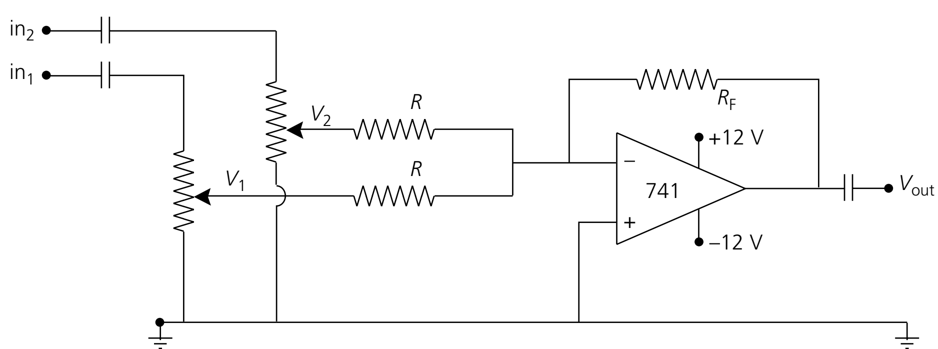
Figure 24.28
For question 27.



*****28** Figure 24.29 shows a 741 Op-Amp in a single audio frequency mixer circuit with two fader inputs. Consider this circuit and explain:

(a) the function of the input potentiometers;
(b) a mathematical equation linking V_{out} to V_1 and V_2;
(c) if $R_F = R$ in this circuit, what is the amplifier gain and give the equation linking the same quantities as in **(b)**;
(d) what would happen to the output voltage if the non-inverting input was connected to a small DC voltage rather than ground.

Figure 24.29
For question 28.



*****29** The circuit of Figure 24.30 contains a 555 timer IC chip and is described as a digital signal injector for testing audio circuits. Describe how this circuit works and determine its likely output frequency range. What is the likely function of the circuit components VR_1 and VR_2? How could the circuit be used as a piece of test equipment in an audio laboratory?

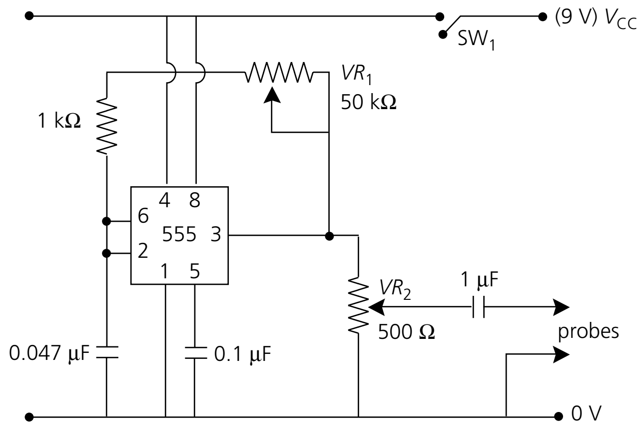


Figure 24.30
For question 29.

The following circuit provides a 2.5 W amplifier based on the LM380 chip suitable for general purpose audio applications:

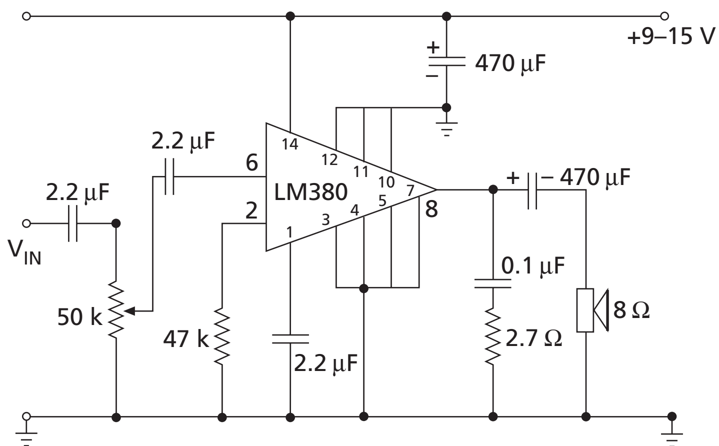
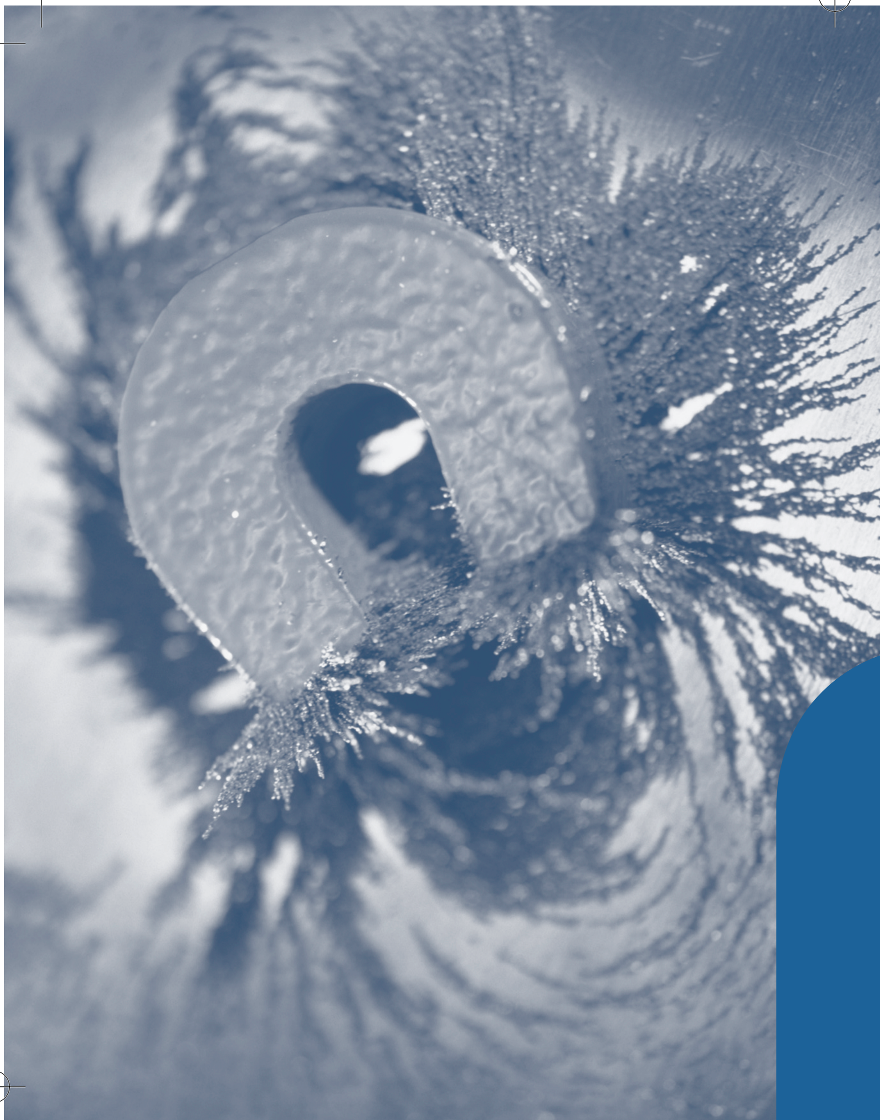


Figure 24.31

Make a list of components in this circuit, and use an electronics supplier catalogue (e.g. Dick Smith) to work out a total price if you wanted to build the circuit.

Ch25-Walding 4th 25/8/04 12:37 PM Page 547



UNIT 08

Magnetism & Electromagnetism

Ch25-Walding 4th 25/8/04 12:37 PM Page 548

CHAPTER 25

Magnetism and Electromagnetism

25.1 INTRODUCTION

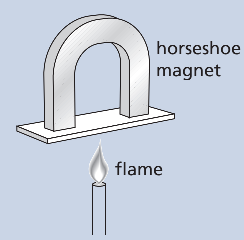
In 1269 a French scholar, Pelerin de Maricourt, also known by his Latin name of Petrus Peregrinus de Maricourt, was taking part in the battle siege of an Italian city. As the action was very slow and dull he wrote a letter to a friend describing his study of magnets. In this letter he described the existence of magnetic poles, regions on the magnets where the force seemed to be most intense, and explained how to determine the north and south pole of magnets, using the fact that the same poles always repelled. He also described how one could not isolate a single pole, for if a magnet were broken in two then each piece would have both a north and a south pole. In the same letter Peregrinus explained that a compass would work better if the magnetic sliver were placed onto a pivot rather than being floated on a cork, and that a graduated scale placed under the sliver would allow more accurate directions to be read. He had described a navigation compass.

Just like Peregrinus way back then, everybody today is fascinated by magnets. In this chapter we will look at the theory and applications of basic magnetism and electromagnetism. These topics were among the earliest scientific investigations and have proven to be extremely valuable areas of research. Some common questions often asked include these:

- Why do compass needles always point north? Have they always done this?
- Why do older recorded tapes always sound worse than brand new ones?
- How do long-distance migrating birds always find their way home?
- Are all metals attracted to magnets or just steel?
- How is it that electric motors are getting smaller but are still getting more powerful?
- Will I lose data from my computer floppy disks if I store them incorrectly?
- Do magnets in pillows and in wristbands really relieve pain and stress?
- Why would you feed a cow a magnet?

NOVEL CHALLENGE

If you heat an iron bar attached to a magnet as shown, at a particular temperature (Curie temperature) the bar falls off. *Why might this be?*

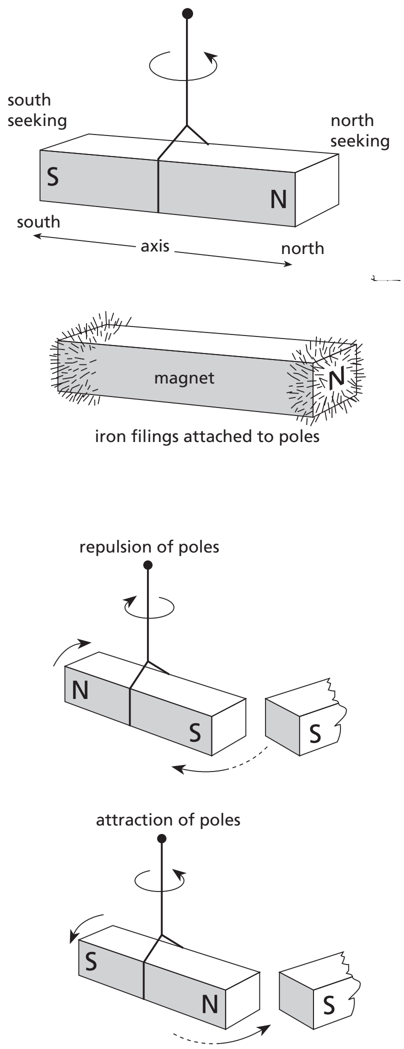


25.2 MAGNETS AND MAGNETIC MATERIALS

Magnetic materials

Magnetic substances are those that can be magnetised. The elements iron (Fe), cobalt (Co) and nickel (Ni), together with certain alloys, display the strongest magnetic properties. Pieces of magnetic mineral ore such as **magnetite** (Fe_3O_4) were probably the earliest magnets discovered and used. In historical times it was recorded that some rocks from a region of Magnesia, now called Turkey, were attracted to each other. These rocks were called magnets. The Chinese used **lodestones** cast in the form of spoons for divination and text from their Han dynasty of about 250 BC describes a south-pointing spoon. In 113 BC details were given on how chess pieces could be made to fight automatically using the lodestone. The term 'lode' seems to refer to the lodestar or guiding star, which refers to how the stone was used in navigation and divination in early Chinese history.

Figure 25.1(a)
Magnetic poles.



Historically, the magnetic phenomenon was regarded as magical, but today we recognise that such forces are due to the fundamental natural force of magnetism. It was in the nineteenth century that it became clear that both electricity and magnetism were related as fundamental forces of nature.

One of the main properties of magnets is their ability to attract objects, chiefly those made of iron. Several naturally occurring minerals are magnetic. Any material able to keep its magnetic properties for a long time is called a **permanent magnet**. The English scientist Michael Faraday showed with sensitive apparatus that, in fact, all substances are influenced by magnets. He classified substances into three types.

- **Diamagnetic** substances, which are very weakly repelled by magnets. This class, in fact, includes most substances. Examples of diamagnetic materials include glass and the metals copper, gold, and bismuth.
- **Paramagnetic** substances, which are very weakly attracted by magnets. Examples include the metals manganese, aluminium and platinum.
- **Ferromagnetic** substances, which are very strongly attracted to magnets. Iron, nickel and cobalt, together with alloys of these metals and aluminium, are the best examples. The majority of small permanent bar magnets used at school are called ALNICO magnets. Can you work out how the name is derived? Ferromagnetic comes from the Latin *ferrum*, meaning 'iron'.

If any magnet is freely suspended by a thread, as shown in Figure 25.1**(a)**, then it will always orient itself so that one of its ends points to the Earth's north pole and one points to the Earth's south pole. Also, as shown in Figure 25.1, if a magnet has small iron filings sprinkled over it then they tend to congregate at both ends where the degree of attraction is strongest.

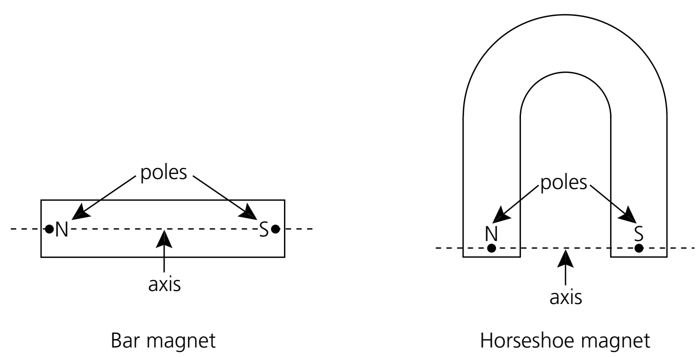
Simple tests between two separate magnets always lead to one of two results — the ends are either attracted together or repelled apart. This is also represented in Figure 25.1. The forces attracting the iron filings or repelling opposite ends of the magnets seem to be strongest at the **magnetic poles** of any magnet. The pole that points north is called the north-seeking pole and is by convention labelled as N, while the pole that points south is called the south-seeking pole and is labelled S. The term north comes from the Italian *nertro*, meaning 'left' because north is to the left when one is facing the rising sun. The line through the poles of a magnet is called the **magnetic axis**. A simple statement of the law of magnetic poles is:

Unlike poles attract while like poles repel.

As seen in Figure 25.1**(b)**, the magnetic axis of a bar magnet passes through the magnetic material itself, whereas in a horseshoe magnet, the bar is bent into a U-shape so that the magnetic axis passes across the gap created between the poles.

The development of magnetic materials has been progressing at a rapid rate since the simplest ferrous magnetics and ALNICO-type alloys of the early twentieth century. Most recently, the development of **rare earth magnets** has occurred, with world leadership roles being taken in research by Australian scientists. In 1992, for example, one of the world's most powerful magnet facilities was opened in Sydney, called the National Pulsed Magnet Laboratory. This facility is used in developing high-tech electronic devices on a sub-atomic level including quantum wires, dots and switches. It houses enormously powerful supercooled magnets capable of producing fields up to a million times stronger than the Earth's magnetism.

Figure 25.1(b)
Magnetic axes.



Rare earth magnets have been produced since about the 1980s, but only recently have become relatively inexpensive to manufacture. The term 'rare earth' is used because they are made from alloys of the rare earth elements or lanthanides. They are mostly made from the alloy **neodymium iron boride** (NdFeB) by a sintering process, meaning they are formed with intense heat and pressure. The element neodymium is found in the mineral sands component called monazite, which is mined in various coastal locations around Australia — a good reason for the environmentally sensitive process of mineral sands mining. After being moulded into various shapes the magnets are coated in zinc to protect against corrosion as the material is highly susceptible to oxidation. One of the big advantages of rare earth permanent magnets over ALNICO alloys is that they retain their full magnetic strength almost indefinitely. The magnetic material, however, will begin to lose magnetisation at high temperatures so these have to be avoided, but their performance is enhanced at very low temperatures. The rare earth super magnets, as they are dubbed, are revolutionising the magnet applications industry. Because rare earth magnets contain much stronger magnetism in smaller volumes of alloy, they are being used in components such as mini hi-fi speakers, mini electric motors and generators, robotics instrumentation, wrist-watches and hearing-aids. In fact, new technology applications are being developed continually. It has been estimated that up to 30 fundamental components in the modern electronically controlled motor vehicle alone will benefit from the use of very small rare earth magnet technology.

NEI Activity 25.1 MAGNETIC MOVES

1 Figure 25.1**(a)** illustrates a magnet sprinkled with iron filings. If you actually did this your teacher would not be pleased — explain why this might be so! Use a magnet placed on the viewing glass of an overhead projector. Place another thin glass plate on top of it and a piece of clear acetate plastic on top of this. Now you can sprinkle iron filings over the magnet, but on to the acetate sheet. Give the sheet a gentle tap and observe the pattern of iron filings produced. Try to explain this to the rest of the class.

2 The *Guinness Book of Records* lists the world's largest magnet and electromagnet. Research these and find out their characteristics as well as what they are used for. In what field of physics–engineering are very large magnets required?

Inducing magnets

If a piece of iron or steel, such as a paper clip, is allowed to come into contact with one of the poles of a permanent magnet then it also will become magnetised and attract iron filings, as shown in Figure 25.2. The paperclip will remain a magnet while in contact with the bar magnet, but once separated will probably lose most of its magnetic attraction properties again. The paperclip has become an **induced magnet**, as opposed to a permanent magnet. Pure iron, or 'soft' iron as it is called, becomes quite a strong induced magnet while in contact with another magnet. The **induced poles** are oriented as shown in Figure 25.2 and this is most easily tested with a third magnet whose poles are marked in some way, using the observed forces of repulsion or attraction.

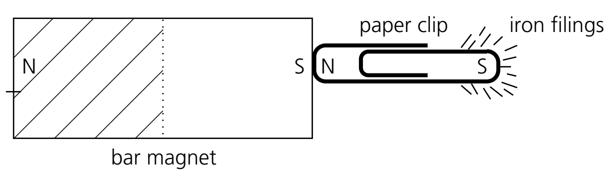
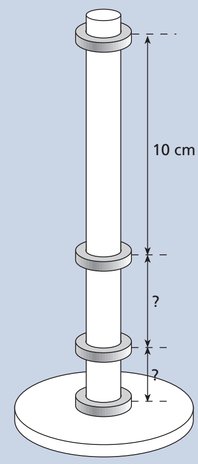


Figure 25.2
Induced magnetism.

NOVEL CHALLENGE

Four ring magnets are placed on a wooden pole as shown.
If the distance between the top two is 10 cm, calculate the other spacings?



PHYSICS UPDATE

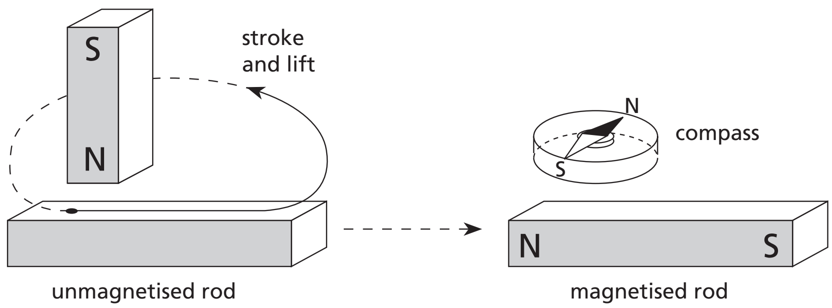
In May 1997, using a Ni/Sn coil, scientists at Lawrence Berkeley National Laboratory in California achieved the highest ever magnetic field of 13.5 T. That's big!

Superconducting magnets have become an important tool in the application of nuclear magnetic resonance (NMR) for materials and medical research, especially magnetic resonance imaging (MRI). With increasing magnetic field strengths, scientists are able to view materials with higher clarity and resolution. The National High Magnetic Field Laboratory (NHMFL) in the USA is working on building the largest NMR magnet in the world, capable of field strengths of 25 T. Now that's even bigger!

It is possible to permanently magnetise a piece of steel alloy using the properties of induced magnetism by using a stroking technique, as shown in Figure 25.3. In fact, this method works for any ferromagnetic material but several steps need to be completed:

- Always move the permanent magnet in the same direction along the rod.
- At the end of a stroke move the magnet in a circle back to the beginning.
- Always keep the same pole of the permanent magnet in contact with the rod.

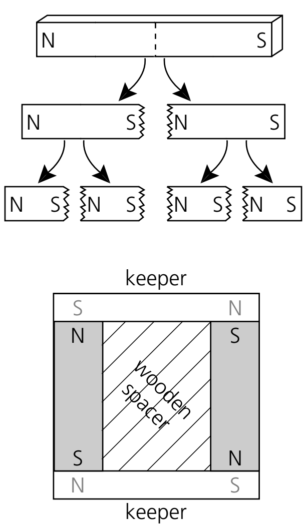
Figure 25.3
Producing a magnet.



After completing from 20 to 50 strokes it will be found that the rod has retained some degree of permanent magnetism. This technique works because small microscopic zones within the rod called **magnetic domains** behave like miniature bar magnets and become aligned so that their axes of magnetism are parallel. This effect will be further explained later in this section. You can also magnetise a rod of steel by holding it in a north–south direction and hitting it repeatedly on one end with a hammer.

Magnetic compass needles are, of course, small bar magnets and as such can be a useful test device for magnetic poles. A compass needle can be used to test the polarity of the bar magnet induced by the stroking technique. If a permanent bar magnet is gently broken in half, each piece itself becomes a smaller, weaker bar magnet, which can also be tested using the compass needle (Figure 25.4).

Figure 25.4
Breaking and storing a magnet.



An iron rod can also be magnetised by being placed into the centre of a long cylindrical coil or solenoid that is carrying a large electric current. This electromagnetic coil and its properties are further discussed in Section 25.4, but for now it is sufficient to realise that the powerful inducing magnetic field produced by the solenoid is responsible for permanently rearranging the domains in the iron rod and producing some permanent magnetism. In fact, laboratory bar magnets are produced in this way while the bar magnet alloy is still in a high temperature state, which allows easy rearrangement of the domains.

Theory of magnetism

Materials can be classified according to how well they retain their magnetism. Those materials that are difficult to magnetise initially but retain their magnetism once induced are called magnetically **hard** materials. Those that are easy to magnetise but lose their induced magnetism almost immediately when removed from the magnetising source are called magnetically **soft** materials. Examples of hard materials, such as steel, are used in long-life applications such as magnetic recording heads on audio and video cassette recorders, loudspeaker magnets and the ferric and ferrochrome alloys used in audio and video cassette magnetic tapes. Examples of soft materials, such as iron and iron–nickel alloys called **mu-metals**, are used as electromagnet cores, in relays and switching solenoids and in magnetic shielding cases surrounding sensitive electronic instruments. You probably have magnetic shielding in your watch. These materials, when used as electromagnet cores, become strongly magnetised only when a very strong electric current is passed through the electromagnet coil.

Audio and video magnetic tapes consist of a flexible polyester film onto which is placed a very thin magnetisable layer of iron oxide or chromium oxide. The polyester is, in fact, the same material from which plastic soft drink bottles are made. The process of recording uses a pulsing electromagnet in the recording head to align the domains on the tape relative to each other in a specific pattern. When the playback head of the recorder is subsequently passed across the tape, this same pattern of aligned domains induces electrical signals into the pickup coil of the head. Recorded tapes should obviously not be placed in the near vicinity of strong magnetic forces such as hi-fi speakers, VCRs and television sets, and also should not be subjected to high temperatures, otherwise the prealigned domains will be disordered and the recorded information lost from the tapes. This loss of recorded information is called fade out.

The same alignment of domains occurs when computer disks (floppy or hard disks) are recorded. (See Photo 25.1.) Modern devices, such as credit cards and telephone phonecards, contain a magnetic strip with encoded information stored on it in the form of permanently aligned domains in particular patterns. Deliberate erasing or rerecording of the information on computer disks and credit cards simply involves electromagnetic recording heads realigning the necessary magnetic domains in the magnetic material. This will be further discussed in Section 25.4. Once again it should be realised that these devices, just like tapes, are easily erased or made unusable if they remain close to strong magnetic fields for any period of time. It would be a pity if you lost your physics assignment, done on a word processor and stored on floppy disk, just because you left your disk on your hi-fi speaker!

Photo 25.1
Magnetic tape and disk.

Even a hard magnet will become **demagnetised** if it is heated strongly. The temperature at which it loses its magnetism is called the **Curie temperature**, named after Pierre Curie who investigated this phenomenon in 1895. For iron, this temperature is 773°C. Demagnetisation also occurs if magnets are dropped or hammered, which is why they need to be handled with care in the laboratory and not continually knocked around. Even if left to stand on a shelf, a single isolated magnet will eventually lose its magnetism due to the combined effects of both temperature and the Earth's magnetism influencing the alignment of the domains within the magnet material. For this reason magnets are purchased in pairs with soft iron or mu-metal **keepers** and should be stored as a complete magnetic circuit, illustrated in Figure 25.4. In this configuration there are no free magnetic poles and the alignment of domains is a continuous N–S orientation.

As illustrated in Figure 25.4, cutting a magnet in half simply produces smaller bar magnets. This seems to indicate the fact that magnetism is connected to the microscopic, and even the atomic, structure of matter. It is known that each electron in an atom does, in fact, act as a small magnet due to its motion of rotation and spin. The electrons are known to always spin on their axes in a very exact manner relative to the atom. In most materials the combined effect of many spinning electrons within the atom cancels out any net magnetism surrounding an individual atom or collective region of atoms within the material. Physicists would say that the net **magnetic spin** is zero. In the case of the ferromagnetic materials, however, the spin of the electrons does not cancel out but produces a magnetic effect associated with the atom. Adjacent atoms subsequently affect each other and become aligned over small zones or regions that are the magnetic domains of the material.

NOVEL CHALLENGE

A Stainless steel is about 80% iron but is non-magnetic because the chromium atoms lock the iron atoms in place. How does this prevent stainless steel from being magnetised?

B Submarines can be detected by the residual magnetism of their iron hulls. Why not make a submarine from stainless steel? Shipbuilders need to use steel with a very high tensile strength to stand the water pressure at depth, and therein lies a clue.

Normally the direction of the N and S poles of adjacent domains point in different directions so that the individual magnetic forces cancel out over the material and it is unmagnetised. See Figure 25.5**(a)** in which the boundaries represent the contact between different magnetic domains in which all atoms have similar alignment. The arrows indicate the strength and direction of the magnetism within each domain. If the material is stroked by an inducing magnet, or placed in a current-carrying electromagnetic coil, the magnetic axes of more atoms become aligned with the outside field direction. Some order is superimposed on

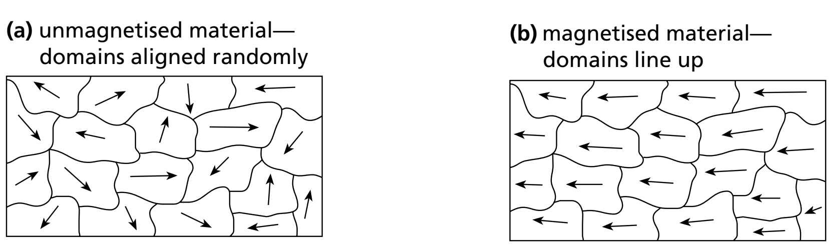
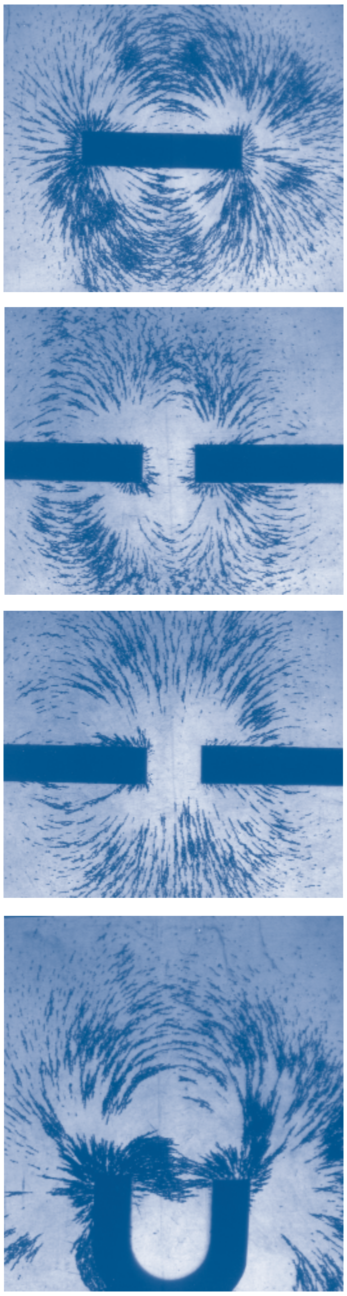


Figure 25.5
Magnetic domains.

Photo 25.2
Magnetic fields.



the material, and the boundaries of the domains, with fields parallel to the outside field, expand. Not all domains have their field in the same direction, but the majority become aligned with the outside field. This has the combined effect of producing a large magnetic field strength within the material and thus it has become magnetised. (See Figure 25.5**(b)**.)

If the magnetic domain explanation and the source of magnetism as related to electron spin are in fact correct, then we would always expect to see magnetic poles occurring in pairs. North and south magnetic poles existing independently should not be possible. The theoretical existence of **magnetic monopoles** or single north or south poles in isolation was postulated by the physicist Paul Dirac in 1931 and research has been going on ever since to try to prove their existence.

Questions

1 What is a magnetic substance and does it exist naturally?
2 What is the difference between a magnetic pole and an induced magnetic pole?
3 Describe how you might tell the difference between two similar metallic rods, only one of which you know to be a magnet, but the other is magnetic material.
4 What is the difference between the types of magnetic substances?
5 What is the importance of modern research into magnetic alloys and supermagnets?
6 Ferrochrome audio tapes generally require a much stronger signal from the recording head of cassette recorders. Explain what might be the reason for this in terms of domain theory of magnetism.

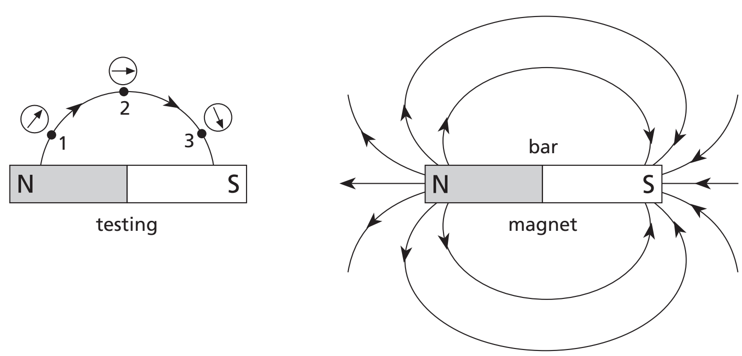
Magnetic forces and fields

If you carried out the iron filings activity with the overhead projector you would have obtained patterns similar to those in the set of photos shown here.

These are the patterns produced by a single bar magnet, a single horseshoe magnet and pole repulsion or attraction. These patterns of iron filings indicate a zone of influence surrounding the magnets called a **magnetic field** of force. This magnetic force field is in fact three-dimensional and the iron filings represent the cross-section through the full 3D field. You should try to visualise what the field would look like in 3D or — alternatively — your school laboratory might have a 3D model of magnetic fields using small iron filings suspended in vegetable oil. This piece of equipment makes viewing the full 3D magnetic field quite easy. Figure 25.6 illustrates the conventional method of drawing magnetic field diagrams using **magnetic field lines** or lines of **magnetic flux** with directional arrows indicating the direction of the force on a small test magnetic north pole placed into the field. The word flux comes from the Latin *fluere*, meaning 'to flow'.

The lines will consequently always be oriented from north pole to south pole about a typical bar magnet. The direction of the force at any point in a magnetic field diagram is given by the tangent to a field line at that point. This direction can always be tested with a

Figure 25.6
Lines of magnetic field (lines of flux).



small compass needle, and it is important to realise that the lines of magnetic flux never cross because this would be indicating two independent force vector directions at the same time, a situation that is, of course, impossible.

To fully describe the nature of a given magnetic field at any point within it, we need to describe both the field's magnitude and its direction. Hence the magnetic field is a vector quantity and is represented by the vector symbol $\boldsymbol{B}$. The magnitude of the magnetic field can be represented by the flux lines being drawn closer together or further apart, as shown in the field representation of Figure 25.7, while its direction as before is given by the arrowhead showing the direction of the force on a single north pole.

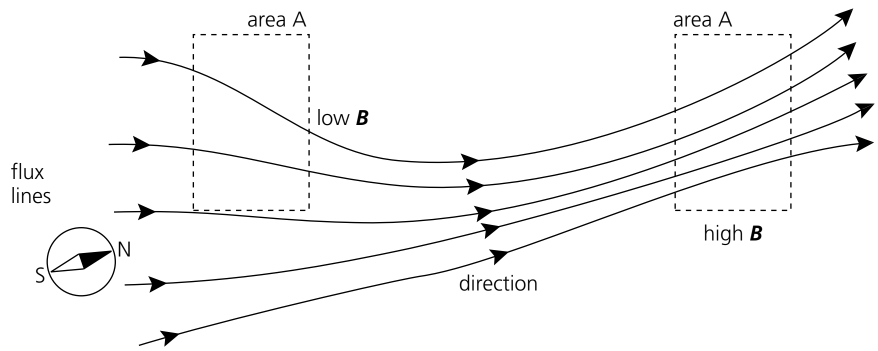


Figure 25.7
Magnetic field strength (flux density).

NOVEL CHALLENGE

A magnetic domain contains about 10^6 iron atoms. If the atoms are 10^{-10} m apart, how big a cube would a domain be?

It is helpful to establish a convention when drawing magnetic field diagrams. The lines of magnetic flux within the field $\boldsymbol{B}$ are drawn so that the number of lines per unit area is proportional to the strength of the field. Recall this is the same concept as for electric field lines. If this convention is used then we can define **magnetic flux**, ϕ, as the total number of lines passing through any given area, and the magnetic field strength, called **magnetic flux density**, $\boldsymbol{B}$, as the magnetic flux per unit area. Hence:

$$\phi = \boldsymbol{B} \times A$$

where ϕ = magnetic flux measured in weber (Wb); $\boldsymbol{B}$ = magnetic flux density or field strength in tesla (T); A = area of the field being considered in square metres (m^2) perpendicular to the flux lines.

Note, therefore, that with SI units, one tesla is the equivalent of one weber per square metre:

$$1.0 \text{ T} = \frac{1.0 \text{ Wb}}{1.0 \text{ m}^2} = 1.0 \text{ Wb m}^{-2}$$

The weber (Wb) is named in honour of **Wilhelm Weber** (1804–1890), a German physicist and close collaborator of Karl Friedrich Gauss (1777–1855), who mathematically modelled electric and magnetic fields. An alternative but older unit of magnetic field strength, $\boldsymbol{B}$, is the gauss (G), where $1.0 \text{ G} = 1 \times 10^{-4} \text{ N A}^{-1} \text{ m}^{-1} = 1 \times 10^{-4}$ T.

The tesla (T) is named in honour of **Nikola Tesla** (1856–1943), a Croatian-born physicist who became a citizen of the United States in 1889, and whose research and patents led to the development of AC power supply and transmission. He was an engineering assistant to the famous Thomas Edison, and also designed the first electric chair in New York in 1905.

Example

In a uniform magnetic field, the field strength is 5.5×10^{-4} T. If an area within the field is defined as having a length of 0.2 m and a width of 0.1 m, calculate the magnetic flux, ϕ.

TEST YOUR UNDERSTANDING

Propose a way to show that the force between two magnets is not electrostatic.

Solution

Given that $\boldsymbol{B} = 5.5 \times 10^{-4}$ T, $A = 0.2$ m $\times$ 0.1 m = 0.02 m^2, use:

$$\phi = \boldsymbol{B} \times A$$
$$\phi = 5.5 \times 10^{-4}\ \text{T} \times 0.02\ \text{m}^2$$
$$\phi = 1.1 \times 10^{-5}\ \text{Wb}$$

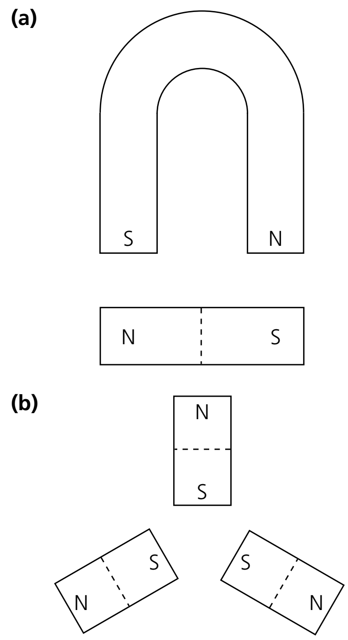


Figure 25.8
For question 9.

Questions

7 Explain the difference between magnetic flux and magnetic flux density.

8 Calculate the magnetic flux density in a region where 2.5×10^{-5} Wb cut through an area whose dimensions are 0.15 cm $\times$ 0.75 cm.

9 Draw the set of magnetic field lines surrounding the formation of magnets and rods in Figure 25.8.

Physicists have developed extremely sensitive magnetic field detectors that rely on the voltages produced as electric currents flow in conductors within the magnetic field. These detectors can measure magnetic field strengths as low as 1×10^{-16} T, which allow research not only into basic magnetisation of materials but also for mapping small variations in the Earth's magnetic field. These data may yield information on underground mineral ore deposits. Within the human body, very weak magnetic fields are generated by various organs and medical diagnosis instruments detect these fields and couple them with imaging processes using computers, in order to look at abnormal tissue such as cancerous tumours. The magnetic effect of atoms in individual cells and tissues can also be detected using associated magnetic technology such as magnetic resonance imaging (MRI). Some of these medical applications are further discussed in Chapter 33.

MAGNETIC FIELD OF THE EARTH 25.3

NOVEL CHALLENGE

The geographical north–south axis (the 'geodesic' pole) is marked with a brass plaque at the South Pole. It has to be shifted about 10 km each year. Why is this, if the geodesic pole doesn't shift?

The Earth has a magnetic field surrounding it which is called the **magnetosphere**. The magnetic field originates from an internal magnetic polarity similar to that of a large bar magnet whose poles are roughly aligned with the geographical north–south axis of the Earth itself. The angular difference existing between the Earth's magnetic axis and its spinning geographical axis is called the **angle of declination**, as shown in Figure 25.9**(a)**.

At present, the north magnetic pole is situated at 101° W longitude and 75° N latitude on the Earth's surface. What is called the magnetic north pole of the Earth is actually a magnetic south and, of course, vice versa for the so-called south magnetic pole.

The Earth's magnetic field also affects a freely suspended compass needle in different ways at different latitudes. (See Figure 25.9**(b)**.) If a compass needle is suspended on a pinpoint it will only come to rest horizontally at the Earth's equator. At all other latitudes the suspended needle will show a downward angle of dip or **inclination**. In fact, at regions close to the magnetic poles of the Earth, the angle of inclination approaches 90° and, of course, makes the normal operation of navigation compasses quite useless.

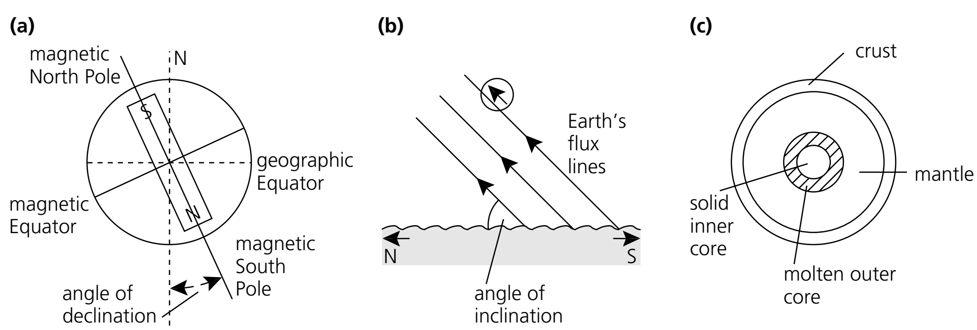


Figure 25.9
Earth's magnetic field: **(a)** declination; **(b)** inclination; **(c)** Earth's interior.

The magnetic poles of the Earth have had a tendency throughout geological history to wander around over the surface of the Earth, in terms of latitude and longitude. This means that the internal processes causing the formation of the magnetic field and the subsequent effective bar magnet have moved around considerably. It has been estimated that the exact position of the magnetic poles may change by as much as ten to twenty kilometres per day. There is considerable geological evidence, based on natural magnetic mineral orientations within lavas that were originally molten, that the magnetic poles of the Earth have even reversed numerous times since the Earth's formation. The real processes that form the Earth's magnetosphere as well as control its motions and pole orientations are as yet not well understood by physicists and geologists.

Research based primarily on seismic studies and analysis of earthquake recordings has produced an internal view of the Earth, presented in a simplified way in Figure 25.9**(c)**. It is thought that the Earth's magnetic field is caused by the cycling motion of the molten material, mainly iron and nickel, that makes up the outer core of the Earth. The motion of this molten material surrounding a solid inner core of iron, together with the spinning of the Earth itself, produce electric currents that flow through the Earth and maintain the magnetic field. Magnetism and flowing electricity are closely related, as seen in the next section.

Figure 25.10 shows a wider view of the Earth's **magnetosphere**, including the magnetic field lines and the region of trapped radioactive particles that come primarily from the Sun's **solar winds**. This region comprises the Van Allen radiation belts, named after **James A. Van Allen**, born 1914, whose work was very important in establishing the International Geophysical Year (1957–58) and the launching of the satellite Explorer 1 which detected the belts. The Van Allen region surrounding the Earth appears to be divided into two zones separated by what is called the 'slot'. The inner zone, comprised mostly of high energy protons, reaches its maximum intensity at a height of about 4800 km. This inner zone is much more stable than the outer zone, which is comprised mostly of highly energetic electrons that are strongly affected by the solar activity that supplies the electrons. This outer zone reaches its maximum intensity at a height of 16 000 km. The Van Allen region, as shown in Figure 25.10, does not completely envelope the Earth but extends from latitude 75°N to 75°S on the daylight hemisphere and from 70°N to 70°S on the night hemisphere. The comet-like shape of the Earth's magnetosphere is caused by the pressure of the solar wind particles being flung away from the Sun in all directions. This magnetosphere is very important for all life on Earth as it acts like a radiation protective shield. It rapidly decelerates charged particles travelling through space and deflects them into the radiation belts and toward the magnetic poles of the Earth. In periods of intense solar particle bombardment of the Earth from the Sun, these charged electrons and protons are forced to spiral downward along the Earth's magnetic field lines close to the poles. These particles collide with atoms present in the upper atmosphere, producing visible light in the form of spectacular polar light shows called **aurorae**.

Scientists believe that the Earth's poles have 'flipped' or reversed many times in geological history. During a reversal the Van Allen belts disappear and the Earth receives enormous bursts of solar wind, causing havoc to the life forms on Earth. It has been postulated that dinosaurs became extinct during one such reversal of the poles.

NOVEL CHALLENGE

In 1986 scientists discovered that the yellowfin tuna has 10 million magnetic crystals in its skull. How could you test whether tuna use these crystals to aid navigation, as has been suggested?

NOVEL CHALLENGE

Which would reach the higher temperature when dissolved in acid: a piece of magnetised steel or a piece of unmagnetised steel? Explain.
If a piece of magnetised steel dissolved in such acid, would the solution of iron chloride be attracted to a magnet?
If not, why not?

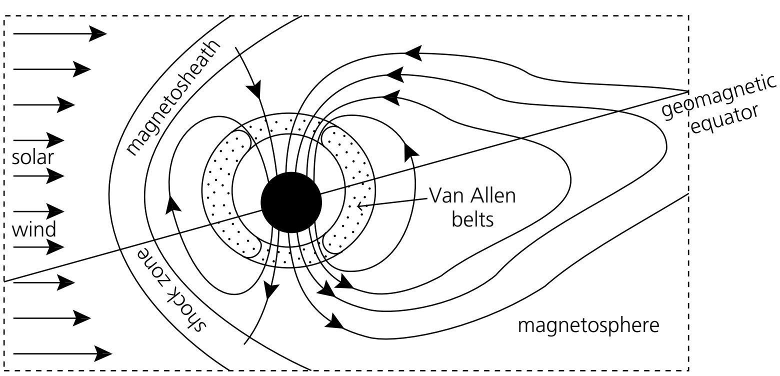


Figure 25.10
Earth's magnetosphere.

NOVEL CHALLENGE

Which would reach the higher temperature when dissolved in acid: a steel spring when it is compressed or when it is relaxed? Why?

NOVEL CHALLENGE

We read on the Internet that magnets are fed to cows to attract bits of wire and nails that they eat with the grass. This seems a bit far fetched when you consider how long wire would last in a cow's acidic stomach. Test it — use 0.17 M HCl (pH = 0.8).

NEI Activity 25.2 MAGNETISM APPLICATIONS

Use reference sources, such as a CD-ROM encyclopaedia or other library material, to research each of the following topics and prepare written reports:

1. The study of biomagnetism and the relationship between migratory birds and the Earth's magnetosphere. Why do certain bird species appear to have a high concentration of iron in particular parts of the brain?
2. The phenomenon of a solar flare and its relationship to telecommunication difficulties on Earth with radio and television.
3. The use of military applications, such as magnetic mines, which were employed successfully as recently as the Gulf War crisis in 1991.
4. Geologists make use of the change in the Earth's magnetic field strength surrounding large iron ore deposits as a means of detecting them. Find out what a magnetometer is and how it is used to survey large expanses of land for possible mineral ore deposits.
5. Do the people who sell magnetic pillows and wristbands offer any scientific evidence for their healing claims, or is it just mumbo-jumbo? Some horse magazines offer magnetic rugs for the comfort and protection of the animal! What do you think?
6. Several species of aquatic bacteria swim along magnetic field lines. They have tiny chains of magnetite crystals of one domain each. When stirred up, bacteria swim north, which is towards the bottom (in the northern hemisphere where they live). What would they do if this experiment was carried out on the equator or in the Southern Hemisphere?

ELECTROMAGNETS 25.4

Permanent magnets are, in general, not as useful as electromagnets because their magnetism cannot be turned on and off at will. In order to look at the practical applications of electromagnets, let's start by looking at the basic principle of interaction between electric current and magnetic fields.

If the circuit in Figure 25.11 is set up with a cardboard support on which is placed several small magnetic field plotting compasses, then a very interesting effect is seen when the current is switched on. Note that the best voltage source to use in this circuit is a 12 V car battery. The experiment should be performed quickly as the large current flowing, greater than 20 A, causes rapid heating of the copper conductor wires. The Danish physicist Hans Christian Oersted (1777–1851) published the results of a similar experiment in 1820 in the British scientific journal *Annals of Philosophy*, in an article called 'Experiments on the effect of electricity on the magnetic needle'. In this work he described the way the compass needle follows the almost circular pattern of magnetic field lines around the current-carrying wire.

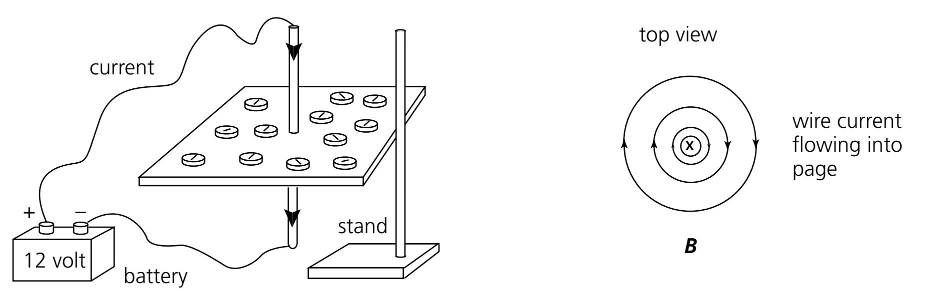


Figure 25.11
Circuit for a current-carrying conductor.

Ch25-Walding 4th 25/8/04 12:37 PM Page 558

Figure 25.11 also shows the typical field diagram used to represent the nature of the field. The main characteristics of the magnetic field in this situation are:

- the field lines are circular and concentric around the wire
- the strength of the field decreases away from the wire, that is, it decreases with radius, r, in metres
- the direction of the field reverses if we reverse the direction of the current
- the strength of the field is proportional to the magnitude of the current, I, in amps.

The magnetic field direction can be easily remembered by making use of **Maxwell's screw rule**. This rule makes use of the right hand and conventional flow of current. It should be noted at this point that all rules associated in physics with electromagnetism normally make use of your right hand. Always keep this in mind or else, if you use your left hand, you'll be predicting the exact opposite of what occurs in nature. Point your thumb along the direction of the current and then curl your fingers around the wire. The direction in which your fingers are pointing represents the direction of rotation of the magnetic field lines. (See Figure 25.12.)

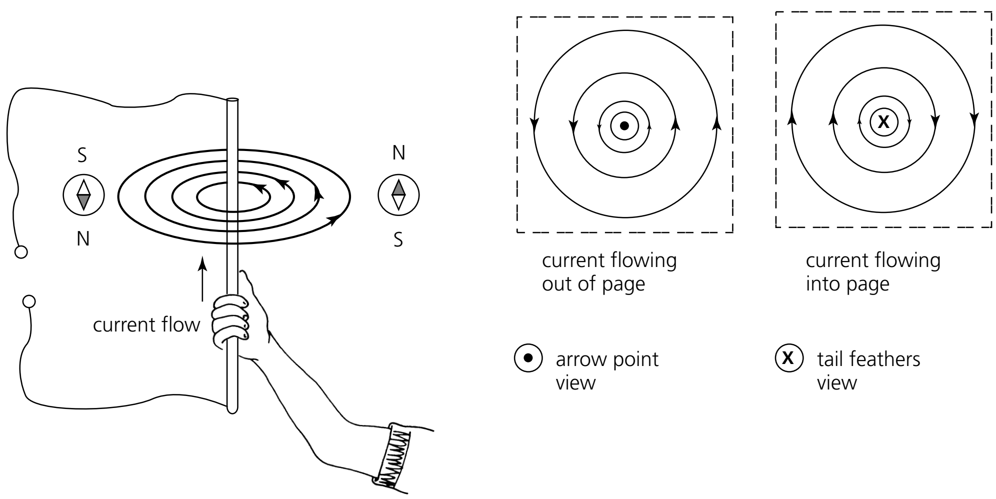


Figure 25.12
Right-hand screw rule.

When visualising the field in three dimensions, remember that the concentric circles create a series of cylinders surrounding the wire, with the strength of the field decreasing radially away from the wire itself. The characteristics of the field allow the proportionality to be stated as, at any point:

$$\boldsymbol{B} \propto \frac{k}{r}$$

but actual measurements made allow a constant of proportionality to be established such that:

$$\boldsymbol{B} = \frac{\mu_0 I}{2\pi r}$$

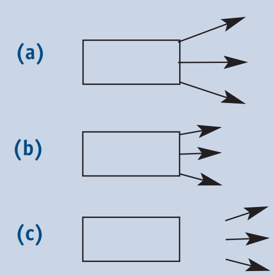
where $\boldsymbol{B}$ represents the value of the magnetic strength at some radial distance r from the centre of a wire carrying current I amps. The constant μ_0 is called the **permeability of free space** and has a value of $4\pi \times 10^{-7}$ T m A^{-1}, hence:

$$\boldsymbol{B} = \frac{2 \times 10^{-7} \times I}{r}$$

for a single current-carrying conductor.

NOVEL CHALLENGE

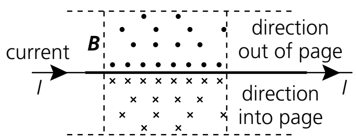
Figure **(a)** shows a part of the field about an electromagnet. *If it was turned off, would the field lines change as in Figure* ***(b)*** *or* ***(c)****?*



Ch25-Walding 4th 25/8/04 12:37 PM Page 559

Figure 25.13 illustrates another method of representing the field, with magnetic lines entering the page as crosses and magnetic field lines exiting the page as dots. Consider these as the flight feathers and the tips of the vector arrows.

Figure 25.13
Field surrounding a current-carrying wire — cross-section.



In a vacuum or air, the value of the constant, 2.0×10^{-7} N A^{-2}, is referred to as the **magnetic constant**, ***k***, and thus the vector magnitude of the magnetic field is:

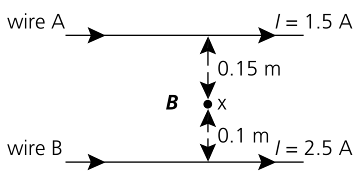
$$\boldsymbol{B} = \frac{kI}{r}$$

Don't get this confused with the electrostatic constant *k* from Chapter 21.

Note that when more than one wire exists, the magnetic fields have to be vectorially added to find a resultant, such as is illustrated in the next example.

Example

Figure 25.14
Example of wire A and B parallel conductors.



Consider Figure 25.14. In the figure, two separate wires are in close proximity. If wire A carries a current of 1.5 A and wire B carries a current of 2.5 A, calculate the value of the magnetic field strength at a point x between the two wires.

Solution

The magnetic field due to wire A at point x is:

$$\boldsymbol{B}_\text{A} = \frac{kI}{r} = \frac{2 \times 10^{-7} \times 1.5}{0.15} = 2 \times 10^{-6} \text{ T into the page}$$

The magnetic field due to wire B at point x is:

$$\boldsymbol{B}_\text{B} = \frac{2 \times 10^{-7} \times 2.5}{0.1} = 5 \times 10^{-6} \text{ T out of the page}$$

But $\boldsymbol{B}_\text{tot} = \boldsymbol{B}_\text{A} + \boldsymbol{B}_\text{B}$ if we choose out of the page as the positive direction. Hence:

$$\begin{aligned}\boldsymbol{B}_\text{tot} &= -(2 \times 10^{-6}) + (5 \times 10^{-6})\\ &= 3 \times 10^{-6} \text{ T out of the page}\end{aligned}$$

Multi-turn coils

Figure 25.15
Single loop coil field.

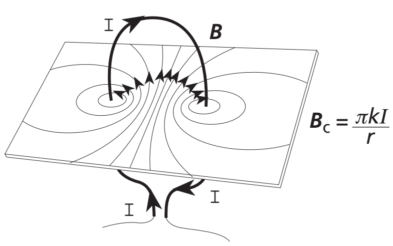


Figure 25.15 shows a wire bent into a single **circular loop**. This loop can be considered as made up of many small, straight segments each adding its individual magnetic field together at the centre of the loop where the field will be the strongest and will be directed through the loop as shown. The direction is once again determined by the right-hand rule. The magnetic field strength at the loop's centre is given by:

$$\boldsymbol{B} = \frac{\mu_0 I}{2r} = \frac{\pi kI}{r}$$

where *r* is the coil radius.

Using a cylindrical former made of cardboard or plastic and winding many hundreds of turns of wire side by side, as shown in Photo 25.3, produces a device called a **solenoid**. The word solenoid comes from the Greek *solen*, meaning 'tube'. This concentrates the magnetic field lines into a region of space that produces an almost perfectly uniform magnetic field within the hollow body of the solenoid. The magnetic field at the centre of a very long solenoid is constant and is found to depend only on the current flowing in the coil as well as the number of turns per unit length of the solenoid. This type of field is illustrated in Figure 25.16, and the formula for the magnitude of the field strength in the solenoid's centre is:

$$\boldsymbol{B} = \frac{2\pi kNI}{L}$$

where N is the number of turns; L is the coil length in metres.

The polarity of the solenoid's magnetic field is often predicted with the right-hand **grip rule**, which states that if you grip the solenoid in the right hand so that your fingers naturally curl around the solenoid in the direction of conventional current flow then the thumb extended will point to the effective north pole of the solenoid magnetic field. The field lines are then drawn in such a way that they flow externally from the north pole toward the south pole at the coil's opposite end. Externally, the solenoid field has a very similar shape to that of a bar magnet (Figure 25.16). The magnetic field lines are continuous and extend down through the centre of the solenoid to create the uniform field.

The solenoid can be made into an **electromagnet** if the hollow core contains a magnetically soft material. The core concentrates the lines of force and increases the magnetic strength through the induction principle. Iron–nickel alloys are the most commonly used material in the physical construction of electromagnet cores, where they can increase magnetic field strengths several hundred times above that produced by the solenoid itself. The greatest advantage of electromagnet assemblies is that the magnetic field can easily be switched on or off simply by breaking the flow of current through the coil turns. They have many practical applications.

Photo 25.3
A solenoid and magnetic field.

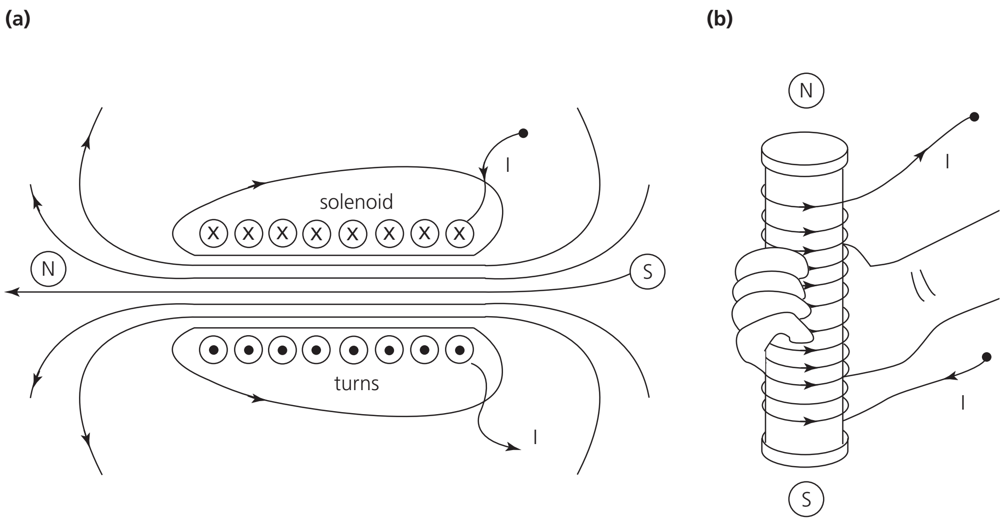


Figure 25.16
(a) Field of a solenoid;
(b) the grip rule.

Questions

10 Describe the difference between the angles of declination and inclination when referring to the Earth's magnetic field. What are the Van Allen belts?

11 Determine the magnetic field strength at a distance of 15 cm from a wire if it carries an electric current of **(a)** 5.5 A north; **(b)** 25 A west.

12 A solenoid has a length of 20 cm and contains 8000 turns. If it carries a current of 15 A, what is the magnetic field strength at the centre of the coil?

PHYSICS FACT

At the Boyne Island smelter at Gladstone, huge currents are used to extract aluminium from its ore by electrolysis. The currents are so great that workers wearing steel-capped shoes find their feet pulled in the direction of the magnetic field. When former Prime Minister Paul Keating visited the smelter his car wouldn't start until they pushed it away from the smelter. (The electronic ignition was affected by the field.) The huge magnetic fields are so big that they can drag large iron objects along the floor.

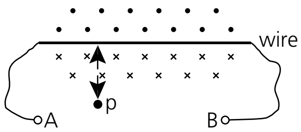


Figure 25.17
For question 13.

13 In Figure 25.17, the direction of the magnetic field around a current-carrying wire is shown. If the magnetic field at point p is 1.5×10^{-3} T and it is 1.0 cm from the wire, what is the magnitude and direction of the current in the wire AB?

14 Determine the direction and magnitude of the magnetic field at points P_1 and P_2 in the diagram shown in Figure 25.18.

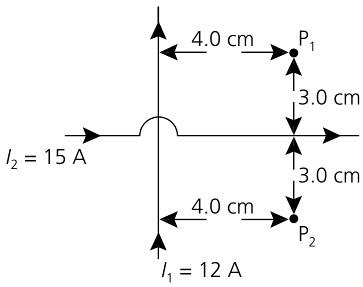


Figure 25.18
For question 14.

Electromagnet applications

There are many examples of electromagnets in devices around school, home and industry. Special electrical switches that are controlled by small actuating electromagnets are called **solenoid switches**. In these devices usually a small electric current flowing in the solenoid coil of the switch opens and closes a set of contacts that are designed to carry much larger currents from the household 240 V supply. A good example is the switching solenoid in a washing machine that 'clunks' at certain stages of the washing cycle when the electric motor controlling the washing and spin-drying functions turns on. The 'clunk' is the electromagnet core moving and opening or closing heavy-duty contact points. A smaller example is the typical **DC bell** or buzzer that you will find in the physics laboratory. In this circuit the DC current initially begins to flow through the electromagnet, which magnetises the core and attracts the iron armature bar through induction. This in turn breaks the contact point, which switches off the electromagnet, allowing the armature bar to swing back. The to and fro motion occurs rapidly and causes the ringing sound as the hammer strikes the bell.

Electromagnetic household **overload circuit-breakers** were discussed in Chapter 22. These devices automatically break the 240 V circuit if the current flowing through the sensing solenoids becomes excessive due to an inappropriate number of appliances being operated on the one household power circuit.

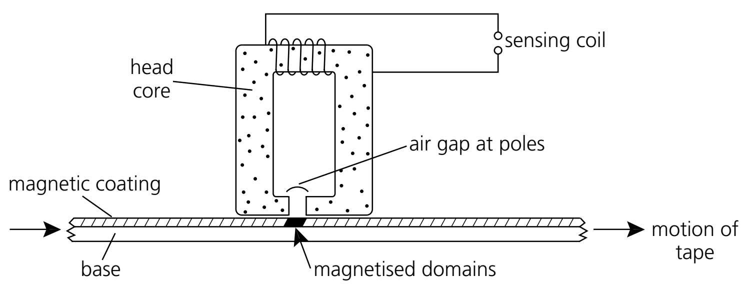


Figure 25.19
A recording head.

Figure 25.19 illustrates the basic mechanism of an electromagnetic **write head**, which forms the basis of the writing or recording head of cassette tape decks, video recorders, and computer floppy and hard disk drives. The magnetic tapes or computer disks are made of a flexible plastic base coated with a magnetic iron oxide particle layer. When small, rapidly changing electric currents are sent to the recording head by the drive electronics, the small electromagnet energises and develops a magnetic field at the air gap poles. This magnetic field induces a pattern of domain alignment in the magnetic layer of the tape or disk and this represents the recorded information or stored computer data. The tape or disk itself passes rapidly across the air gap of the head so that a great amount of information can be recorded. If this tape is subsequently passed across an even more sensitive **read-head**, the pattern of aligned domains induces electric currents back into the pickup electromagnet coil and this represents the AC current, which can be amplified and converted back into sound or data required by a computer program.

Also associated with computers are the devices called **dot-matrix printers**. These printers either use a 9 pin or a 24 pin print head, which contains either 9 or 24 small electromagnetic solenoid print hammer rods. The devices are called impact printers because the

Ch25-Walding 4th 25/8/04 12:37 PM Page 562

metal firing pin is pushed forward as the electromagnetic firing solenoid is controlled by the digital firing pulse. The pin strikes or impacts onto the paper through the print ribbon and makes a mark. Very high quality print type can be produced, with each letter being formed by a certain combination of all 9 or 24 pins being fired simultaneously. The solenoids are highly sensitive and designed to fire rapidly, with recovery times in the order of milliseconds. The typical life of printer heads is usually several hundred million strokes per firing pin and they have a reputation for very low maintenance.

PHYSICS UPDATE

Most modern printers use ink-jet technology. Find out if these use electromagnetic firing.

NEI Activity 25.3 SOME MORE APPLICATIONS

Numerous other technology applications of magnets and electromagnets exist around us. Research the method of operation of the following:

1. Large electromagnetic cranes.
2. Magnetic switches for intruder alarms.
3. Library security tags on books and sensing coils at the registration desk.
4. Magnetic levitation of certain diamagnetic materials as a research oddity.
5. Magnetic levitation on very fast trains.

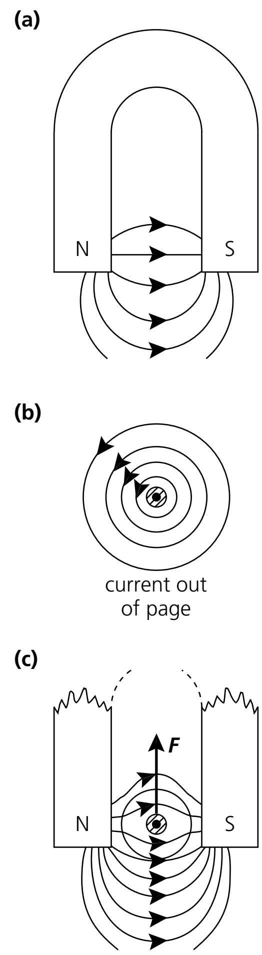
25.5 FORCES ON CURRENT-CARRYING CONDUCTORS

We have seen that a conductor carrying an electric current produces a magnetic field around it. It is logical to expect that this field would interact, by attraction or repulsion, with another magnetic field of a permanent magnet placed within it. The fact that this occurs was originally verified by Michael Faraday and is commonly referred to as the **motor principle** because it is the basis of operation of modern electric motors.

The diagrams of Figure 25.20 illustrate the nature of the force exerted on a current-carrying conductor placed within a magnetic field. The action of the force on the conductor is described by the following statement:

> *The force exerted on a current-carrying conductor within an external magnetic field is mutually perpendicular to both the direction of the current flow and the direction of the magnetic flux lines.*

Figure 25.20
The motor principle force:
(a) external field; **(b)** conductor field;
(c) the force on the conductor.

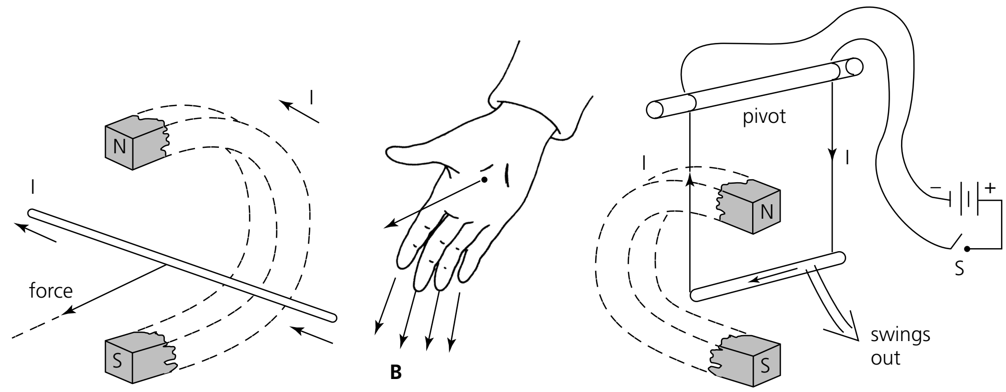


The reason for the force exerted in Figure 25.20 can be seen from the concentration of the field lines on one side of the conductor and the cancellation of field lines on the opposite side. Once again it is possible to use a simple right-hand rule (RHR) to predict the direction of the force, as shown in Figure 25.21. This rule is often called the RH **motor rule**. A simple, freely pivoting conductor set up between the poles of a horseshoe magnet is enough to demonstrate the nature and direction of the force. Point your fingers in the direction of the external field lines, your extended thumb in the direction of the flow of conventional current and the palm of your hand pushing in the direction of the induced force.

The magnitude of the force on a current-carrying conductor can be shown to depend on four factors:

- The strength of the external field, $\boldsymbol{B}$, in tesla.
- The magnitude of the current, I, in amps.
- The length, L, of the conductor within the magnetic flux in metres.
- The angle, θ, of the orientation of the conductor to the lines of magnetic flux in degrees.

Figure 25.21
The right-hand motor rule.



NOVEL CHALLENGE

When you 'jump start' a car with a flat battery, the leads from the good battery are connected to the same polarity terminals of the flat battery. Will these leads move together or apart when the charge flows?

Hence, the force, $\boldsymbol{F}$, in newtons (N) will be given by:

$$\boldsymbol{F} = \boldsymbol{B}\, I\, L \sin\theta$$

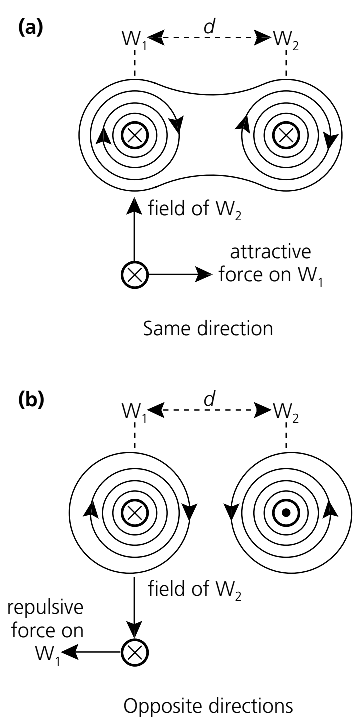
Parallel conductors carrying current

If two conductors are placed within close proximity of one another running in parallel directions, then, as they are made to conduct electric current, the magnetic field of the first conductor will produce a force on the second current-carrying conductor and vice versa. This is shown in Figure 25.22. The diagram illustrates the direction of the forces acting mutually on each conductor. Two force cases may exist, which can be identified by the use of the right-hand rule.

Note that:

For two wires each carrying current in the same direction, the force is attractive, while for two wires carrying currents in opposite directions, the force is repulsive.

Figure 25.22
Forces on parallel conductors.



The force between two wires of length L, separated by a distance, d, carrying currents I_1 and I_2 respectively, can be found using the relation for the field produced at the centre of the left-hand wire, 1, by the right-hand wire, 2.

$$\boldsymbol{B}_2 = \frac{kI}{d}$$

where k is the magnetic constant, and the relation between the force on wire 1, its current and the magnetic field due to wire 2, $\boldsymbol{F}_1 = \boldsymbol{B}_2 I_1 L$. Hence, the magnitude of the force acting on wire 1 will be:

$$\boldsymbol{F}_1 = \frac{(kI_2)}{d}\, I_1 L = \frac{kI_1 I_2 L}{d}$$

but this is the same magnitude as and the opposite direction to the mutually acting force on wire 2.

By measuring very accurately the value of this mutually acting force on a pair of current-carrying conductors, a precise definition of the electric current unit, the **ampere**, has been established as a standard in SI units. Consequently, the ampere is defined as:

The current which, when flowing through two infinitely long, parallel straight thin wires, placed one metre apart in a vacuum, produces a force of 2×10^{-7} newtons on each metre of wire.

Questions

15 A current-carrying wire passes perpendicularly through a magnetic field as shown in Figure 25.23. If the magnetic field strength is 1.5×10^{-3} T and the wire carries a current of 8.0 A, calculate the force on the wire in both magnitude and direction.

16 A conductor of length 8.5 cm is placed between the poles of a large magnet as shown in Figure 25.24. If the wire conductor carries a current with direction shown of 25 amperes, calculate the force on the conductor.

17 In each of the situations of Figure 25.25, predict the directions of the induced force acting on the conductor if the direction of current flow is as shown.

Figure 25.23
For question 15.

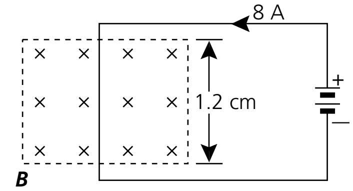


Figure 25.24
For question 16.

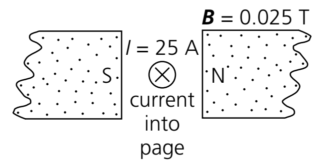
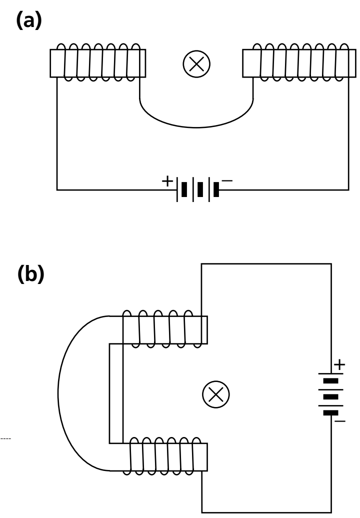


Figure 25.25
For question 17.



The loudspeaker

A **moving coil loudspeaker**, as is commonly found in small transistor radios or home stereo systems, is designed to change electrical signals from the output of an amplifier back into sound waves. The device is an output transducer and relies on the force produced by a flowing current in a conductor within a magnetic field. A movable coil attached to a strengthened paper cone is placed over the central shaft of a permanent magnet. The magnetic field is radial, so that any movement produced will be backward and forward as shown. The amplifier output supplies variable frequency currents and as these flow through the speaker voice coil it is forced to vibrate in the same way as the currents. The paper cone also vibrates backward and forward, moving the air and producing sound waves that match the amplitude and frequency of the original electric current signals. (Refer back to Section 16.10.)

25.6 ELECTRIC MOTORS AND METERS

The force produced on a current-carrying conductor leads to some very important practical applications in physics and engineering. In this section we will firstly look at the combined turning effect of a magnetic field on a coil of conducting wire, called **electromagnetic torque**. The word torque comes from the Latin *torquere* meaning 'to twist'. We will use this concept of torque as the basis for understanding both electrical measurement meters, such as the voltmeter and ammeter, as well as the basic principles of operation of electric motors. The action of the motor principle on a coil that is free to rotate in a magnetic field on an axis produces the turning torque and can be regarded as the rotational equivalent of a force. The magnitude of a torque, τ (tau), where τ is the Greek letter symbol used, is given by the product of the force involved and the perpendicular distance from the line of action of the force to the axis of rotation:

$$\tau = \boldsymbol{F} \times d_{\perp}$$

Units of torque are newton metres (N m).

Refer to Figure 25.26. When torque is applied to a coil rotating in a magnetic field, you can see that each side of the coil has a motor force, $\boldsymbol{F} = \boldsymbol{B}IL$, acting on it to cause a rotation about the coil axis. Most importantly the coil side AB has a force acting upward whereas side CD has a force acting downward so that the torque causes a clockwise rotation as seen from the front of the diagram. The magnetic flux runs from left to right or north to south.

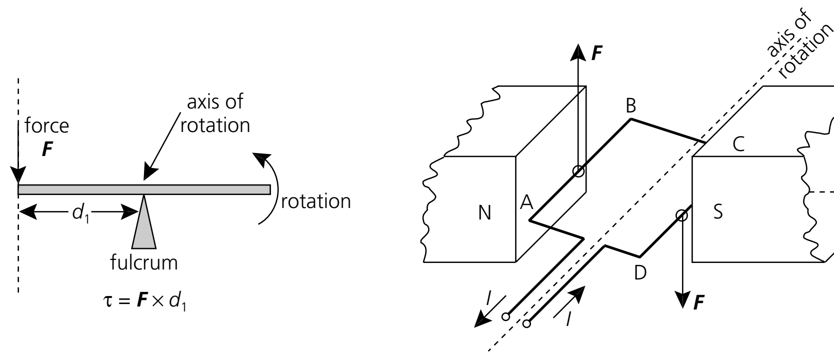


Figure 25.26
Torque on a coil.

It is interesting to consider what happens when coil side AB rotates around to the top of the arc and begins to come down again. With the current flowing from B toward A, this part of the coil now has a force acting upward and the section CD has an opposite force acting downward. The coil now rotates anticlockwise as seen from the front of the diagram. The fact that a coil can only rotate through 180° before the torque changes direction is a problem and prevents the coil from continuously rotating. It can be overcome and we will see how quite soon! It should be realised that the coil sides BC and DA are, in fact, always parallel in alignment to the magnetic field lines and thus will have no force acting on them at all and consequently will not play a part in producing a torque on the rotating coil.

Mathematically, if the coil has side length AB and side width BC in metres, then about the axis of rotation, as shown in Figure 25.26, with the coil lying horizontally at first:

$$\tau_{tot} = \tau_{AB} + \tau_{CD}$$

but using $$\left(\boldsymbol{F}_{AB} \times \frac{BC}{2}\right) + \left(\boldsymbol{F}_{CD} \times \frac{BC}{2}\right) = \tau_{tot}$$

$$(\boldsymbol{F}_{AB} + \boldsymbol{F}_{CD}) \times \frac{BC}{2} = \tau_{tot}$$

$$2BI \times AB \times \frac{BC}{2} = \tau_{tot}$$

$$BIA = \tau_{tot}, \text{ where area A = AB} \times \text{BC}$$

Thus, the total torque acting on a coil within a magnetic field is given by:

$$\tau = \boldsymbol{B}AIN$$

where $\boldsymbol{B}$ = magnetic field strength threading coil in teslas; A = cross sectional area of the coil in square metres; I = current passing through the coil in amps; N = number of turns on the coil. Note, that at some angle, θ, of the coil to the flux lines, the torque will no longer be a maximum as it was at the horizontal. In fact, the torque at the vertical orientation is zero instantaneously, because the force is acting through the axis of rotation. Thus, a more general formula for the torque acting is given by:

$$\tau = \boldsymbol{B}AIN \cos \theta$$

where θ is the angle in degrees between the coil and the lines of magnetic flux.

Example

A coil contains 20 turns of conductor carrying an electric current of 150 mA. If the plane of the coil is at an angle of 45° to the lines of magnetic flux between the poles of a magnet whose field strength is 5.5×10^{-4} T, determine the magnitude of the total torque on the coil. The coil dimensions are a length of 6 cm and a width of 4 cm.

Solution

- Given $B = 5.5 \times 10^{-4}$ T, angle $\theta = 45^0$, N = 20 turns, current $I = 150 \times 10^{-3}$ A.
- Area of the coil will be $A = l \times b = (6.0 \times 10^{-2} \times 4 \times 10^{-2}) = 2.4 \times 10^{-3}$ m^2.

Use:

$$\tau = \boldsymbol{B}AIN \cos \theta$$
$$\tau = 5.5 \times 10^{-4} \times 2.4 \times 10^{-3} \times 150 \times 10^{-3} \times 20 \times \cos 45^0$$
$$\tau = 2.8 \times 10^{-6} \text{ Nm}$$

Note that the direction of the torque could only be established through the motor rule, knowing the direction of current flow.

Galvanometers

Both ammeter and voltmeter electric circuit measurement meters make use of the motor principle. They are a form of **galvanometer** called the moving coil galvanometer, which uses the current flowing through a coil placed in a magnetic field to generate a torque. The current causing the torque is the circuit current being measured or at least a small portion of it, as described in the calibration discussion in Chapter 22. The torque winds up a small spring until the restoring force of the spring balances the torque. A needle attached to the coil registers the final deflection position and allows the reading on a calibrated instrument scale. The larger the circuit current being measured, the larger the needle deflection. The moving coil galvanometer is illustrated in Figure 25.27.

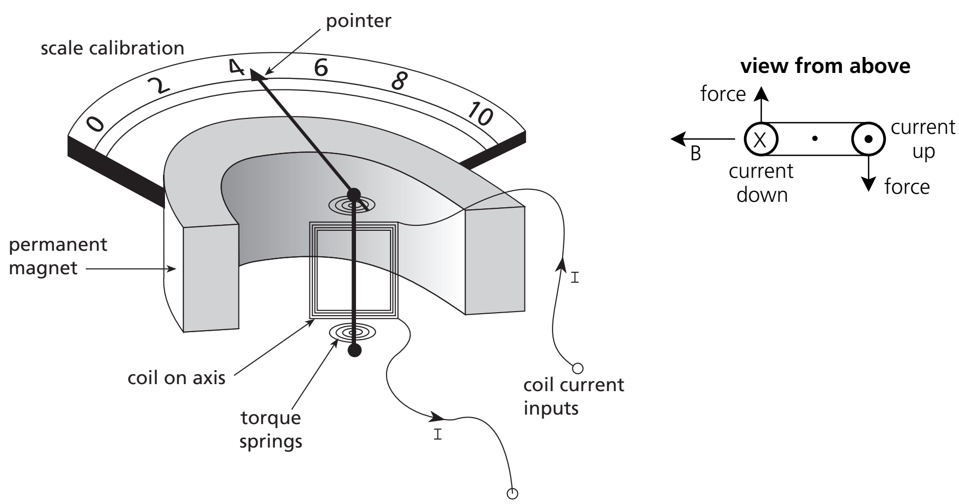


Figure 25.27
The galvanometer.

The coil usually moves freely around a fixed soft iron cylinder or armature. When used in conjunction with curved magnetic pole pieces, which produce a radial magnetic field, the galvanometer is much more sensitive. This also ensures that some of the field always lies at right angles to the coil, which means that the turning force on the coil is almost independent of the coil's position. It was John Schweigger of Germany, in 1928, who constructed the first working galvanometer of the moving coil design.

Electric motors

Electric meters use the torque on a current-carrying coil in a magnetic field to move a scale pointer. The turning force, as we've seen, would cause the coil to reverse on itself, if the coil was allowed to rotate too far. A **DC motor** is a device that makes use of the motor principle

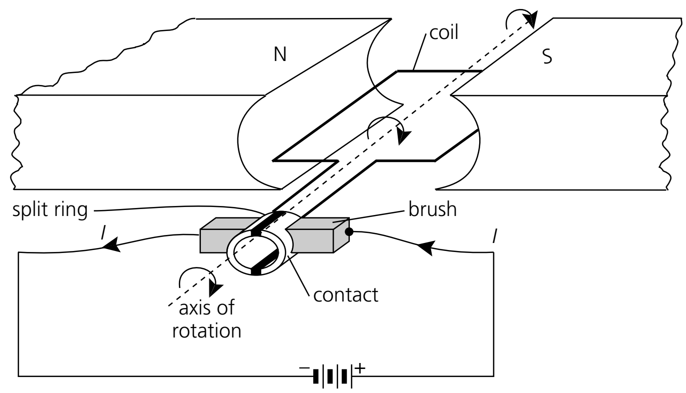


Figure 25.28
A split-ring commutator.

but contains a special switch assembly on the rotating coil shaft that allows the direction of the current through the coil to be reversed every 180°. This switch is called a **split-ring commutator** and is shown in simplified form in Figure 25.28.

This commutator consists of two semicircular contacts mounted on the motor shaft. These contacts are connected directly to the turns of the motor coil. Electric current from an external circuit flows into contact via a **carbon brush**, which presses against the contact and slides over it as the shaft rotates. Current leaves the coil through a second brush on the opposite side of the shaft. As the motor rotates through the position where the coil lies across the field, each brush loses contact with one side of the commutator split ring and almost immediately reconnects to the other commutator contact. This occurs twice in every single rotation, at which time the current flow through the coil reverses and the motor continues to be rotated under the action of a torque of constant direction. The carbon brushes will eventually wear out due to friction but can easily be replaced in more expensive electric motor designs. In cheaper motors, such as found in some toys, the brushes are actually just metal sliding contacts to the commutator and will eventually wear down so that the whole electric motor becomes useless.

The earliest known examples of a patent for an electric motor is US patent No. 132, granted on 25 February 1837, to Thomas Davenport of Brandon, Vermont. The patent was titled, 'Improvements in propelling machinery by magnetism and electromagnetism'. According to the description contained in the specification, '*the motor, which is intended to be driven by a galvanic battery, is constructed on sound electromagnetic principles*'.

In large commercial DC motors the permanent magnets are usually replaced with electromagnets because these can produce much greater magnetic field strengths. The electromagnetic coils are called the **stators** and are fixed in relation to the rotating armature windings, which are wound onto a segmented commutator. Multiple winding sets are used on a laminated armature in order to obtain a very smooth output torque from the motor.

Three major types of DC motors exist: permanent-magnet, shunt-wound and compound-wound motors.

Permanent-magnet DC motors have an armature winding and use permanent magnets for the field. In this type of motor you need to be able to reverse the connections to the armature winding in order to have the motor run in reverse. This can be done in practice with a contactor (high current relay) or four power modules in a bridge configuration, which would allow for electronic reversal of the armature voltage.

Shunt-wound DC motors have an armature winding via brushes and commutator, and a separate field winding that provides a magnetic field in which the armature rotates. Forward, reverse or no current is applied to the field winding in order to control rotation direction in the motor. It is usual for the field winding current to be much less than the armature current, but the field winding coil has a higher inductance so it stores a lot of energy.

In the **compound-wound** DC motor there is an armature winding, a field winding that is in series with the armature and a separate shunt field winding. This configuration allows even more motor rotation control but is more expensive to manufacture.

The use of DC motors in wheelchairs and motorised 'gophers' has enabled people with physical disabilities to move about with greater freedom. A typical motorised gopher incorporates a pair of 12 V motorcycle-type batteries connected to a permanent magnet 24 V DC motor of high torque characteristics. A switch system allows the connection of either 12 V or 24 V for low and high speed operation. Braking is usually provided by the torque of the DC motor itself and wheel rotation is by a common chain drive system. These vehicles are becoming more common in the community, providing assistance to arthritis sufferers and others with immobilising disabilities. The 'gopher' vehicle has a 24 V DC motor made by the Rae Corporation of McHenry Illinois, USA, and is rated at a maximum of 9.88 A, 354 rpm and a torque of 0.4 N m.

Photo 25.4
Gopher vehicle.

Compact disc players need to provide a constant linear velocity of the pickup laser head across the surface of the CD. Because the encoded information on the disc has to be read at a constant rate, the disc motor has to spin faster when the head is at the centre compared with at the outer rim of the disc. A typical CD motor needs to spin the disc at between 200 and 500 rpm. This is achieved through microprocessor control electronics and a variable speed DC servo-motor. This is a good example of the interaction between modern digital electronics and motor technology.

Stepper motors are used in robotics and control systems. A stepper motor is designed to rotate through a given series of angles in small steps when driven by pulsed DC. Again, very small, powerful permanent magnets and electromagnets are used as the basis for stepper motors.

Victoria was the first state in Australia to use electric trains in 1918. Modern Victorian electric trains use 1500 V DC motors rated at 124 kW. In a typical train of about six carriages there will be 16 motors developing a total power output of about 2 MW. Electric trains are now commonplace across other parts of Australia and world-wide. The most common motor design for their use is an AC induction motor, which is more efficient and, because brushes are not necessary, requires less maintenance.

AC induction motors

The **AC induction motor** rotates because of the interaction of magnetic fields of the rotor and the stator. In this type of motor, the stator windings are connected to an AC supply in one or three phase form. By applying a voltage across the winding, a radial rotating magnetic field is formed. The rotor has layers of conductive strands along its periphery. These strands are short-circuited to form conductive closed loops. The rotating magnetic fields produced by the stator induce a current into the conductive loops of the rotor. Once that occurs, the magnetic field causes forces to act on the current-carrying conductors, which results in a torque on the rotor.

The simplicity of the AC induction motor is that the currents in the rotor do not have to be supplied by commutator, as they do in a DC motor. The velocity of the rotating magnetic field of the stator can be calculated with the formula below:

$$V = 120 \,.\, f\,/\,p, \text{ where } p \text{ is the number of poles and } f \text{ is the frequency}$$

The rotor reacts to the magnetic field, but does not travel at the same speed. The rotor speed actually lags behind the speed the magnetic field. The term 'slip' quantifies the slower speed of the rotor in comparison with the magnetic field. The rotor is not locked into any position and therefore will continue to slip throughout the motion. The amount of slip increases proportionally with increases in load, thus open loop induction motor systems are not particularly stable in rotation speed.

Photo 25.5
AC induction motor.

There is a variety of different types of induction motors, differing mainly by the number of phases and the winding type. Some of the more common names are shaded pole, split phase, capacitor start, two value capacitor, permanent split capacitor, two phase, three phase star, three phase delta and three phase single voltage. We will not get into the differences here. You might like to find out the differences on the Web.

AC induction motors have had greatest use in industrial applications where precise speed control is not needed (such as pumps, fans and conveyors). The induction motor can be connected directly to a 50 Hz or 60 Hz commercial main, making a system very inexpensive. Today, more and more induction motors are being controlled by AC variable speed drives (inverter). These drives can control the frequency of the AC supply fed to the windings, making the induction motor a controlled velocity device more like the DC motor previously discussed. Refer to Photo 25.5 for the internal design of a typical 18 W electric fan motor.

MOVING CHARGES IN MAGNETIC FIELDS 25.7

An electric current is the result of a flow of charge. Therefore, if a current in a conductor experiences a force due to the presence of a magnetic field, then so should a single charge as it moves through a field. The fact that this is so can be easily demonstrated by carefully bringing a magnet up to the screen of a black and white TV. (Best not to do this with a colour TV!) The electron beam producing the TV image will be shifted considerably by the nearby magnet and a distortion of the picture will be seen. The direction of movement of a beam of negative electrons in a magnetic field can be predicted using the right-hand motor rule, but remember that electrons move in a direction opposite to that of conventional positive charges, so you must point your extended thumb in the opposite direction to the travel of the electron beam. An electron beam consists, of course, of many individual electrons and their behaviour when in a constant magnetic field is shown in Figure 25.29.

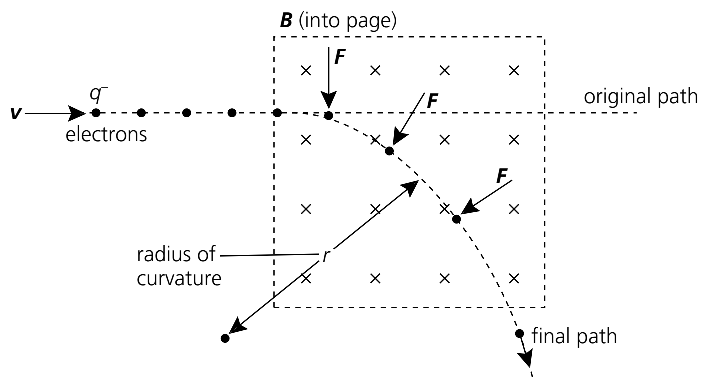


Figure 25.29
Force on charged particles.

The size of the force exerted is found to depend on:

- the size of the charge C in coulombs
- the velocity of the charges in m s^{-1}
- the strength of the field, $\boldsymbol{B}$, in tesla.

Recall that the maximum force produced on a current-carrying conductor is:

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

but $I = \frac{q}{t}$ or rate of flow of charge, hence:

$$\boldsymbol{F} = \frac{\boldsymbol{B}qL}{t} = \boldsymbol{B}q\boldsymbol{v} \text{ where } \boldsymbol{v} \text{ is the particle velocity}$$

The force on the moving charge will be acting constantly to change its direction of travel. In fact, the force is always directed at right angles to the instantaneous direction of travel. Recall that this is the precise requirement for centripetal motion. Thus, if a moving charge enters a magnetic field it will be curved into a circular path of travel. If the strength of the field is strong enough the charged particle may be constrained to move in a completely circular path within the field and may never escape. Mathematically:

$$\boldsymbol{F} = q\boldsymbol{vB} = \boldsymbol{F}_C = \frac{m\boldsymbol{v}^2}{r}$$

where r is the radius of path in m; m is the mass of the charged particle in kg.

Thus, to calculate the radius of curvature of any charged particle in a magnetic field we obtain:

$$r = \frac{m\boldsymbol{v}}{q\boldsymbol{B}}$$

Note that this effect will occur for all charged particles, not just simple electrons. This effect has very useful practical applications, especially where charged particles or ions are concerned.

Example

In the diagram of Figure 25.29 the charged particles are protons of mass 1.67×10^{-27} kg. They enter the magnetic field of strength $\boldsymbol{B} = 3.0 \times 10^{-2}$ T at a velocity of 2.5×10^5 m s^{-1}. Determine:

(a) the force acting on the protons

(b) the radius of curvature of their path in the field.

Solution

(a) Use the charge on the proton as 1.6 ¥ 10–19 C and the formula:

$$\boldsymbol{F} = q\boldsymbol{vB}$$
$$\boldsymbol{F} = 3.0 \times 10^{-2} \times 1.6 \times 10^{-19} \times 2.5 \times 10^5$$
$$\boldsymbol{F} = 1.2 \times 10^{-15} \text{ N upward}$$

(b) Use the radius formula:

$$r = \frac{m\boldsymbol{v}}{q\boldsymbol{B}}$$
$$r = \frac{1.67 \times 10^{-27} \times 2.5 \times 10^5}{1.6 \times 10^{-19} \times 3.0 \times 10^{-2}}$$
$$r = 8.7 \times 10^{-2} \text{ m}$$

The mass spectrometer

The **mass spectrometer** is an instrument used by physicists and chemists to separate gaseous ions or isotopes in a magnetic field according to their masses. The technique allows the measurement of atomic and molecular masses (Figure 25.30). Charged ions are produced by electron bombardment. If a mixture of ions or isotopes of different mass and charge enter the magnetic field region, they will be curved into paths of different radii. The position at which they strike the detector and the number of particles in a particular strike location on the photographic plate or electronic detector can be used to determine relative masses.

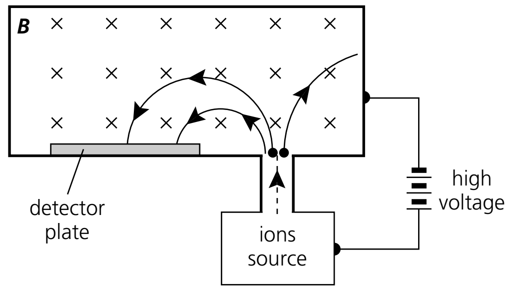


Figure 25.30
Mass spectrometer.

Particle accelerators

In the study of atomic particles, or high-energy physics, scientists require atom-smashing machines. These large **accelerators**, used in places such as Fermilab near Chicago in the USA and the Centre for European Nuclear Research (CERN) in Geneva, are used to study the collisions of highly energetic particles in order to learn about the structure of matter. You might look ahead to Chapter 29. Many of these large accelerators use powerful magnetic fields to deflect the charged particles into circular paths. One of the earliest devices, called a **cyclotron**, is shown in Figure 25.31. It was originally developed in 1930 by Ernest O. Lawrence (1901–58) at the University of California, Berkeley. The cyclotron was used to keep protons and other charged particles moving in circles. For its design, Lawrence won the 1939 Nobel prize for physics. A cyclotron has two D-shaped cavities called dees. As protons cross the region between the dees, a high voltage accelerates them. They move in an ever-increasing radius path until finally, at very high energy, they are allowed to exit the cyclotron and interact with special target nuclei. More modern versions of the original cyclotron are called synchrocyclotrons, synchrotrons, tevatrons and supercolliders.

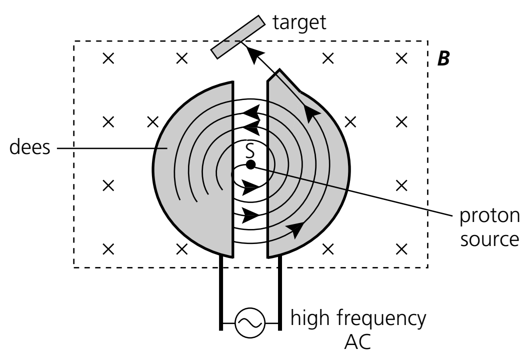


Figure 25.31
Cyclotron accelerator.

Nuclear research

One of the most active areas of research today is in the field of nuclear fusion. Scientists try to create and maintain the nuclear fusion reaction that drives the Sun. In order to do this, physicists need to hold extremely hot deuterium plasma (10^9 K) inside a closed container. Not an easy job! Some success has been gained with devices like the Joint European Torus (JET) experimental fusion reactor, which is basically a **magnetic bottle** container in which the hot charged plasma is confined within a highly evacuated toroidal chamber by extremely powerful superconducting electromagnets. These types of reactors are based on the Tokomak field shape in which the plasma circulates around the torus. Presently these reactors require more energy input than is released during the brief periods of actual fusion that take place; however, they could prove to be an extremely valuable energy resource in the future.

NEI

Activity 25.4 BIG MACHINES

1 Use your library resources to research the following structures and find out their purpose:
- An AC induction motor.
- A Tokomak design nuclear fusion reactor.
- The world's most powerful atom smasher, the 2 km diameter (TeV) proton synchrotron at Fermilab, near Illinois, USA.
- The world's largest particle accelerator, the CERN electron–positron supercollider in Geneva, Switzerland at 26.7 km diameter.

2 Find out what a medical synchrotron instrument is capable of doing. In what field of medical diagnosis and treatment is it used? Are there any located in Australia?

Questions

18 Figure 25.32 shows a simplified DC motor assembly. If length AB is 0.07 m and the magnetic field strength is 0.35 T:
(a) what direction will the coil rotate when a current of 10.5 A flows;
(b) redraw the diagram to illustrate the type and purpose of a commutator;
(c) what is the maximum force acting on coil side AB?

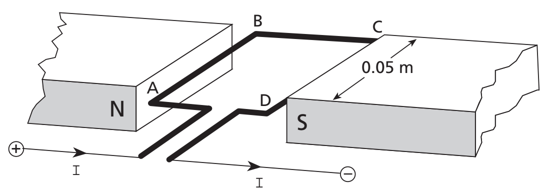
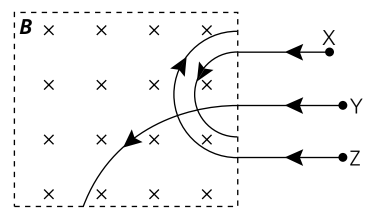


Figure 25.32
For question 18.

19 A charged particle $q = 3.0 \times 10^{-19}$ C and of mass 5.65×10^{-27} kg moves through a uniform magnetic field of 0.35 T at a velocity of 0.001*c*. Deduce:
(a) the maximum force which may act on this particle;
(b) the particle centripetal acceleration.

20 Particles X, Y and Z are all equally charged and moving at the same speeds into a magnetic field oriented as shown in Figure 25.33. The paths of curvature in the field are shown. Hence determine:
(a) the polarity of the charges X, Y and Z from the information given;
(b) the particle that has the greatest mass. Explain your reasoning.

Figure 25.33
For question 20.



Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***21** The Earth's magnetic field at three different locations in Australia is given below. Explain the differences and suggest a reason for the strongest field being at Hobart. Estimate a value for the field strength at Brisbane and at Melbourne. What factors might affect the actual value at these locations?

Darwin	46×10^{-6} T	at dip angle of 39°.
Perth	58×10^{-6} T	at dip angle of 66°.
Hobart	66×10^{-6} T	at dip angle of 73°.

***22** Using the domain theory, explain why you should not keep your constantly-used audio and video tapes under your bed in a cardboard box into which you have to constantly rummage around in order to find the one you want.

Ch25-Walding 4th 25/8/04 12:37 PM Page 573

Figure 25.34
For question 24.

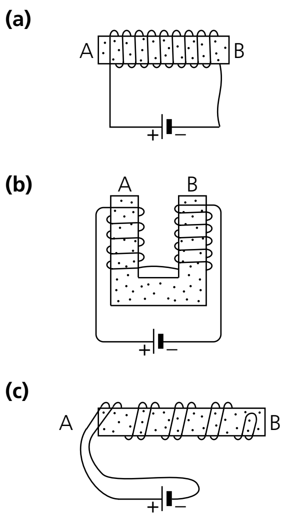
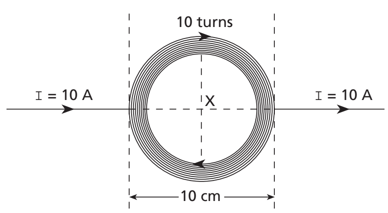


Figure 25.35
For question 28.



***23** What is the total magnetic flux passing through an area measuring 15 cm by 15 cm if the flux density, $\boldsymbol{B}$, is 5.0×10^{-4} T?

***24** The diagrams of Figure 25.34 (**a**, **b** and **c**) illustrate possible electromagnets. In each case state what the magnetic polarity of ends A and B are.

***25** Two parallel conductors are separated by a distance of 1.5 cm. If they carry currents of 2.5 A and 3.5 A respectively, in the same direction, calculate the force acting per metre of length and state the direction of the force as attractive or repulsive.

***26** A power line carries a current of 100 A from east to west. The Earth's magnetic field is 40 μT directed from south to north, inclined downward at 60° to the horizontal. What is the magnitude and direction of the force on each 10 m length of the power line?

***27** A long straight conductor carries a constant DC current in a uniform magnetic field of 30 μT north. Calculate the magnitude and direction of the DC current if at a point 5.0 cm above the conductor the net magnetic field is zero.

***28** Figure 25.35 shows a 10 turn circular loop carrying an input current of 10 A flowing east. If the current circulates clockwise, what is the field strength at point X, the centre of the loop?

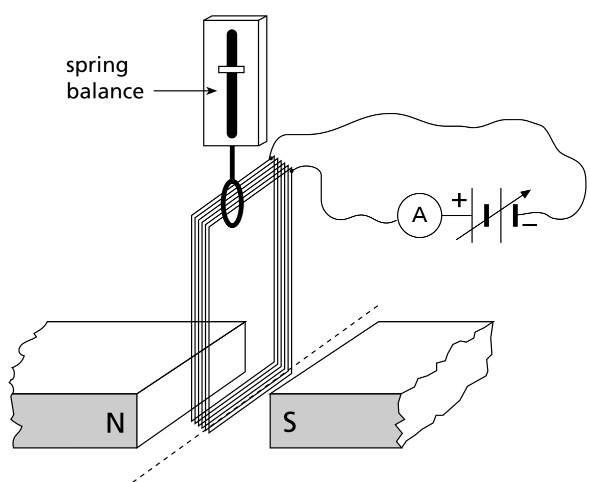
***29** Make neat sketches of the magnetic fields surrounding the following devices:
(a) A horseshoe magnet.
(b) Two bar magnets with their south poles facing each other end-on.
(c) A long multi-turn air-cored solenoid coil.

***30** Explain how a moving coil loudspeaker works. Predict the changes that are necessary to convert a moving coil loudspeaker into a microphone input transducer.

****31** Suppose that your wrist-watch remains undamaged by a magnetic flux density of less than 8.0 T. If you were a tourist in Melbourne, is it safe to walk under the tramway overhead cables, which carry currents of 500 A? Explain your reasons.

****32** A coil of wire is suspended from a spring balance between the poles of two magnets. The coil is rectangular with dimensions 80 cm high and 10 cm wide, and has 100 turns of wire. In an experiment the spring balance readings were recorded for different currents. The apparatus and results are given in Figure 25.36.
Plot a graph of force (N) versus current (A) and answer these questions:
(a) What is the weight of the coil?
(b) What is the magnetic field strength?
(c) Is the current direction clockwise or anticlockwise?
(d) If the current is adjusted so that the balance reads zero, what current flows in the coil and in which direction does it flow?

Figure 25.36
For question 32.



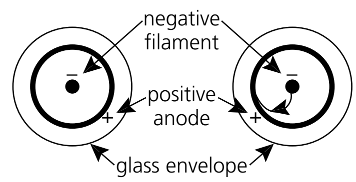
Force (N)	Current (A)
3.5	0.5
5.0	1.5
5.6	2.0
6.8	3.0
7.5	3.5
8.3	4.0
9.5	5.0

Extension — complex, challenging and novel

*****33** In an experiment to determine the mass of an electron, a vacuum valve tube with a central negative filament and an outer positively charged electrode is used to provide a source of electrons. The potential difference between this electrode and the filament is 200 V. This apparatus is placed into the magnetic field of an open solenoid coil so that the electrons leaving the filament are curved into a circular path, as shown in Figure 25.37. Deduce:

(a) the direction of the magnetic field;
(b) the kinetic energy of the electrons reaching the outer electrode;
(c) the equation for the radius of curvature of the path of the electrons assuming circular deflection:
(d) the mass of an electron as determined by this apparatus if $\boldsymbol{B}$ = 0.02 T and a measured radius of curvature is 2.5 mm.

Figure 25.37
For question 33.



*****34** A metal rod XY is 5.0 cm long. It lies on two metal rails connected to a DC supply. The rod and rails are balanced on a flat insulator base in a magnetic field of strength 0.20 T. A current is then passed through the rod causing a downward movement. If a mass, *m*, of 1.0 g is needed to restore the system to a level position, as shown in Figure 25.38, calculate the direction and magnitude of the current in rod XY. Use $\boldsymbol{g}$ = 9.8 m s^{-2}.

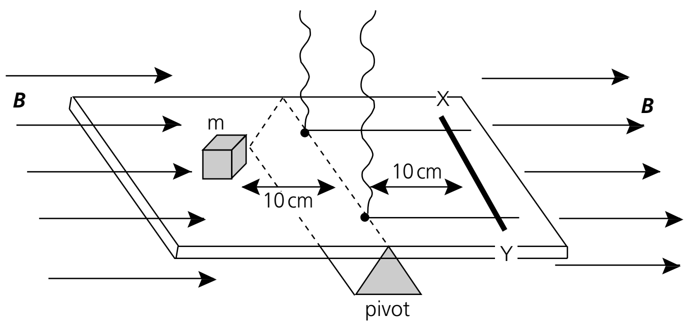


Figure 25.38
For question 34.

*****35** An electric motor consists of a 100-turn coil of wire on a former that is free to rotate in a radial field, as shown in Figure 25.39. The field strength is 25.0 mT and the coil dimensions are 0.32 m length and 0.1 m width. The effective voltage driving the motor is 120 V DC. If each of the turns on the coil has an effective resistance of 0.018 Ω, determine:

(a) the current flow from the power supply;
(b) the effective torque supplied by the motor.

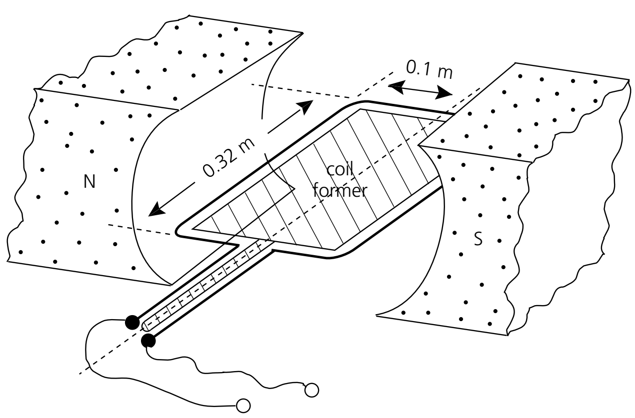


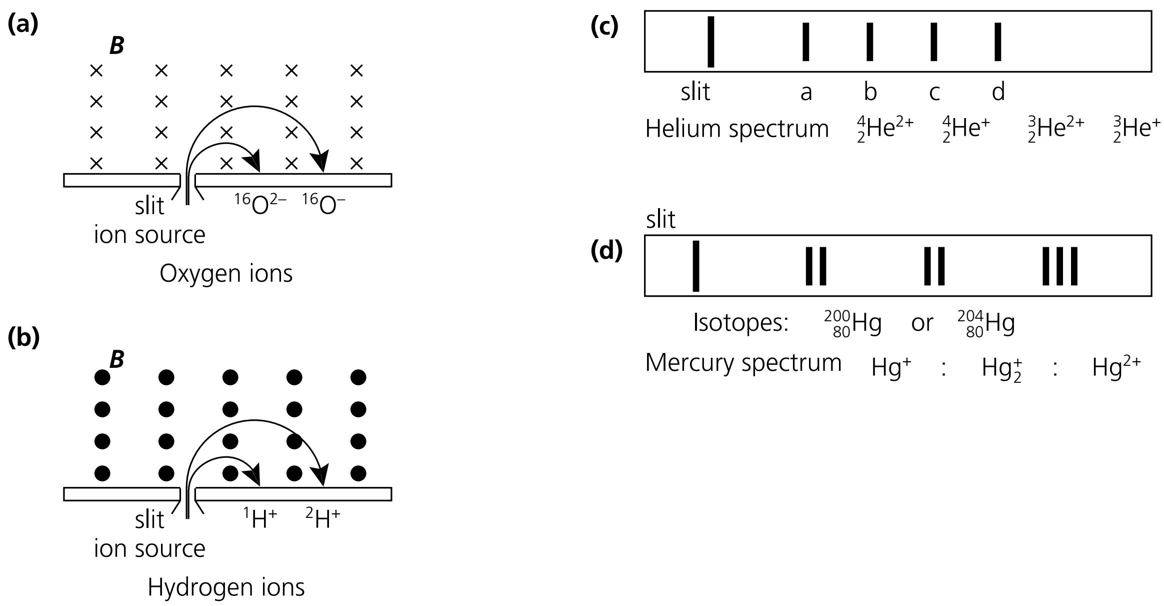
Figure 25.39
For question 35.

***36 The mass spectrometer dates back to the work of J.J. Thompson in England in 1912. A specimen to be analysed is ionised and passed into a magnetic field. Physicists use an *m/e* ratio to label the spectrum produced. Although the ions are bent into a circular path by the field, heavy ions are bent less than light ones and ions with a low charge, for example 1^+, are bent less than ions with more charge, for example 2^+. (Figure 25.40**(a)** and **(b)**.) The *m/e* ratio is the atomic mass divided by the charge. In figure **(a)** the *m/e* ratios are 8 and 16 respectively. The greater the *m/e* ratio, the greater the distance from the slit.

(a) Figure **(c)** represents the spectrum of helium consisting of four ions. Label each ion ${}^4_2He^{2+}$: ${}^4_2He^+$: ${}^3_2He^{2+}$: ${}^3_2He^+$ with the corresponding letter from the spectrum.

(b) Figure **(d)** represents a spectrum from a sample mixture of mercury ions, $Hg^+ : Hg_2^+ : Hg^{2+}$ from either of the isotopes of mercury. Label the formula of the ions to each of the seven spectral lines.

Figure 25.40
For question 36.



CHAPTER 26

Electromagnetic Induction

26.1 INTRODUCTION

Since physicists like Oersted and Ampere had shown that an electric current could produce a magnetic motor effect, **Michael Faraday** predicted that the reverse situation must be true; hence quite a lot of his research was into **electromagnetic induction**. This is where a magnet is made to produce a flowing electric current as the result of an induced EMF. The discovery of electromagnetic induction is credited to Faraday on 29 August 1831 with a device described in an entry in his diary, 'Experiments on the production of electricity from magnetism'. Physicist Joseph Henry is reported to have discovered the phenomenon of self-induction in 1830, but through his failure to publish his research, credit was given to Michael Faraday.

It is interesting to look at Faraday's original notes in his diary referring to his experiments because the discovery of electromagnetic induction has led directly to the development of the rotary electric generator, which converts mechanical motion into electrical energy. This discovery has certainly led to the modern electrical age as we know it today. His diary entry reads:

> Have had an iron ring made, round and one-eighth inch thick and ring 6 inches in diameter. Wound many coils of copper wire round one half, the coil is being separated by twine and calico — there were 3 lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring A. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length, the direction being as with the former coils: this side call B.
>
> Charges a battery of 10 pr plates 4 inches square. Made the coil on B side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from the iron ring). Then connected the ends of one of the pieces on A side with battery; immediately a sensible effort on needle. It oscillated and settled at last in original position. On breaking connection of A side with battery again a disturbance of the needle happened.
>
> Made all the wires on A side one coil and sent current from the battery through the whole. Effect on needle much stronger than before.

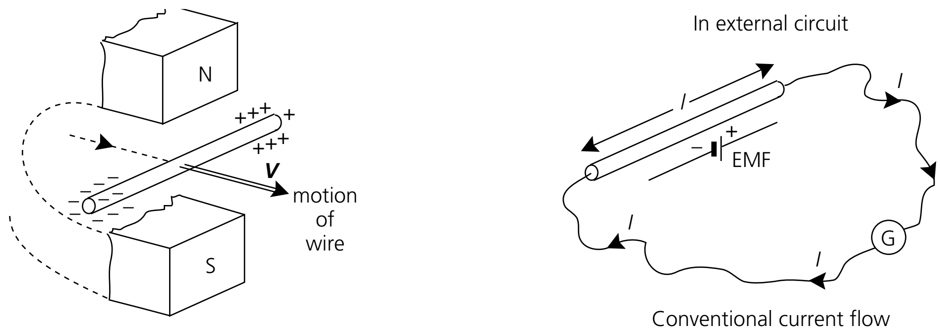
These words of Michael Faraday describe the very first transformer experiment and really represent the start of electrical technology. In this chapter we will look closely at electromagnetic induction, generators and transformers, as well as electrical power transmission.

26.2 LAWS OF ELECTROMAGNETIC INDUCTION

Faraday's law

As we found in the previous chapter, when a conductor, such as a wire, moves through the pole gap of a magnet, the electrons in the wire that are free to move will experience a force along the length of the wire (Figure 26.1). As electrons shift to one end of the wire, we are left with a net excess of negative charge at that end and a net excess of positive charge at the other. This leads to a potential difference or EMF across the ends of the wire and current will flow in any external circuit, which can be shown with a sensitive galvanometer. The direction of flow can be determined using the right-hand rule.

Figure 26.1
Electromagnetic induction.



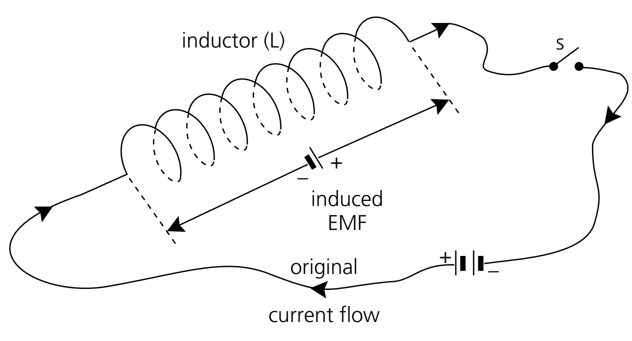
Notice in Figure 26.1 that for a maximum induced EMF and its consequent induced current, the conductor needs to be moved perpendicular to the lines of magnetic flux. We say that the magnetic flux lines are being cut in a perpendicular direction for maximum induced voltage. Alternatively, there will be no induced voltage across the ends of the wire if the movement of the conductor is parallel to the lines of flux. It is also important to realise that the effect also occurs if the conductor is held still and the magnetic field is moved perpendicular to the wire. Hence, what is really important is relative motion between the magnetic field and the conductor. This principle is often stated as Faraday's law of electromagnetic induction:

> *When the magnetic field in the region of a conductor changes, an electromotive force or EMF is induced across the ends of the conductor. If the conductor is made part of a complete circuit then an induced current will flow.*

An easy way to visualise this rule in terms of the separation of charge is to imagine a small test positive charge sitting within the conductor. As the conductor is moved this test charge moves in a direction as given by the thumb in the RH motor rule. The extended fingers point in the direction of the magnetic flux lines and the palm pushes in a direction that shows the way the test positive charge itself moves in the wire. Hence, the polarity of the EMF is established within the conductor and this will determine the direction of flow of conventional current in any external circuit, as shown in Figure 26.1.

Self-induction

Figure 26.2
Inductor self-induction.



Another method of varying the magnetic field is by using an electromagnet that provides a variable magnetic field. This method has greater application in electric generators and in transformers and will be discussed later in the chapter. Before we consider Faraday's law mathematically, consider another interesting situation. When a current is made to flow through any conductor, especially a solenoid coil, the external magnetic field produced will itself induce a voltage across the ends of the coil. This is called **self-induction** and was discussed in relation to inductor components in Chapter 23, Electronics. This self-induced voltage is always opposite in direction to the applied voltage causing current to flow through the solenoid coil in the first instance (Figure 26.2). The induced voltage tends to limit the original current, resulting in a form of resistance. Electric self-induction is analogous to mechanical inertia. An induction or choke coil tends to smooth out varying currents in a circuit just like a flywheel tends to smooth out the jerky rotation of an engine. The amount of self-inductance of any coil is measured in a unit called the henry, named after Joseph Henry. The self-inductance property of an inductor coil is determined solely by the geometry of the coil and by the magnetic properties of its core. You might like to refer back to the section on inductors in Chapter 23.

Faraday's law quantitatively

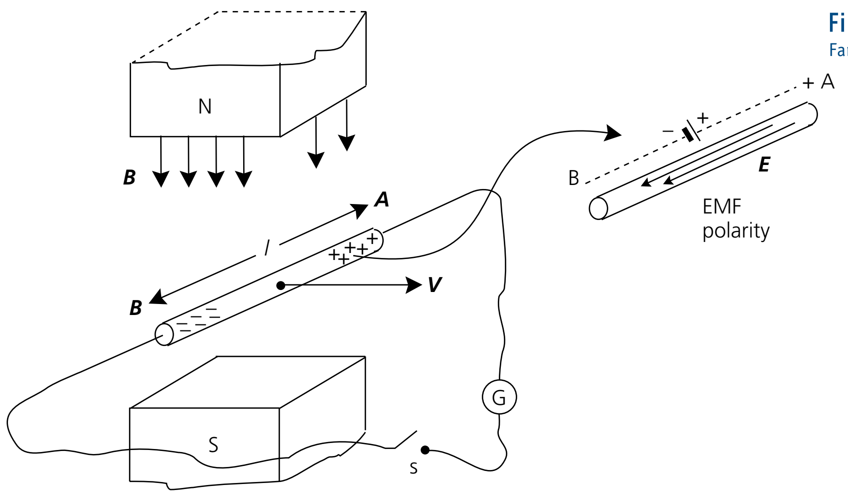


Figure 26.3
Faraday's law.

Let us now consider Faraday's law mathematically. Refer to Figure 26.3, showing a conductor in motion between the poles of a horseshoe magnet. Consider the free electron charges within the metallic conductor of length l being moved perpendicular to the field lines at a velocity $\boldsymbol{v}$. The magnetic force acting on each moving electron is given by:
where $\boldsymbol{B}$ = magnetic field strength in teslas; q = charge on the electron in coulombs; $\boldsymbol{v}$ = velocity of motion in m s^{-1}.

$$\boldsymbol{F}_\text{B} = q\boldsymbol{v}\boldsymbol{B}$$

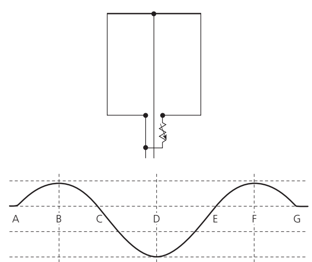
This force has a direction such that the electrons are pushed towards end B as the conductor is moved. This leaves end A equally charged but positive. As the charges build up at each end, AB, of the conductor, an electric field $\boldsymbol{E}$ is set up within the conductor, which begins to oppose the free flow of electrons. Movement of electrons and separation of charge will continue to occur until the magnitude of the electric force is equal to the magnetic force. That is, until:
but as the electric force is given by:

$$\boldsymbol{F}_\text{E} = \boldsymbol{F}_\text{B}$$

$$\boldsymbol{F}_\text{E} = q\boldsymbol{E}$$

NOVEL CHALLENGE

Enemy submarines used to be detected by electromagnetic induction. The process involved laying long lengths of electrical cable on the seafloor at the entrance to important harbours. Submarines, like all ships, become magnetised as they are being built and as they travel though the Earth's magnetic field. When they pass over these loops of cable a small voltage is induced (in the order of microvolts) and this is recorded at the shore station. One of these shore stations still exists on Bribie Island (see Photo 26.1). The layout of the loops is as follows. Imagine that a submarine was magnetised north on its underside (as they mostly were), and it passed over the loop arrangement from left to right. The following 'magnetic signature' would be obtained if a CRO was connected to the two wires at the bottom.



Explain how the location of the ship in its passage over the loops matches with the signature. More information, photos and a solution can be found on the textbook website under the link 'antisubmarine loops'.

Photo 26.1
Shore station at Bribie Island, Queensland.

hence:

$$qE = qvB, \quad \text{or} \quad E = vB$$

but as the electric field strength is also given by the relationship:

$$E = \frac{V}{d} = \frac{\text{EMF}}{L}$$

where EMF is the **induced voltage** in volts and L is the conductor length in metres.
We can arrive at an expression for the magnitude of the induced EMF as:

$$\text{EMF} = BLv$$

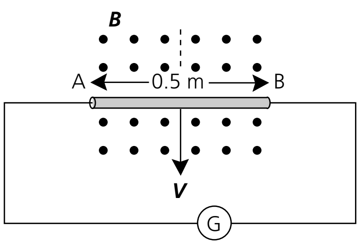
or if the conductor is not at right angles to the field, the net EMF is reduced, becoming:

$$\text{EMF} = BLv \sin \theta$$

If the switch in Figure 26.3 is closed, conventional current will flow from end A through the galvanometer to end B but, again, only while the conductor is physically being moved in relation to the magnetic field.

Example

Figure 26.4



Consider Figure 26.4, showing a conductor of length 50 cm moving at right angles to a magnetic field directed out of the page as shown. If the magnetic field strength is 5.6 mT and the conductor is moving at a velocity of 4.5 m s^{-1} across the flux lines, calculate:

(a) the induced EMF across the conductor AB;
(b) the direction of current flow in the galvanometer circuit AGB.

Solution

(a) Use the equation for induced EMF:

$$\text{EMF} = BLv$$
$$\text{EMF} = 5.6 \times 10^{-3} \times 0.5 \times 4.5$$
$$\text{EMF} = 1.26 \times 10^{-2}\ \text{V}$$

Note that the polarity will be A positive, B negative and thus:

(b) in the external circuit conventional current will flow from A through G to B or anticlockwise around the external circuit.

Solenoids

Figure 26.5
Induction with a solenoid.

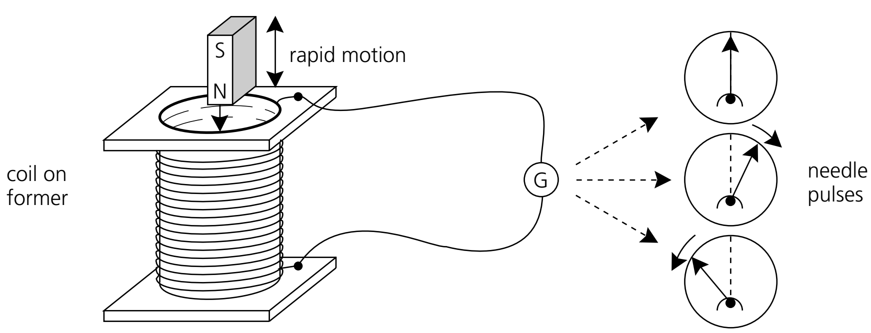


Figure 26.5 illustrates a solenoid connected to a galvanometer. This equipment is easy to obtain in the laboratory and you might like to set it up as a demonstration while going over the next bit of theory. An interesting effect is noted on the galvanometer as a bar magnet is rapidly inserted into, and withdrawn from, the air core of the solenoid. The needle of the meter is seen to flick one way and then the other, as shown. How can we explain this in terms of induction? Recall that relative motion between a conductor and a magnetic field is important if an EMF is to be induced. In this apparatus, the coils of the solenoid cut the lines of magnetic flux and while the magnet is moved a pulse of electric current will be produced. When the magnet is withdrawn, a second, opposite polarity pulse is induced as the solenoid again cuts the lines of flux. When the coil and magnet are stationary, even if for an instant, there is no induced current and the galvanometer needle returns to its zero point. A quite sensitive galvanometer is needed to show this effect, as the EMF and currents produced are very small.

Lenz's law

In the operation of the apparatus of Figure 26.5 it will be observed that the direction of the galvanometer needle pulse of current is reversed when the magnet is pushed into and pulled out of the solenoid. It will also be noticed that if the north pole of the magnet is pushed into the solenoid firstly, and then the magnet reversed and the south pole pushed in, the directions of the galvanometer needle deflections are reversed also. Check this out. Furthermore, if a very powerful magnet is used and a very large solenoid coil with thousands of turns, it will be noticed that considerable force and effort would be needed to push the magnet into the solenoid at all! It appears that nature is trying to prevent, or oppose, the induction of current flowing through the coil. The Russian physicist Heinrich Lenz (1804–64) first explained the direction of the induced current in a solenoid coil as the result of a changing magnetic field. He used the notion of nature trying to oppose any applied force. **Lenz's law**, as it is referred to today, really states that nature does not provide something for nothing! In the field of electromagnetics, Lenz's law states that:

> *The current induced in a conductor by a changing magnetic field is in such a direction that its own induced magnetic field opposes the change that produced it.*

Refer to Figure 26.6.

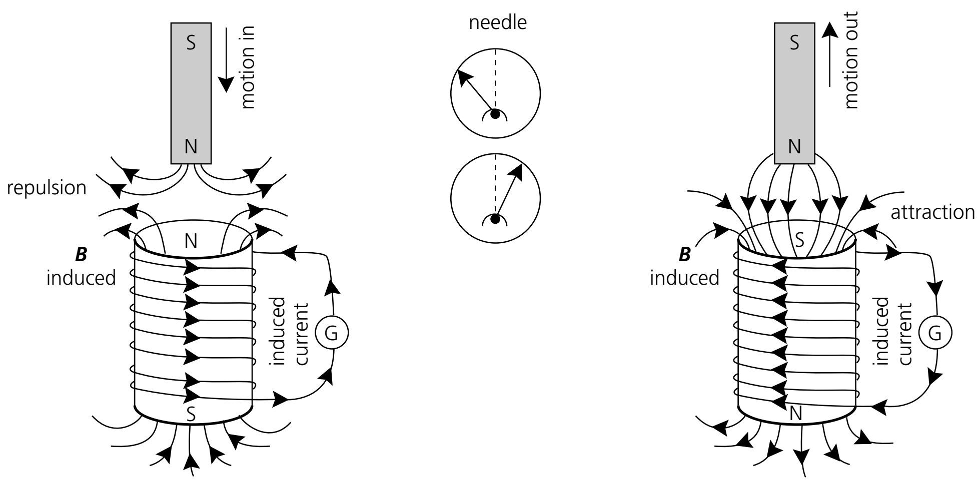


Figure 26.6
Lenz's law.

Lenz's law is only one expression of a fundamental law of nature. In biology you will meet negative feedback — if a light is shone in your eye, your pupil opposes the change and closes up. In chemistry you will meet Le Chatelier's principle in chemical equilibrium. They all work in the same way. When an induced current flows through the solenoid and the galvanometer, the magnetic field produced by the solenoid has a polarity that repels the incoming permanent magnet pole.

NOVEL CHALLENGE

When you swing an aluminium baseball bat through the Earth's magnetic field a small voltage is induced. Estimate the voltage produced, by making some assumptions about the size of the field, its direction, the direction and speed of the swing, and the size of the bat. These values should all be stated in your answer.

Of course, Lenz's law operates just as effectively when the permanent magnet is pulled out of the solenoid. Now the direction of the induced current in the solenoid is such that the magnetic field produced tries to attract the withdrawing permanent magnet back inside the coil again. In both situations a force needs to be exerted and work needs to be done in order to continue to move the magnet. It is this work done that is the origin of the induced electrical energy. We are really transforming mechanical energy into electrical energy and this principle is the basis of all electric generators. The big problem is how to organise for a continuous flow of electrical energy from a generator apparatus and not just single electric current pulses. This problem is examined in Section 26.3.

Magnetic flux

Recall the concept of **magnetic flux**, ϕ, from Chapter 25. The magnetic flux in a region of space is the product of both the magnetic field strength, $\boldsymbol{B}$, and the effective area, $A \cos \theta$, which is perpendicular to the direction of the field, namely:

$$\phi = \boldsymbol{B} A \cos \theta$$

and is measured in webers (Wb).

A mathematical statement of Faraday's law of electromagnetic induction can now be made, involving the notion of induced EMF and current direction as formulated by Lenz, namely:

The EMF induced in a loop is directly proportional to the rate at which the magnetic flux through the loop changes with time

or mathematically as:

$$\text{EMF} = \frac{\Delta\phi}{\Delta t} = \frac{-\Delta(\boldsymbol{B}A)}{\Delta t}$$

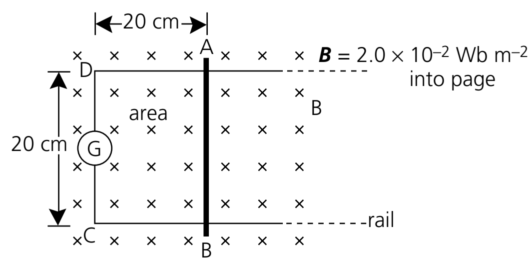
where the negative sign indicates the opposition factor.

If the loop contains N turns, then the EMF is increased by a factor of N. Notice that this equation states that an EMF will be induced if either the magnetic field strength or the area threaded changes with time or, in fact, if both factors change. This law really gives a method for producing an electric generator as all that is needed is a continuously rotating coil.

Example

Consider the apparatus of Figure 26.7. It shows a metal rod AB resting on conducting rails connected to a galvanometer CD. The apparatus is sitting in a magnetic field, as shown.

Figure 26.7



(a) Will there be an induced EMF across AB, if the magnetic field changes from a value of 2.0×10^{-2} Wb m^{-2} to zero in a time of 5.0 ms?

(b) Calculate the value of the EMF and the direction of the induced current around the loop ABCD.

(c) If the rod now is moved to the right at a velocity of 5×10^{-2} m s^{-1}, with the magnetic field remaining constant at its initial value, will there be an induced EMF across AB now?

(d) Calculate the value of the EMF and the direction of the induced current around ABCD for the case in **(c)**.

Solution

(a) Yes, there will be an induced EMF across AB, as the magnetic field strength, ***B***, changes with time.

(b) Use:

$$|\text{EMF}| = \frac{\Delta(\boldsymbol{B}A)}{\Delta t} = \frac{\Delta \boldsymbol{B} \times A}{\Delta t} \quad \text{as area is constant}$$

$$= \frac{2.0 \times 10^{-2} \times 0.2 \times 0.2}{5.0 \times 10^{-3}}$$

$$|\text{EMF}| = 0.16 \text{ V}$$

Lenz's law will require the decreasing magnetic field to be opposed, hence the induced current around the loop will produce its own field to reinforce the original field lines. This requires a clockwise flow of conventional current around ABCD.

(c) Yes, there will be an induced EMF across AB again, because this time the area *A* is changing even though the magnetic field is constant.

(d) Its value is calculated by using:

$$|\text{EMF}| = \frac{\boldsymbol{B}\Delta A}{\Delta t} \quad \text{as } \boldsymbol{B} \text{ is constant}$$

but area is changing because side DA is increasing at the rod velocity. Hence:

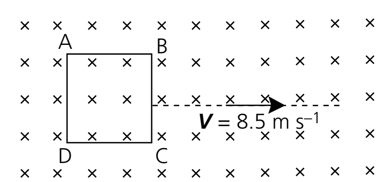
$$|\text{EMF}| = \boldsymbol{B} \times DC \times \boldsymbol{v}_{AB}$$
$$\text{EMF} = 2.0 \times 10^{-2} \times 0.2 \times 5.0 \times 10^{-2}$$
$$\text{EMF} = 2.0 \times 10^{-4} \text{ volts}$$

In this instance, current will flow through AB so as to produce a force that opposes the applied force moving the rod to the right. Using the RH motor rule, the direction of conventional current must be from B to A. Hence, current flow is anticlockwise around the loop BADC.

Questions

1. Define the terms electromagnetic induction, Faraday's law, Lenz's law, induced voltage.
2. A conductor of length 55 cm is moving through a magnetic field of 3.6×10^{-2} T. What is the EMF induced between the ends of the conductor:
   (a) if it is moved perpendicularly to the field at a velocity of 12 m s^{-1};
   (b) if it is moved at an angle of 30° to the field at the same velocity?
3. Figure 26.8 illustrates a conductive metal square coil positioned within a magnetic field of strength 150 mT. If the coil has side AB = 5 cm and is moved sideways at 8.5 m s^{-1}, calculate:
   (a) the voltages induced across each of the sides of the coil AB, BC, CD, DA;
   (b) whether an induced current will flow around this coil as it is moved. Explain.
4. Consider the apparatus of Figure 26.9. Predict the nature of the galvanometer deflection when switch S is closed. What would occur if switch S is opened and closed repeatedly?

Figure 26.8
For question 3.



Ch26-Walding 4th 25/8/04 12:59 PM Page 583

Figure 26.9
For question 4.

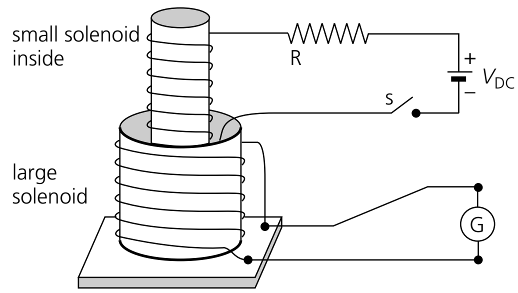
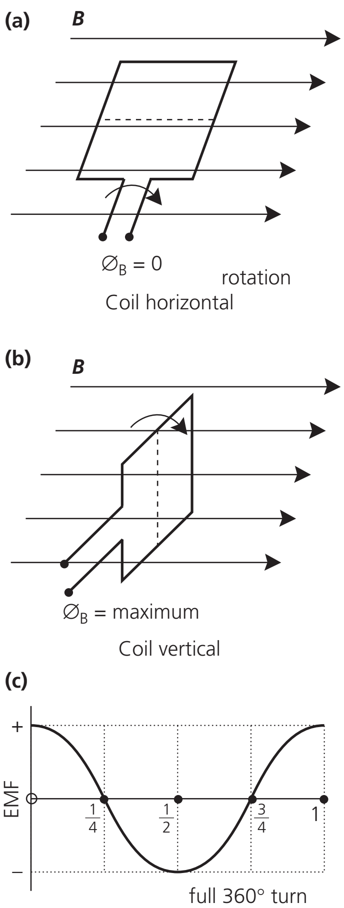


Figure 26.10
Change in flux in a rotating coil:
(a) coil horizontal; **(b)** coil vertical;
(c) the EMF waveform.



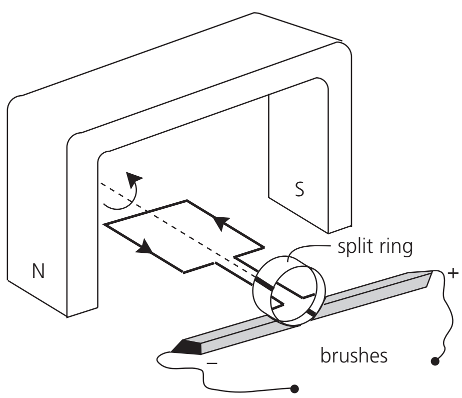
GENERATING ELECTRICITY 26.3

A machine that converts mechanical energy into electrical energy is called a generator, alternator or dynamo. All these machines rely on the induction principle between a coil and a magnetic field (Figure 26.10). A laboratory demonstration generator is shown in Photo 26.2. In Figure 26.10**(a)** the instantaneous flux through the flat coil is zero, but the rate at which the flux is changing, $\Delta\phi$, is maximal, so the induced EMF is maximal. In 26.10**(b)**, however, the instantaneous flux through the flat coil is maximal, but now the rate of change of flux, $\Delta\phi$, is minimal, so the induced EMF is at minimal. If the flat coil is made to rotate through a full 360°, the output induced EMF waveform at the commutator's split ring will be very close to a sinusoidal waveform. (Refer to Figure 26.10**(c)**.)

The magnetic fields of permanent magnets are usually only strong enough to operate small practical generators or motors. The larger devices employ electromagnets. All electrical generators contain two structures:

- the **field** coils, which are the magnetic field-producing coils or magnets
- the **armature**, which supports the coiled conductors that cut the magnetic field lines and carry the induced current externally to commutators or sliprings. The armature is usually made of a laminated soft iron core around which are wound the coils.

Direct current (DC) generators



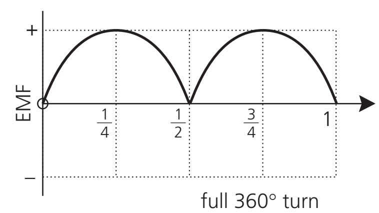


Figure 26.11
A simple DC generator and output.

Photo 26.2
Demonstration generator.

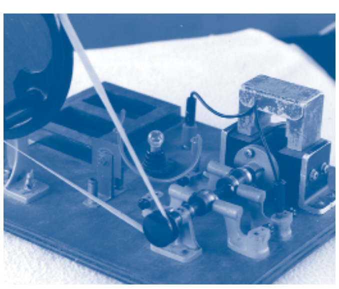


Figure 26.11 illustrates the basic operation of a DC generator. To produce a steady flow of induced current in one direction only from this device, it is necessary to provide a way of reversing the current flow outside the generator every half-cycle of rotation. In older machines this is accomplished by the **split-ring commutator**. The two halves of the metal ring are insulated from each other and, through slipping contacts (brushes), serve as the terminals of the armature coil. In newer machines this reversal is often accomplished using power diode rectifiers, so in fact the DC generator internally is exactly the same as an AC generator. In order to provide even further smoothed DC output, modern DC generators often

use drum armatures that consist of a large number of coil windings in longitudinal slits in the armature core. These coils are then connected to an appropriate multiple segmented commutator that allows contact to each coil in turn as the armature rotates. DC generators are usually operated at low voltages to avoid the electrical discharge sparking that will occur between brushes and commutator at higher voltages.

NEI Activity 26.1 YOUR BICYCLE

Use the library or reference books to find a diagram of a bicycle dynamo. You might like to compare this with an actual device from a bike, and try to answer these questions:

1. How is the dynamo turned?
2. How does the device produce electric current?
3. How could the output current be increased?
4. Does the voltage vary with speed?

Alternating current (AC) generator

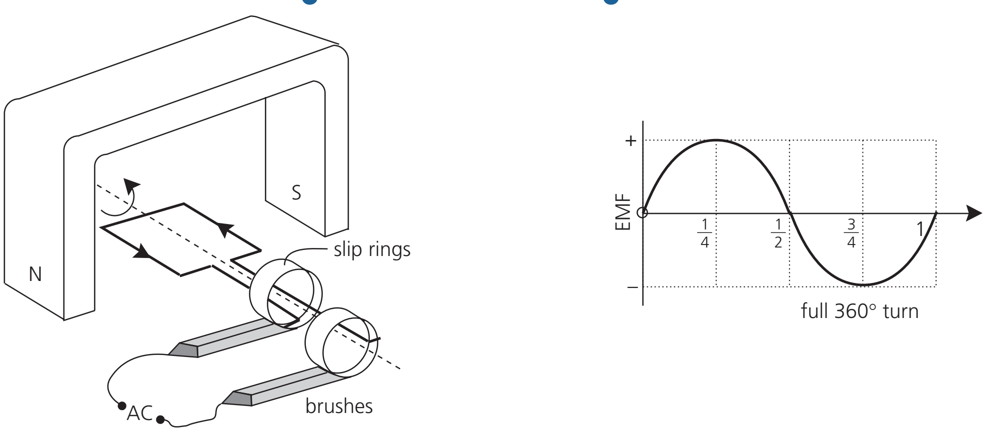


Figure 26.12
A simple AC generator.

Figure 26.12 illustrates the difference between the DC and AC generator design. This AC design has no commutator, with the armature coil being terminated simply at two sliprings. In practical electrical engineering, the transmission of generated electrical energy is in the form of alternating current, and so most large commercial generators are of the AC type, with the permanent magnet being replaced by electromagnetic field coils, energised by external DC sources of EMF. Low speed AC generators are often built with up to 100 poles to improve their efficiency and produce the required output AC frequency, but high speed AC generators most often are simple two-pole machines. It is common to refer to any AC generator as an **alternator**.

In Figure 26.12, if the coil is made to rotate continuously, then a continuous sinewave AC voltage will be produced. Mathematically, the output AC voltage produced is given by the sinusoidal equation:

$$E = E_0 \sin \omega t$$

where E_0 is the peak output voltage, given by:

$$E_0 = N A \boldsymbol{B} \omega$$

where N = number of turns in the coil; A = cross-sectional area of the coil in square metres; $\boldsymbol{B}$ = magnetic field strength in tesla (T); ω = angular velocity = $2\pi f$, where f = frequency in hertz (Hz).

NOVEL CHALLENGE

Car stereo amplifiers generally need a higher voltage than the 12 V DC provided by the car battery. How is this achieved if transformers can only convert AC. The term 'oscillating switch' or 'switched-mode power supply' may give you a clue.

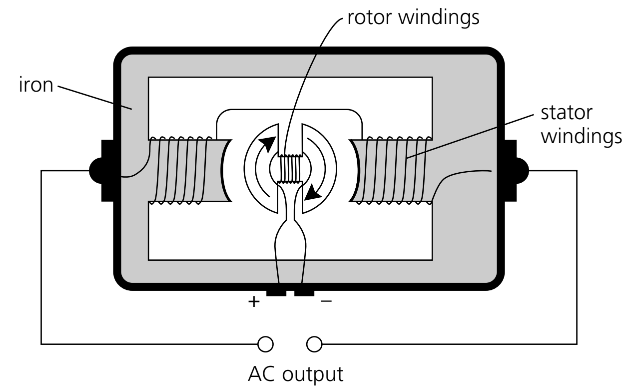


Figure 26.13
A simplified AC alternator.

High voltage alternators often use a design in which a stationary armature set of windings remains fixed in place, with field magnetic coils positioned on a rotor that revolves inside the armature windings. This design reduces sparking and helps to prevent mechanical failures (Figure 26.13). The current generated by all simple alternators follows a single sine waveform called **single phase** alternating current. If the armature windings are composed of a triple set of windings aligned at 120° to each other, the alternator will produce **three phase** alternating current. This three phase system is the most commonly used for electrical power generation and distribution. Three phase alternating current distribution will be discussed in Section 26.5. Typical specifications for the large industrial AC generator that is used in power stations throughout Australia are as follows:

- Generator weight — 350 to 400 tonnes
- Generator rotation — 3000 rpm at rotor
- Rotor assembly — electromagnet on steel, 15 m × 1 m
- Electromagnet field — 2500 A producing 1.68 T
- Generator output — 500 MW at 24 kV (3 phase at 50 Hz)

Usually the rotor assembly is connected to a superheated, steam-driven, four cylinder turbine, which itself may weigh 500–600 tonnes. Steam pressure at the turbine is 16.8 MPa at 540°.

— Mini generator applications

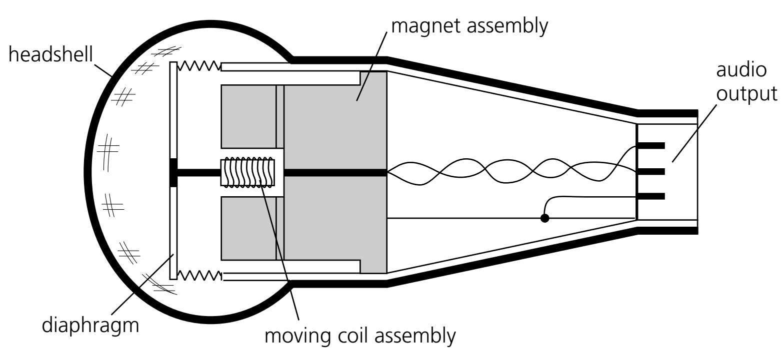


Figure 26.14
A moving coil microphone.

Numerous applications for the principle of induction exist, with most of them being simple AC generators. For example, a loudspeaker may be used in reverse to produce a **moving coil microphone**. However, these microphones are more usually constructed as in Figure 26.14. Sound vibrations cause the diaphragm plate to oscillate and move the coil within the poles of the permanent magnet. This generates a small AC voltage, which is fed to an amplifier.

A similar structure is the basis of a magnetic stylus and cartridge for a record player. Of course, today, CD players are more common in audio systems but high quality turntables are still used by audiophiles. The principle of operation of the **magnetic cartridge** is that an alternating voltage is induced as the pickup stylus is forced to vibrate when it passes along the record grooves. This small AC voltage is again fed to an amplifier.

Some automatic marine navigation buoys generate electricity using the induced current generated by relative motion of coils and magnets. As these devices bounce up and down in the waves the inertia differences between a solenoid coil and a spring-loaded magnet allow generation of an induced current, which is rectified and fed to batteries. These power navigation lights are attached to the buoys. You might like to sketch how they would be made.

26.4 MUTUAL INDUCTION AND TRANSFORMERS

Electromagnetic induction will also occur in the situation where the expanding or collapsing magnetic field of an electromagnet solenoid coil cuts through a stationary conductor or second coil. This type of induction between closely separated coils is called **mutual induction**. It is important in the construction of **ignition coils** and **transformers**.

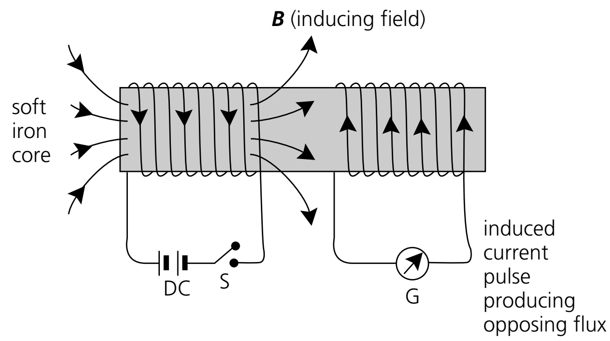


Figure 26.15
Mutual induction principle.

NOVEL CHALLENGE

A Year 9 student asks you: 'Does a transformer convert voltage to current?' What would you say, in language the student could understand?

Refer to Figure 26.15, showing two solenoid coils wound onto a common soft iron core. When the switch is closed in the left-hand circuit a magnetic flux that is expanding is produced and cuts the right-hand circuit. This induces a voltage pulse across the ends of the right-hand coil in such a way as to oppose this expanding flux, according to Lenz's law. No voltage is induced when the current in the left-hand coil is constant. However, if the switch is opened again, a collapsing magnetic flux now cuts the right-hand coil and again an induced voltage, opposite in direction to the original pulse, is produced. In a car electric ignition coil, the continuous switching action of the distributor points in the coil's primary circuit induces very high spark voltages in the coil's secondary circuit. This high voltage is passed in correct sequence via the distributor again to the engine's spark plugs to fire the fuel–air mixture in the cylinders.

Transformers

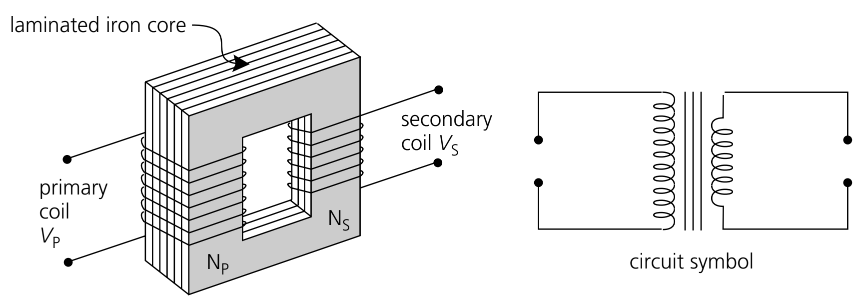


Figure 26.16
A simple transformer and its symbol.

Another obvious method of producing continuously expanding and collapsing magnetic fields is to use a primary circuit driven by AC current. This produces a transformer device that uses mutual induction to vary AC voltages. Two coils called the **primary** and **secondary** are wound onto a common soft iron core (Figure 26.16). The soft iron core concentrates the magnetic flux lines threading both coils. If the primary coil is fed with AC voltages at a particular frequency, an induced AC voltage of equal frequency will occur across the secondary coil. Notice that there is no physical electrical connection between the two sets of coils or windings. If the ratio of turns in the windings is varied, either a **step-up** or a **step-down** transformer is produced. For any transformer operating under ideal conditions the following relationship is determined by experiment:

$$\frac{V_P}{V_S} = \frac{N_P}{N_S}$$

where V_P = AC voltage across the primary; V_S = AC voltage across the secondary; N_P = number of primary turns; N_S = number of secondary turns.

- If $N_P > N_S$, then the transformer is a step-down, which reduces AC voltage.
- If $N_P < N_S$ then the transformer is a step-up, which increases AC voltage.

This discussion makes transformers sound like marvellous devices for varying AC voltages and they are; however, energy is not created in these devices as the electrical power available at the output is never greater than the electrical power supplied to the input of the transformer; namely:

$$\text{Input power} = \text{output power}$$
$$V_P \times I_P = V_S \times I_S$$

In practice, although transformers are very efficient devices, there is always some energy loss. When an AC current passes through the primary coil, tiny circulating currents called **eddy currents** are set up in the soft iron core. This causes heating and represents lost energy. The coil conductors also lose heat through ohmic heating due to coil resistances. Practical transformers are constructed on a laminated soft iron core, where the core is made from insulated flat iron sheets of correct shape. This technique helps to reduce eddy current losses. Typical electromagnetic transformers are about 90–95% efficient in operation.

Example

A transformer purchased from an electronics store is labelled as 240 V AC input, 56 V CT @ 120 VA output. Calculate:

(a) the voltages available at the output;
(b) the maximum output current able to be drawn;
(c) the current drawn from the mains supply at maximum output.

Solution

(a) The term 56 V CT means that the secondary winding is 'centre tapped' and thus the 56 volts is divided into a positive and a negative 28 volts with respect to the zero volts centre tap. The full voltage available across the secondary would be 56 V.

(b) Use power rating $P = VI = 120$ V A.
Thus:

$$V_S \times I_S = 120$$

or

$$I_S = \frac{120}{V_S} = \frac{120}{56} = 2.1 \text{ A}$$

(c) Use:

$$V_P \times I_P = V_S \times I_S = 120 \text{ V A}$$

or current drawn from the mains supply. $I_P = \frac{120}{V_P} = \frac{120}{240} = 0.5$ A

Transformers are extremely useful engineering devices as they allow for the changing or transforming of AC voltages, not only in small construction circuits but on an electrical energy production and distribution level. Mains distribution high voltage transformers are a common sight on the poles of our suburban electricity networks. The earliest patent covering the construction of a transformer appears to be that applied for in Britain by the team of Carl Zipernowski, Max Deri and Otto Titus Blathy, all of Budapest, Hungary, on 27 April 1885, numbered Patent No. 5201 under the title, 'Improvements in induction apparatus for transforming electric currents'. Max Deri, in the same year, applied for and received the first patent for an electrical distribution transformer.

NEI Activity 26.2 TRANSFORMERS

1 Transformers are one of the easiest components to recognise in old radios and broken electrical appliances. See if you can get hold of one of these in order to examine it closely. Make diagrams of what you see. Can you identify the windings and the laminated core?

2 The grey boxes located on power poles in your street are the transformers. Make a diagram of these boxes showing where all the wires go. Can you hear a humming noise? What causes that?

Questions

5 Describe the differences in construction and design between a DC generator and an alternator.

6 Why would it be dangerous to connect a step-down transformer in reverse, that is, with the primary voltage connected across the secondary windings?

7 Explain the statement 'A motor run in reverse can act as a generator'.

8 A square loop of wire of side 8 cm is rotated through 90° in a magnetic field of 2.5×10^{-2} T in 0.1 s. Calculate the average EMF induced.

9 If neon lights require at least 12 kV for their operation and operate from a 240 V line, what is the turns ratio required of the transformer used? Would it be a step-up or a step-down transformer?

10 The armature of a 50 Hz AC generator rotates in a 0.15 T magnetic field. If the area of the coil is 2.0×10^{-2} m^2 and the coil contains 150 turns, calculate the peak output voltage, E_0.

26.5 POWER TRANSMISSION AND DISTRIBUTION

Because electricity generating stations are usually located at the sites of the primary fuels for the turbine generators, most electricity needs to be distributed over very long distances. This requires very large distribution networks, or 'grids' as they are called. For example, very large power stations are located near coal fields in both Queensland and Victoria, and in New South Wales and Tasmania they are located near large hydroelectric facilities. Often several separate power stations are combined to supply power to the distribution grids. Figure 26.17 illustrates the typical Australian distribution network from the generating power station to the industrial or domestic consumer.

Modern power station AC generators deliver power at between 11 kV and 33 kV. In order to distribute the electrical power, transformers at the station step up the voltage to typically 200 kV, 330 kV or 500 kV. These very high voltages are used because over long cable distances, the loss of energy through ohmic heating is reduced. Recall that electrical power loss is proportional to the square of the current ($P = I^2R$), so electrical engineers use very

NOVEL CHALLENGE

What does a power station sell you: power, voltage, current or energy (joules)? Who owns the electrons in the wires from the power station to your power point?

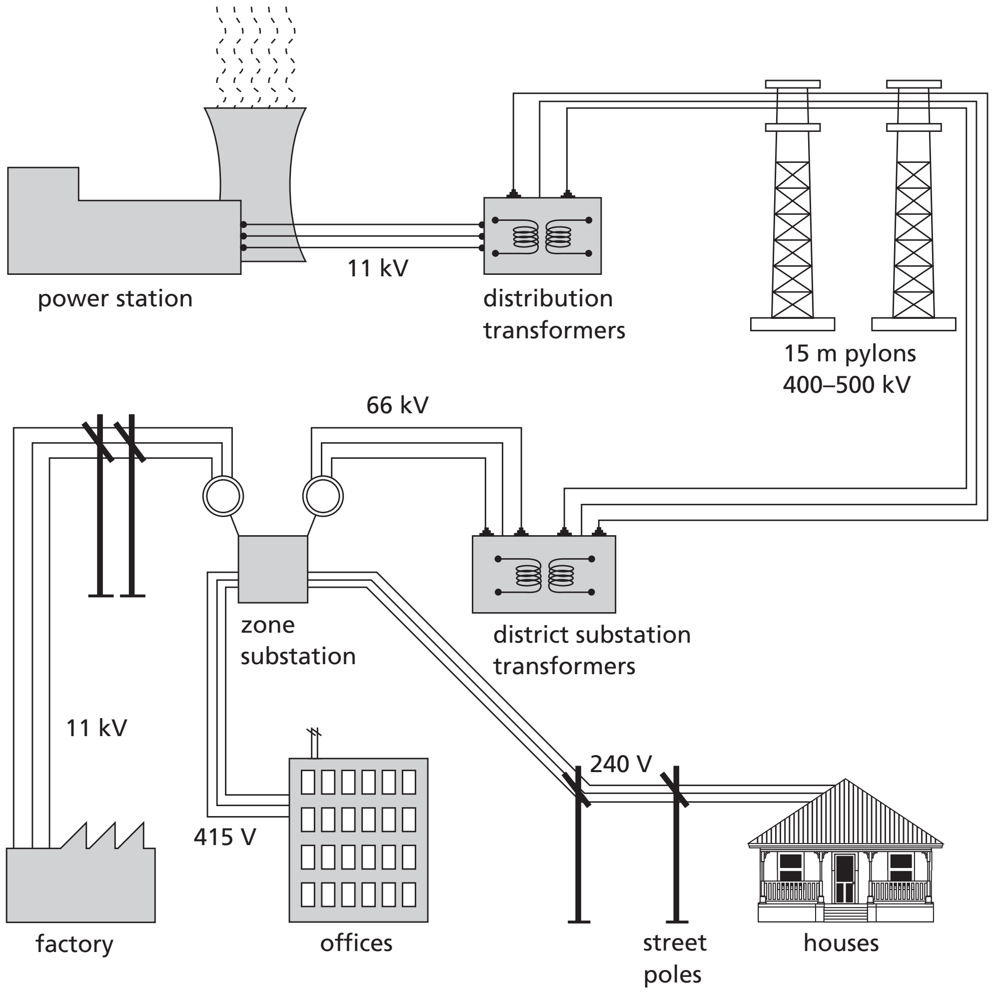


Figure 26.17
A typical Australian distribution network.

POWER STATIONS

Power stations typically generate AC at 11 kV to 33 kV, which is then transformer increased to 220–500 kV. Power switches called circuit breakers are placed throughout the distribution network to help to avoid damage due to lightning strikes. Street poles also often have a fused link to help to avoid damage caused by voltage surges and spikes.

INVESTIGATING

As mentioned above, when electricity is distributed it is stepped up to high voltages (e.g. 275 kV) to reduce 'ohmic losses' and then later stepped down for household or industry use.

Why don't they just generate it as 275 kV and cut the first step out? Engineers at your nearest power station may be able to help. Why wouldn't you contact Energex?

high voltages in order to keep the currents small and the power losses as heat to a minimum. Refer to the data contained in Table 26.1 illustrating the power losses in a distribution system of 150 km, assuming a total cable resistance of 6 Ω and a total power distributed of 500 MW. The table shows the losses incurred as a percentage for different voltages.

Table 26.1 POWER LOSSES IN A DISTRIBUTION SYSTEM

System voltage	500 kV	220 kV	66 kV
Current ($P = VI$)	1000 A	2270 A	7580 A
Power loss (I^2R)	6 MW	30.1 MW	345 MW
Power loss (%)	1.2	6.0	70

At even higher voltages than 500 kV the losses would be less, but the increased chance of high voltage discharge from the power cables to ground is regarded as unacceptable. Atmospheric conditions such as moisture, rainfall and high winds also make power distribution at even higher voltages impractical. In Australia the grid distribution voltages are stepped down again at substations to voltages of 22 kV and 11 kV for distribution to local areas or zone substations. Within the local neighbourhood, local electricity companies distribute power via street poles at typically 415 V_{RMS} three phase or 240 V_{RMS} single phase to factories, schools and houses. Local telegraph pole transformers do this job of stepping down from 11 kV and occasionally you may hear crackling in the vicinity of these pole transformers, especially in wetter climate conditions. This effect is generally nothing to worry about. The electricity supplied to our houses is AC at 240 V_{RMS} 50 Hz through a pole fuse often protecting a group of houses. Refer back to Chapter 22 for a discussion of household electricity and circuits.

If you investigate the nature of distribution electricity poles and towers in your local neighbourhood, you will certainly notice quite a large number of porcelain or glass insulator spacers and large encased oil filled transformers. Remember that even dry air breaks down under high voltage and will conduct. This again could lead to distribution power losses if the cables were not kept apart far enough. In the 500 kV transmission lines conductors need to be at least 600 mm away from pylons and each other. At smaller voltages, the separations can be reduced.

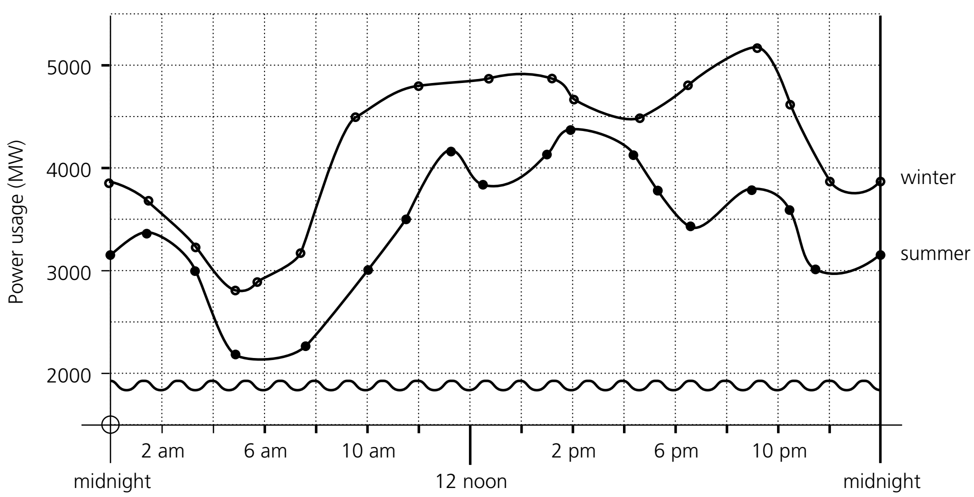


Figure 26.18
Power usage in Australia.

It is interesting to consider the changing demands of the electricity consumer over time. The graphs in Figure 26.18 illustrate typical electrical power usage curves on a summer and winter weekday for comparison. The electricity company is responsible for providing consumers with a supply of electrical energy that is as constant as possible. Our efforts at electrical energy conservation should help to lighten the load, but occasionally supply cannot be maintained and we experience an electrical **brownout** where the supply voltage suddenly reduces. This causes lights to dim, motors to slow down and is generally damaging to equipment. The opposite is a **power surge** where the average voltage increases. This is also damaging to equipment, leading to fuses blowing and transformers and motors overheating. Luckily, both of these occurrences are infrequent in Australia today. Power **blackouts** occur when physical damage, such as lightning strikes or trees breaking or shorting power lines, trips out circuit-breakers or fuses. In times when electricity workers go on strike or during power generating station mechanical failure, the electricity company will often shed the load by disconnecting sections of the distribution grid temporarily. At these times it is important to turn off all major appliances, so that when the power comes back on, a large surge will not cause even further problems. Electricity generating companies and engineers use the larger capacity AC generators at power stations to supply the **base load**. This usually represents about 70% of required demand. These large generators, usually coal fired, take up to 16 hours to reach maximum capacity. The rest of the load is covered by much more flexible oil-fired, gas-fired and hydroelectric generators. These smaller generators can come on-line very quickly, often in a matter of minutes. Overall, the generation of electricity for our community involves a fine balancing act between supply and demand, involving different generation methods, costs of development and maintenance, and accurate forecasting of future trends in consumer requirements. Being an electrical power engineer is a worthwhile profession.

Ch26-Walding 4th 25/8/04 12:59 PM Page 591

Questions

11 Why is a DC electrical distribution not favoured by engineers in Australia? There are some places in the world where DC distribution does take place. See if you can find out where they are and present a report.

12 Using Figure 26.18, explain the following features:
(a) The dip and sudden rise in power consumption at about 6 am.
(b) The higher average demand on a winter's day.
(c) The dip and sudden rise of the graphs at about 6 pm.

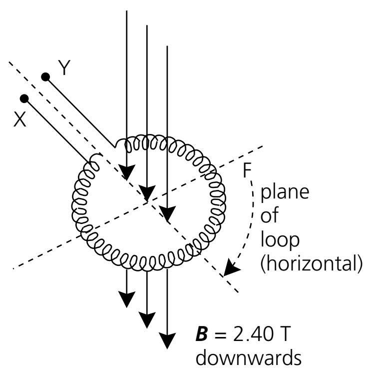
13 Suppose the total resistance of a power line distribution system is rated at 120 Ω. Calculate the power loss in the system if 600 kW is generated and transmitted at either 11 kV or 66 kV.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

Figure 26.19
For question 14.



***14** Figure 26.19 shows a circular spring loop perpendicular to a magnetic field. If the spring is released and its area changes from 0.55 m^2 to 0.15 m^2 in 0.4 s, what is the EMF induced between the points X and Y shown on the loop?

***15** Find the average EMF induced in a coil having 15 turns when the magnetic flux threading it increases from 0.03 Wb to 0.15 Wb in a time of 0.02 s.

***16** If a dynamo is found to generate an EMF of 200 V when rotating at 60 rpm, what will be the induced EMF at a rotation of 80 rpm?

***17** Why is Lenz's law often stated as a law of commonsense when applied to electromagnetic induction?

***18** Why do the solenoids used in electromagnets and transformers need to have soft iron cores? Why are the cores laminated?

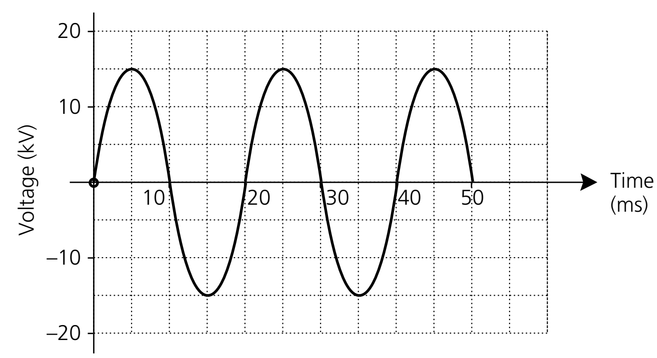
***19** A transformer is designed to step down the mains voltage used in Australia to a voltage of 6.3 V AC. If the primary coil has 2000 turns, how many turns will the secondary coil have?

***20** Why does the operation of a transformer depend on AC rather than DC voltage?

***21** Explain the following terms as applied to the electrical distribution system in the community: **(a)** thermal power station; **(b)** turbine-driven generator; **(c)** brownout; **(d)** power surge; **(e)** generated base load.

****22** Figure 26.20 represents graphically the output of an alternator that is rotating at 50 Hz in a magnetic field of 0.40 T. By superimposing your own curves on top of this one, illustrate the changes to the output waveform if **(a)** the magnetic field changes to 0.80 T; **(b)** the field remains the same but the rate of rotation increases to 100 Hz.

Figure 26.20
For question 22.



****23** Figure 26.21 illustrates two solenoid coils connected in series. Describe the behaviour of the suspended magnet in the right-hand coil after a bar magnet is dropped through the left-hand coil as shown. Explain your reasoning.

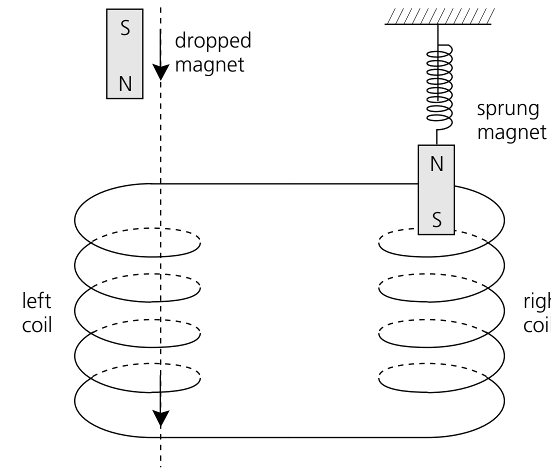


Figure 26.21
For question 23.

****24** What is the frequency and voltage of the AC supply to your home? Sketch a graph of the voltage as a function of time. Carefully note all peak and RMS values.

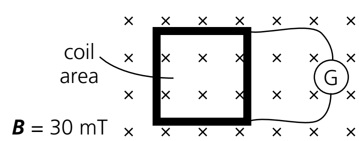
****25** Dynamic or moving coil microphones are generally used for live vocals in rock bands. Draw a diagram that illustrates their basic construction and describe the principle of electromagnetic induction on which they operate. Some household intercom units use a small loudspeaker for both speaking and listening functions. Explain how this is possible.

***26** What are the factors affecting the output EMF of any practical transformer? If these transformers are not perfect, where does the energy lost between input and output go to?

Extension — complex, challenging and novel

*****27** A galvanometer of internal resistance 5 Ω is wired in series with a 200-turn coil of area 50 cm^2, as shown in Figure 26.22. If this assembly is perpendicular to the field and its intensity varies from 30 mT to 10 mT in a time of 0.02 s, calculate the current reading on the galvanometer, assuming a total coil resistance of 20 Ω.

Figure 26.22
For question 27.



*****28** The diagram of Figure 26.23 shows a flexible loop of wire between the poles of an electromagnet that provides a uniform field $\boldsymbol{B}$ in the region of the loop. At time $t = 0$, the current through the electromagnet is turned off and the field $\boldsymbol{B}$ falls to zero at time t_1, as shown in the accompanying graph.

(a) Draw a corresponding graph of the nature of the induced EMF across the ends of the loop as a function of time.

(b) If the loop has an area of 0.04 m^2, with $\boldsymbol{B}_0 = 0.6$ T at time $t = 0$ s, find the average EMF induced if time $t_1 = 2.0$ s.

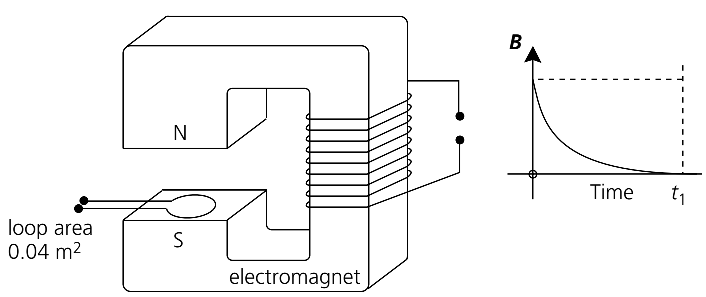
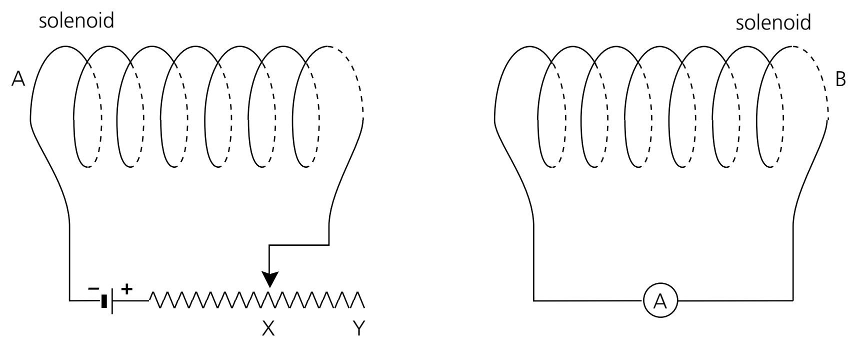


Figure 26.23
For question 28.

*****29** A variable resistor is connected into circuit with solenoid A and a battery as shown in Figure 26.24. If the resistor is varied from position X to Y, what is the direction of induced current in solenoid B? Explain your analysis.

Figure 26.24
For question 29.



*****30** A hospital generator, 600 m from the hospital, generates 40 kW of power at 250 V AC for use in an emergency. The power lines for distribution to the hospital complex have a total resistance of 0.2 Ω.

(a) How much power is lost in the system?
(b) Can the hospital staff use normal 240 V appliances?
(c) What would happen if the generator was sited 4 km away from the hospital and the total line resistance was 1.0 Ω?
(d) How much power would be lost if transformers were used at both ends of the lines changing the voltages up to 10.0 kV and down again?

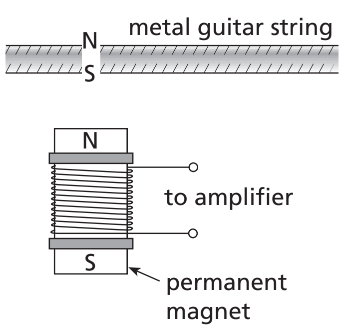
****31** A simple generator has a 100-loop square coil of 8.0 cm side length. How fast must it turn in a 0.50 T field to produce a peak voltage of 20 V AC?

****32** Discuss why the government doesn't want to legislate to force all consumers to use electrically efficient, electronic, compact, fluorescent-style lights in domestic homes.

****33** Photo 26.3 shows a copy of a Fender Stratocaster, the type of guitar used by Jimi Hendrix and many other musicians. Whereas an acoustic guitar depends for its sound on the acoustic resonance produced in the hollow body of the instrument by the oscillations of the strings, an electric guitar is a solid instrument, so there is no body resonance. Instead, the oscillations of the six metal strings are sensed by electric 'pickups', which send signals to an amplifier and a set of speakers.

Photo 26.3
A Yamaha Pacifica — one of the most popular copies of the Fender Stratocaster guitars.

Figure 26.25



The basic construction of a pickup is shown in Figure 26.25. Wire connecting the instrument to an amplifier is coiled around a small magnet. The magnetic field of the magnet produces a north and south pole in the section of the metal string just above the magnet. When the string is plucked and made to oscillate, its motion relative to the coil changes the flux of its magnetic field through the coil, inducing a current in the coil. As the string oscillates towards and away from the coil, the induced current also oscillates.

On a Stratocaster, there are three sets of pickups each with different frequency responses. The musician can choose which set to use. Hendrix would also sometimes rewrap the wire in the pickup coils to make them more sensitive.

(a) How would the frequency of the string compare with the frequency at which the flux changed in the coil? Explain.

(b) In Figure 26.25, as the string moved away from the pickup, would the current to the amplifier move out through the top or bottom wire in the pickup coil?

(c) What change could Hendrix have made to the pickup coil (as mentioned above)? Describe what difference this change would make and why.

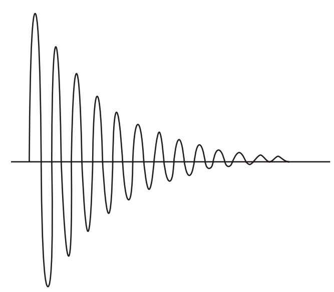
(d) Could you use this sort of pickup on a guitar with nylon strings? Explain.

(e) Another way musicians can change the sound of electric guitars is by using an 'effects pedal', which is plugged into an amplifier and activated by the tap of a foot. Three major types are described below:

- **Big muff** — produces a guitar note of relatively constant amplitude, drawing out the sound until it eventually decays. It does so by clipping off the high and low peaks.
- **Delay** — the effects box stores information about the notes and feeds them to the amplifier a few milliseconds later. Often used in rockabilly music.
- **Wah-wah** — uses a filtering device that changes the volume of different frequencies as the sound decays. For example, as the sound decays the high frequency notes are reduced quickly whereas the lower frequency notes are allowed to predominate to produce the muffled 'wah-wah' sound.

If the guitar produces a note as shown in Figure 26.26, match the three output waveforms (Figure 26.27) with the three types of effects.

Figure 26.26



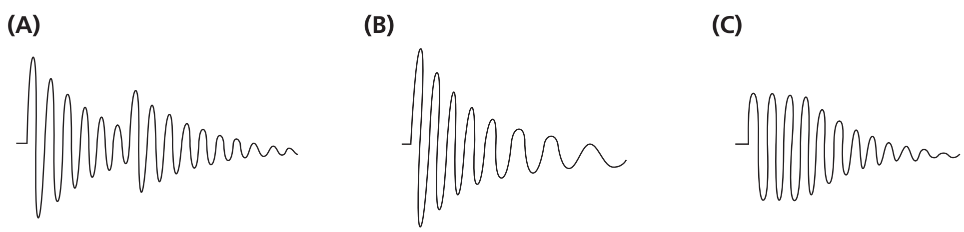


Figure 26.27

Ch27-Walding 4th 25/8/04 1:33 PM Page 595

UNIT 09

Atomic & Nuclear Physics

Ch27-Walding 4th 25/8/04 1:33 PM Page 596

CHAPTER 27

Atomic Structure

27.1 INTRODUCTION

The existence of the atom is widely accepted but its incredibly small size is hard to comprehend. Only recently have scientists been able to see and photograph individual atoms. Every breath you take contains about 10^{24} atoms. The full stop at the end of this sentence is a million atoms wide.

People get very confused about atoms. They ask questions like these:

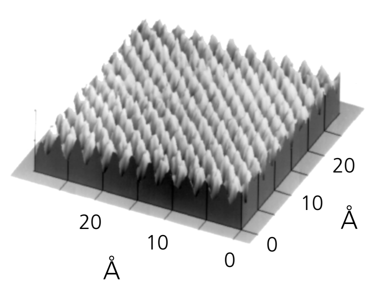
- If atoms are mostly empty space, how come a brick feels so hard?
- What colour is an atom?
- If the electron is negative why doesn't it get sucked into the positive nucleus?
- How many atoms are there in the universe? It must be a mind-bogglingly big number!
- If the nucleus is made up of positive particles, why don't they fly apart?
- How do we know atoms really exist if you can't see them?

Scientists have answered the last question but the rest need careful explanation. That's what this chapter is about.

27.2 FOUNDATIONS OF ATOMIC THEORY

The word *atom* comes from the Greek words *a* meaning 'not' and *tom* meaning 'to cut'; hence, not cuttable, or indivisible. This arose from the ideas of the Greek philosophers Democritus and Leucippus 2500 years ago.

Photo 27.1
Photo of surface of graphite in air taken by Dr K. Finlayson (Murdoch University, WA) using a scanning force microscope (SFM). Individual carbon atoms can clearly be seen at this magnification of 1.5 million times. The unit 'angstrom' (Å) is a non-SI unit equal to 10^{-10} m.



Democritus

The word 'philosopher' was used differently from the way it is now. Until the word 'scientist' was coined in 1830, natural philosophers were people who loved learning about the world (it comes from the Greek *philos* = 'love', *sophia* = 'wisdom'). One of the first philosophers to suggest the idea of atoms was Leucippus; however, not much is known about his work. The earliest writing about atoms was that of **Democritus of Abdera** (460–371 BC). He argued that you could not keep cutting up something into smaller and smaller pieces forever; eventually you would end up with a piece that could not be cut any further — the 'atom' (Figure 27.1). Democritus also argued that atoms were in constant motion and that all atoms were composed of the same substance but differed in size and shape. His model accounted for many observable properties of matter but as he believed that the atom was the fundamental indivisible particle, he did not try to explain its structure.

Greek philosophers did not test their theories by experiments as scientists would today. This wasn't because they did not have the equipment to do so; Greeks had no inclination to conduct experiments because the philosophers came from élite (rich and powerful) families

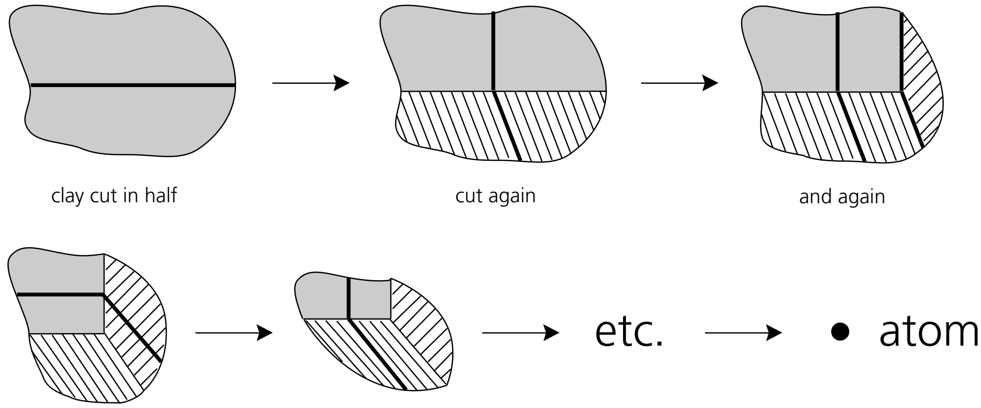


Figure 27.1
Democritus's concept of the atom.

PHYSICS FACT

The word 'scientist' was coined by the distinguished Oxford professor William Whewell in 1834. Up until then they were called 'natural philosophers' (= lovers of learning about nature).

and thought manual work (such as experimenting) was only for slaves. They developed their ideas by reasoning and discussion. Their method of reasoning was to state some important principle or law, often based on observations of the heavens, then draw conclusions based on it. Experimentation generally did not occur — that was a seventeenth-century development.

Aristotle

One of the most famous natural philosophers was **Aristotle**, born in 384 BC. His father was doctor to the king of Macedonia and Aristotle received a good education in Athens under the teaching of another famous Greek philosopher, Plato. His writings were vast and many of his theories went unnoticed until the thirteenth century, when Christian theologians began to endorse his work as being truth. Aristotle argued that matter could be divided an infinite number of times until there was a void, that is, nothing. He taught that matter was made up of four elements — earth, air, fire and water — and that different combinations produced different substances (Figure 27.2). This was at odds with the atomic theory but religious leaders could understand Aristotle's view of matter. They did not like the idea of atoms that moved around seemingly without the control of the gods. So from 300 AD to 1600 AD, the atomic theories lay dormant while Aristotle's ideas flourished.

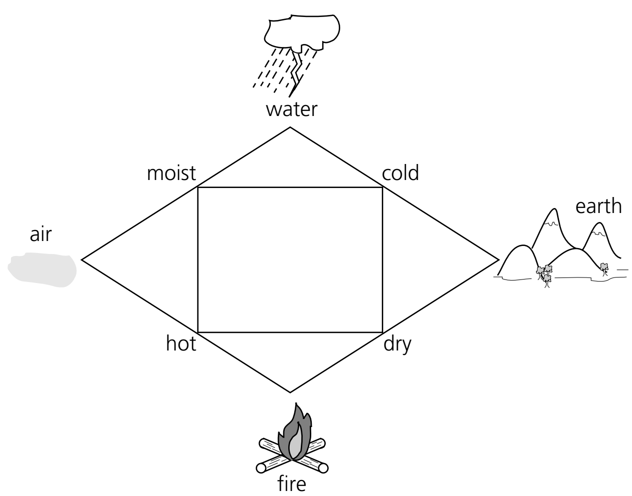


Figure 27.2
Aristotle's concept of the four elements.

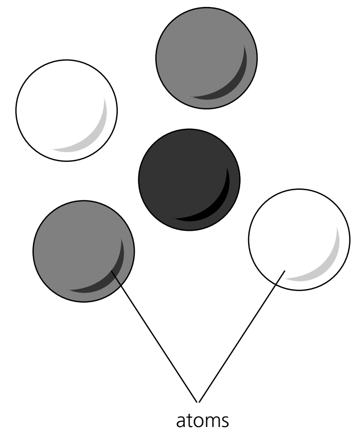
THE WORK OF JOHN DALTON 27.3

A long time passed before the idea of atoms was revived. In the seventeenth century, an Englishman by the name of Francis Bacon (1561–1626) introduced the idea of experimentation

Ch27-Walding 4th 25/8/04 1:33 PM Page 598

as a way of understanding nature. He reasoned that direct observation of nature rather than a study of Aristotle or theology (religious writings) gave a better idea of how the world worked. He is often thought of as the father of modern science. Later, people like Galileo and Descartes supported his idea of the experimental method. This inspired a lot of experimental work through England and Europe. Robert Boyle (1627–91) investigated the gas laws; Joseph Priestley (1733–1804) experimented with the extraction of gases; Antoine Lavoisier (1743–94) discovered the composition of air; and Henry Cavendish (1731–1810) discovered hydrogen. These experiments paved the way for a breakthrough in our understanding of matter. Near the beginning of the nineteenth century, the English scientist **John Dalton** (1766–1844) conducted a series of experiments, and published his atomic theory proposing the existence of individual particles called atoms in all matter (Figure 27.3), with a list of atomic masses. Dalton believed that all atoms of the same element were identical and that compounds were formed by the combination of atoms in small whole-number ratios. Other scientists went on to add to his theories and then in 1897, **J. J. (Joseph) Thomson** discovered the electron. This discovery led to the modern-day theory of atomic structure.

Figure 27.3
Dalton's model.



27.4 DISCOVERY OF THE ELECTRON

The big leap in our understanding of atomic structure came with the use of electricity in the laboratory. In the mid-nineteenth century, the effects of sparks, arcs and electrical discharges through gases were most interesting but of little importance. But after Heinrich Gessler invented the vacuum pump in 1855, electrical discharges through gases at low pressures produced brilliant results. Suddenly, the possibility of using vacuum tubes for electrical lighting (and making a fortune) was investigated and knowledge about the discharge increased dramatically. **Sir William Crookes** (1832–1919) in 1876 designed a number of tubes to study these charges. A variety of discharge tubes based on Crookes' designs are commonly available in physics classrooms today (Figure 27.4).

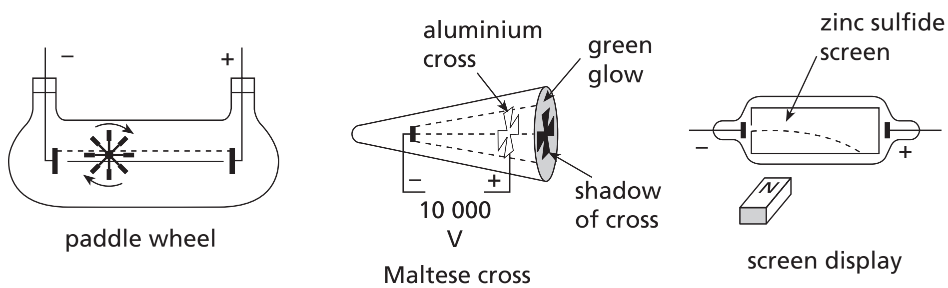
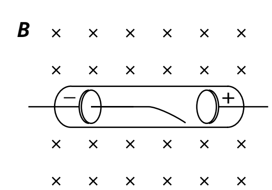


Figure 27.4
Discharge tubes commonly used in school laboratories.

These tubes contained various gases at low pressure and when a high voltage (about 20 000 V) was applied across the terminal at the ends, a purple light was seen; but as the pressure was reduced, the purple faded and the glass glowed with a green light near the positive end. There was dispute about what caused this glow but the invisible rays involved became known as 'cathode rays' as they emanated from the negative (cathode) terminal. The green light was an example of fluorescence (Latin *fluere* = 'to flow' and *esse* = 'to exist') — light given off by a substance (the glass) when being illuminated by energy from an external source (the discharge).

Crookes suggested that cathode rays would be deflected by magnetic fields (Figure 27.5) and by a series of experiments, Thomson was able to show this magnetic field deflection and so proved Crookes' hypothesis to be correct.

Figure 27.5
A magnetic field (into the page) deflects cathode rays downward.



Thomson devised a technique for passing cathode rays through an electric and a magnetic field that were orientated so as to exert opposing forces on the negatively charged rays. By this method, Thompson was able to measure the charge-to-mass ratio of the cathode ray particles, which he named electrons. The rays were deflected by the fields and struck the end of the glass tube, emitting light. The strength of the fields was then adjusted until the beam was not deflected. At this point the magnitude of the force exerted by the magnetic field, $\boldsymbol{F}_B$, was equal to the magnitude of the force exerted by the electric field, $\boldsymbol{F}_E$:

$$\boldsymbol{F}_E = \boldsymbol{F}_B$$
$$qE = \boldsymbol{B}q\boldsymbol{v}$$
$$\therefore \boldsymbol{v} = \frac{qE}{\boldsymbol{B}q} = \frac{E}{\boldsymbol{B}}$$

where $\boldsymbol{v}$ = the velocity of the electrons.

By knowing the strengths of the magnetic and electric fields, Thomson was able to calculate the velocity of the electron. When he switched off the electric field, he knew that the deflection of the electron was due just to the magnetic field. The curved path of the electron was due to centripetal force provided by the magnetic field.

The centripetal force formula:

$$\boldsymbol{F}_C = \frac{m\boldsymbol{v}^2}{r}$$

The magnetic force formula:

$$\boldsymbol{F}_B = \boldsymbol{B}q\boldsymbol{v}$$

(Recall these formulas from Chapters 5 and 25.)

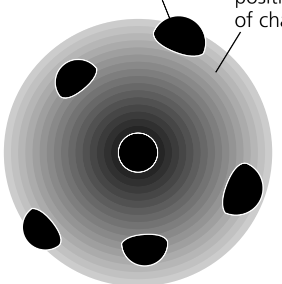
Hence:

$$\boldsymbol{B}q\boldsymbol{v} = \frac{m\boldsymbol{v}^2}{r}$$

or

$$\frac{q}{m} = \frac{\boldsymbol{v}}{\boldsymbol{B}r}$$

Figure 27.6
Thomson's 'plum pudding' model.



Substituting the value for velocity $\boldsymbol{v}$ above:

$$\frac{q}{m} = \frac{\frac{E}{\boldsymbol{B}}}{\boldsymbol{B}r} = \frac{E}{\boldsymbol{B}^2r}$$

Thus the charge-to-mass ratio of the electrons would be given by:

$$\frac{q}{m} = \frac{E}{\boldsymbol{B}^2r}$$

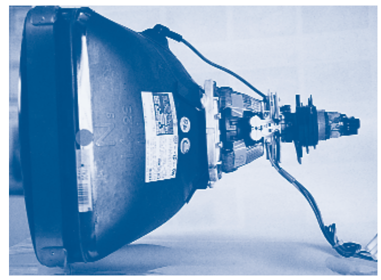
The ratio q/m is more commonly referred to as the e/m ratio, where e stands for the charge on the electron. Thomson was able to measure the e/m ratio to be 1.759×10^{11} C kg^{-1}. However, neither e nor m can be determined individually by this method.

When Thomson determined the e/m ratio for the hydrogen ion to be about 1800 times greater than that of the electron he realised that the electron was much smaller than the smallest known atom (hydrogen) and electrons were identical no matter where they came from. The only conclusion was that the electron was a sub-atomic particle; a conclusion later proved correct. Seven of Thomson's assistants went on to win Nobel prizes.

27.5 THOMSON'S MODEL OF THE ATOM

Thomson knew that an atom was electrically neutral but contained negatively charged electrons so he proposed a 'plum pudding' model of the atom (Figure 27.6). This envisaged the atom as a ball of positive charge (like a pudding) with electrons scattered throughout (like raisins). This model explained many features of the atom but couldn't explain others such as atomic spectra or radioactivity and was eventually replaced.

Photo 27.2
A cathode ray tube. The electron gun at the rear of a computer monitor can be seen.



Modern cathode ray tubes

Cathode ray tubes (CRTs) have undergone continuous development since Crookes's original models. Many modern applications such as televisions, visual display units (VDUs) and cathode ray oscilloscopes (CROs) contain CRTs. In a modern CRT, a heated filament is used in an evacuated tube to produce an electron beam. The part of the tube that accelerates the high-speed electrons is called the **electron gun**.

Activity 27.1 DISCHARGE TUBES

SR

1. If you have an old TV set or computer monitor that is no longer any good, you may be able to take the cover off and look at the tube. Do not plug it in! And if it has been turned on in the past few months don't touch anything — the capacitors could still be charged (zap!!) Try to locate where the electrons are produced (the electron gun) and where the electric field and magnetic field coils could be. Share your findings with your class.
2. If your teacher demonstrates a flat screen discharge tube, suggest how you could test the effect of an electric field on it. Be careful not to stand too close to any operating discharge tube for too long. There is a slight danger from X-rays.

Millikan's experiment

Thomson had worked out the ratio between charge and mass for the electron but it was not until the experiments of American physicist **Robert Millikan** (1868–1953) between 1909 and 1916 that the actual charge (in coulombs) and consequently the actual mass of the electron became known. Millikan's experiment was one of the classic experiments in physics.

A mist of oil was sprayed into the region above a pair of metal plates (Figure 27.7) and eventually a single oil drop fell through the hole in the top plate. When the plates were uncharged, the drops fell through at a steady velocity dependent on their weight. Being light objects, the oil drops reached a terminal velocity because they experienced considerable air resistance. By viewing the droplets through a microscope, Millikan was able to measure the terminal velocity and hence could determine the droplet's weight. When an electric field was applied between the plates, however, the motion of the drop was changed. It could be made to rise and fall depending on the voltage applied to the plates. A drop that remained stationary did so because it became charged during the spraying process and the electric force ($\boldsymbol{F}_E = \boldsymbol{B}qv$) was balanced by the gravitational force ($\boldsymbol{F}_W = m\boldsymbol{g}$). See Figure 27.8. If the upward electric force was greater than the gravitational force, then the charged droplet would move upwards. For a stationary drop:

Figure 27.7
Schematic diagram of Millikan's apparatus.

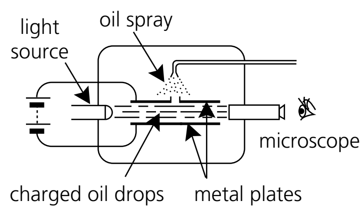
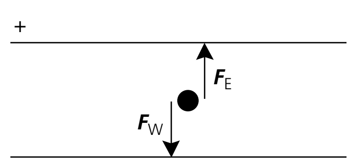
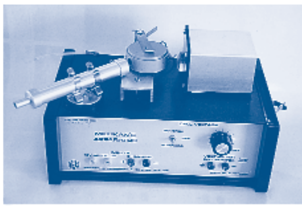


Figure 27.8
An oil drop balanced in the electric field between oppositely charged plates.



$$\boldsymbol{F}_E = \boldsymbol{F}_W$$
$$Eq = m\boldsymbol{g}$$
i.e.
$$q = \frac{m\boldsymbol{g}}{E}$$

Photo 27.3
Millikan's apparatus — the telescope is on the left, the viewing chamber in the middle and the light source on the right.



Millikan studied the behaviour of thousands of oil drops and was able to work out the charges on them. He found that all charges were whole-number multiples of the minimum charge of $1.602\,1892 \times 10^{-19}$ coulombs. It is worth remembering that an uncharged oil drop still has billions of electrons but these are balanced by an equal number of protons. In Millikan's experiment, it is the *excess* electrons that are being counted.

Knowing the charge on the electron enables the mass of the electron to be calculated. If $e/m = 1.76 \times 10^{11}$ C kg^{-1} and $e = 1.6 \times 10^{-19}$ C, then $m_e = 9.11 \times 10^{-31}$ kg.

Example

In a Millikan oil drop experiment, an oil drop of mass 4.0×10^{-16} kg was held stationary between a pair of electric plates held 2.0 cm apart. The voltage across the plates was 120 V. Assume $\boldsymbol{g} = 9.8$ m s^{-2}.

(a) Calculate the magnitude of the electric field between the plates when the oil drop was stationary.
(b) What was the size of the charge on the oil drop?
(c) How many elementary charges does this correspond to?

Solution

(a)
$$E = \frac{\boldsymbol{V}}{d} = \frac{120}{0.02} = 6 \times 10^{3} \text{ V m}^{-1}$$

(b)
$$q = \frac{m\boldsymbol{g}}{E} = \frac{4 \times 10^{-16} \times 9.8}{6 \times 10^{3}} = 6.5 \times 10^{-19} \text{ C}$$

(c)
$$\text{Number of elementary charges} = \frac{6.5 \times 10^{-19} \text{ C}}{1.6 \times 10^{-19} \text{ C}} = 4.$$

In 1910, Millikan was appointed Professor of Physics at the University of Chicago and in 1923 he was awarded a Nobel prize for this work.

At about the time that Thompson and Millikan were experimenting on electrons, the existence of radioactivity became known. The next section describes some of the milestones and people involved in its discovery.

SR Activity 27.2 MATCHBOX MODELS

1. Take four matchboxes and put two marbles in one, three in the next, four in the next and five in the last. If you don't have marbles, any small objects such as 20 cent coins will do.
2. Ask a friend to measure the mass of each box and challenge her to tell you how many marbles or coins are in each box. Can she also tell you the mass of an empty matchbox?
3. How does this relate to Millikan's experiment?

SR Activity 27.3 COMPUTER SIMULATIONS

If your school has Millikan's apparatus, your teacher may demonstrate how it works and collect some data. If not, there are some excellent computer simulations available.
If you don't have access to a commercial package there are several available on the Internet. If you can download one of these, you should find that you get a good feel for Millikan's experiment.

Questions

1 Briefly compare and contrast the methods used by the Greeks and modern scientists in their investigation of scientific problems.

2 In a Millikan oil drop experiment, an oil drop of mass 1.05×10^{-15} kg was held stationary between a pair of electric plates 2.6 cm apart and with a potential difference of 210 V. Assume $g = 9.8$ m s^{-2}.

(a) Calculate the magnitude of the electric field between the plates when the oil drop was stationary.

(b) What was the size of the charge on the oil drop?

(c) How many excess electrons were on the drop?

27.6 THE DISCOVERY OF RADIOACTIVITY

Meanwhile, back to the debate about the nature of cathode rays — two opposing camps had developed. In England, Crookes led the particle theory group by saying the rays were torrents of negatively ionised gas molecules. Across the English Channel, the German scientist Heinrich Hertz led the opposition group, which said that as the rays could pass through metal foil they couldn't be big gas molecules and must be more like electromagnetic (light) waves. The breakthrough was soon to come by accident.

The history of science is littered with examples of scientists stumbling on amazing discoveries purely by chance. Does this mean that these serendipitous (lucky) discoveries are not real science? No! The discovery of radioactivity involved many of these fortuitous events and the talent of four main players — Wilhelm Roentgen, Henri Becquerel, and Marie and Pierre Curie.

Wilhelm Roentgen

The breakthrough into the cathode ray wave versus particle debate came in November 1895. **Wilhelm Roentgen** (1845–1923), an obscure professor of physics at Würzburg, wanted to resolve this controversy so he bought a Crookes tube and soon found that something was happening outside the tube. Although his discharge tube was completely enclosed in black cardboard, he noticed that a piece of paper coated with the fluorescent compound barium platinocyanide, which happened to be lying on the bench, glowed while the tube was in operation. He identified the origin of the radiation in the glass wall of the tube where it was struck by the cathode rays. He didn't know what to call them so he just said they were **X-rays**. They were not charged particles since they were undeflected by magnetic fields. On 23 January 1896, in his one-and-only significant public lecture in Würzburg, Roentgen stated 'for the sake of brevity, I should like to use the term "rays" and to distinguish them from others I shall use the name "X-rays"'. The anatomist von Kolliker, who was present at this meeting, had his hand X-rayed (Photo 27.4) — the audience was astonished. Roentgen's breakthrough in the debate about whether cathode rays were particles or waves came as the consequence of this apparently accidental discovery. Within a few days it was news all over the world; it was the subject of music hall jokes and within a few days almost every physicist was repeating the experiments to admiring audiences.

Photo 27.4
Radiograph of the hand of anatomist Albert von Kolliker, made at the conclusion of Roentgen's lecture and demonstration at the Würzburg Physical-Medical Society on 23 January 1896.

EI

Activity 27.4 WHAT ROENTGEN SAW

See if you are able to try the following experiment (under teacher supervision). Use one of the Crookes gas discharge tubes from the physics laboratory. Place some luminescent mineral (e.g. uranium salt such as uranyl nitrate) about 1 m away from the tube in a darkened room and note the effect. Now place the uranium salt under an ultraviolet light (black light). How does it compare? Can you see how Roentgen was misled?

Henri Becquerel

The early workers advanced many ingenious theories for the origin of the X-radiation. At first they thought it might be connected with the fluorescence of the glass walls of the tubes. **Henri Becquerel** (1852–1908), a professor of physics in Paris, began investigating a variety of fluorescent materials to see if they also emitted these X-rays. All trials with various minerals, metal sulfides, and other compounds known to fluoresce on exposure to visible light gave negative results. Then he remembered he had a sample of potassium uranyl sulfate from 15 years earlier, which he exposed to a bright light and placed on top of a photographic plate wrapped in two sheets of black paper. After several hours the plate had darkened. Cool!

Becquerel soon found out that this amazing behaviour had nothing to do with the fluorescence of the uranium salt. On 16 February 1896 he was going to confirm his result by repeating the experiment (as scientists do) but it was a rainy day so he left the uranium salt on top of the wrapped plate in a drawer, hopefully not near his lunch. A few days later he developed the plate and to his amazement a sharp, clear image was present. The penetrating radiation had come from the uranium itself, not from fluorescence under bright light. He had discovered a new source of radiation but it was not identical to X-rays. The radiation was emitted spontaneously by the uranium and was nothing to do with outside influences such as sunlight. The rays became known as 'Becquerel rays', but he did not do any more research on them, preferring instead to carry out research on X-rays. He discovered the phenomenon of naturally occurring radiation, partly by good scientific deduction and partly by accident. Becquerel rays were later to become known as 'alpha rays'.

Activity 27.5 WATCH THE LUMINESCENCE

If you can get hold of an *old* watch or clock dial with a luminescent dial, look at it under a microscope or with a 10 power (or more) magnifying glass. Luminescent paint contains radioactive radium in the pigment. In a dark room, once your eyes are used to the dark, you should be able to see the watch dial clearly. Under the lens you should be able to see tiny light flashes that are given off whenever a particle is emitted by the unstable isotopes and strikes another chemical (zinc sulfide, ZnS) in the paint. During the manufacture of many of these dials in the United States in the 1920s, the women who worked on them used to make a fine point on their paintbrushes by wetting them between their lips. Those who did eventually ended up with tongue cancer, and fifteen of them died. Luminous dials these days contain H-3, Kr-85, Pm-147, Tl-204.

Marie and Pierre Curie

Marya (Marie) Sklodovska was born in Warsaw (Poland) in 1867 and after high school did a degree in physics, topping the class. While enrolled in a second degree (in maths) in 1894 she met **Pierre Curie** (1859–1906) and they married the following year. On the advice of Becquerel, Pierre suggested Marie undertake a PhD studying the new 'Becquerel rays' as they were known. She began testing uranium compounds and measured their activity by using a device called an electrometer invented earlier by Pierre and his brother Jacques. As radiation passes through the air it can remove electrons from the gas molecules, creating positive ions, which will be attracted to a negatively charged electrometer and alter its charge.

She soon discovered that some uranium minerals were four or five times as active as they should be on the basis of their uranium content. She concluded that new substances must be present. She introduced the term 'radioactivity' (Latin *radius* = 'ray', a spoke of a wheel) to distinguish this form of emission from X-rays. Marie and Pierre began separating the uranium ore pitchblende into various fractions. The Curies bought a truckload of pitchblende and boiled it up in large vats, distilling the vapour. In July 1898 she discovered a new element, which she named 'polonium' (after Poland), and in December she isolated radium.

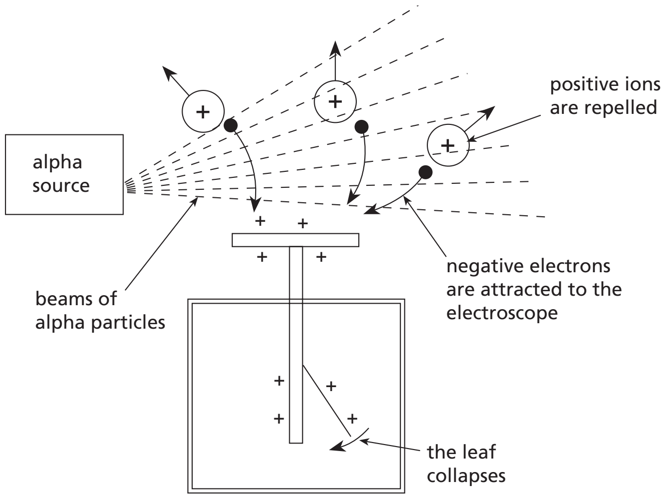


Figure 27.9
A radioactive source discharges an electrometer by ionising the air around it.

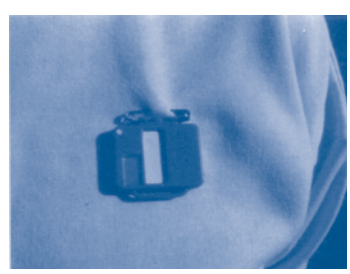
Unfortunately, Pierre was run over by a horse and dray and killed in Paris in 1905. Marie's daughter, Irene, joined the laboratory (The Radium Institute) and with her husband, Frederick Joliot, in 1934 was the first to demonstrate **artificial radioactivity** by bombarding aluminium with alpha rays. Later that year Marie died of leukaemia, a cancer of the blood, in her case caused by handling those radioactive substances for so long. When she died her skin was white from the lack of red blood cells. Today, in the garden at the rear of the Curies' laboratory, white roses grow — a moving metaphor and reminder of Marie's life.

Since then, many other radioactive elements have been discovered, although not all of them occur naturally. We now use the term 'nuclear radiation' instead of Becquerel rays because it is known that the radiation comes from the nucleus.

EI Activity 27.6 RADIATION DETECTORS

1. Charge an electroscope and time how long it takes to discharge into the air. What causes this? Repeat but hold a lighted taper close to the plate. Explain the result. Repeat again but, under teacher supervision, place an alpha particle source such as strontium-90 close to the plate. Explain this result.
2. Besides the electroscope, there are several other types of detectors available today, the simplest being photographic film that is often worn as a badge to monitor personal levels of radiation. Others include the Geiger-Müller tube, the thermo-luminescent dosimeter, the cloud chamber and the scintillation counter. Write a paragraph on one of the above types of detectors and include a drawing of what it looks like.
3. Why is a badge radiation monitor (see Photo 27.5) not the best way to protect you from radiation hazards?
4. Under teacher supervision, place a gamma source (such as cobalt-60) on top of a pinch of sodium chloride to irradiate it. After a while place the salt on top of a heated hotplate in a dark room and watch what happens. Explain how this is related to the thermo-luminescent dosimeter.

Photo 27.5
Radiation badge.



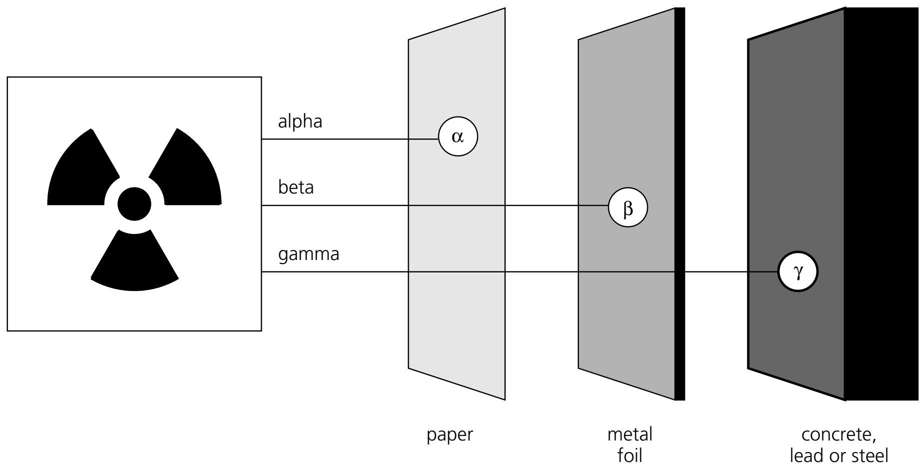
Questions

3. Radioactivity was discovered by accident. Does this mean it is not 'real' science?
4. Radioactivity would never have been discovered, had it not been for the discovery of X-rays. Comment.
5. The measurement of nuclear radiation by the Curies depended on the invention of a particular instrument. Which instrument was this?

NUCLEAR RADIATION 27.7

At the end of the nineteenth century research into radioactivity became one of the most important frontiers of research. One man led the field for many years. He was an experimental physicist, who won a Nobel Prize in chemistry and was perhaps the most productive physicist the world had known — the New Zealand born **Ernest Rutherford** (1871–1937). Rutherford won a scholarship to Cambridge to study under J. J. Thomson, beginning his life's work. He found that radiation could be classified according to its **penetrating ability** and in 1899 wrote: 'There are present at least two distinct types of radiation — one that is very easily absorbed, which will be termed for convenience α (alpha) radiation, and the other of a more penetrating radiation, which will be termed β (beta) radiation' (after the first letters of the Greek alphabet). In 1900, the French physicist **Paul Villard** (1860–1934) discovered a third more penetrating radiation, which he called γ (gamma).

Figure 27.10
The penetration power of radiation.

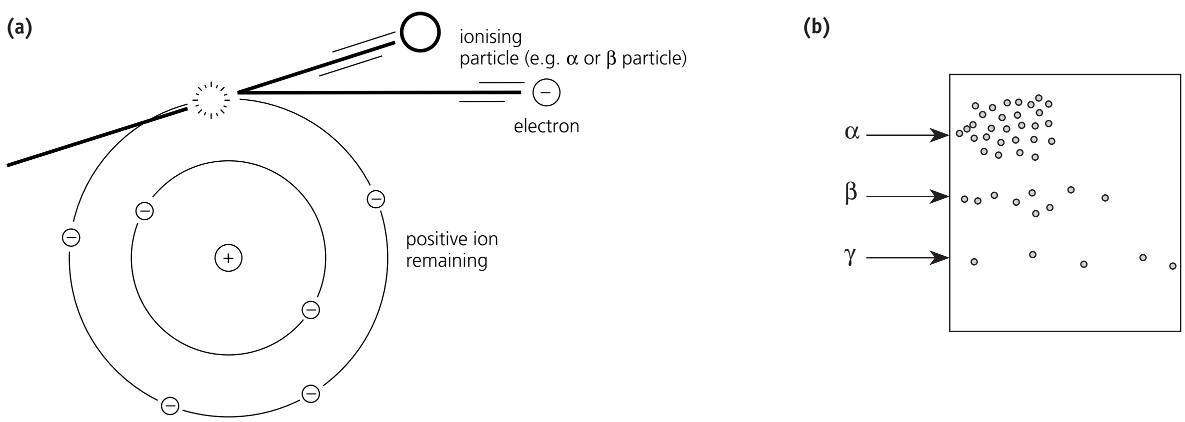


In 1903, Rutherford prepared some data on the relative penetrating power of the radiation. He used the term 'half-thickness' to indicate the thickness of aluminium that would reduce the intensity of the radiation to half.

- Alpha radiation is almost totally absorbed by a piece of paper; it will only travel a few centimetres in air. Aluminium half-thickness = 0.0005 cm.
- Beta radiation is almost totally absorbed by a millimetre of lead or a few millimetres of Perspex; it will travel about 1 m in air. Aluminium half-thickness = 0.05 cm.
- Gamma radiation is never fully absorbed and requires several centimetres of lead or over 1 m of concrete. Aluminium half-thickness = 8.0 cm.

Ionising ability

Figure 27.11
(a) The process of ionisation.
(b) Comparative ionising ability.



Sometimes so much energy is absorbed by an atom that an electron completely escapes from the atom and a positive ion is produced. When this radiation causes ions to form, it is called **ionising radiation** (Greek *ion* = 'to go', hence the idea of ions moving around).

- Alpha radiation is powerfully ionising.
- Beta radiation has lower ionising ability than alpha radiation.
- Gamma radiation is only very weakly ionising.

The next step for physicists was to uncover the exact nature of each type of radiation. They did this by several ingenious experiments.

Magnetic field effects

At the same time as radioactivity was discovered, the effect of magnetic fields on cathode rays was being investigated. It was a logical extension to try similar experiments on alpha, beta and gamma rays.

It was found that three types of deflections occurred when the radiation from a sample of radioactive material was passed through a magnetic field. In Figure 27.12, the rays are shown moving vertically up the page as they pass through a magnetic field into the page (shown by the crosses). Some rays deflect to the left, some to the right and some pass through unaffected. If you apply your hand or palm rules for the effect of a magnetic field on a charged moving particle, you will find that the rays passing to the left must be positively charged (alpha). The small curvature of the alpha rays indicates that they are relatively heavy particles. Similarly, rays to the right must be negatively charged (beta, or electrons), and rays passing straight through must be uncharged (gamma). In the early experiments, the evidence for deflection was collected on a photographic plate. In 1900, Rutherford and his assistant Thomas Royds showed that alpha particles are helium ions.

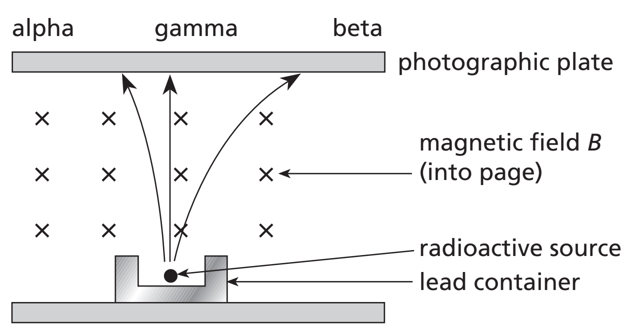


Figure 27.12
The deflection of alpha, beta and gamma rays by a magnetic field.

Questions

6 Alpha particles have a positive charge. How can they form neutral helium atoms in the tube? Where must the electrons come from?

7 How did investigations of cathode rays lead to the discovery of radioactivity?

8 Explain why alpha radiation has a low penetrating ability but is a powerfully ionising radiation, whereas gamma radiation is highly penetrating but a very weakly ionising radiation.

The nuclear atom

The greatest discovery of all was yet to come. Rutherford and his colleague Hans Geiger had experimented on the scattering of alpha particles by thin pieces of heavy metal and found small deflections of about 1°. One day in 1910 at the University of Manchester, Geiger suggested to Rutherford that their young colleague Edward Marsden should begin research.

Ch27-Walding 4th 25/8/04 1:33 PM Page 607

Rutherford jokingly said to let him see if any alpha particles scatter through a large angle, knowing that they probably would not. Rutherford told the story:

> Three days later Geiger came to me in great excitement saying, 'We have been able to get some of the alpha particles coming backwards'. It was the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. (Source: E. Rutherford, *Philosophical Magazine*, Vol. 21, 1911.)

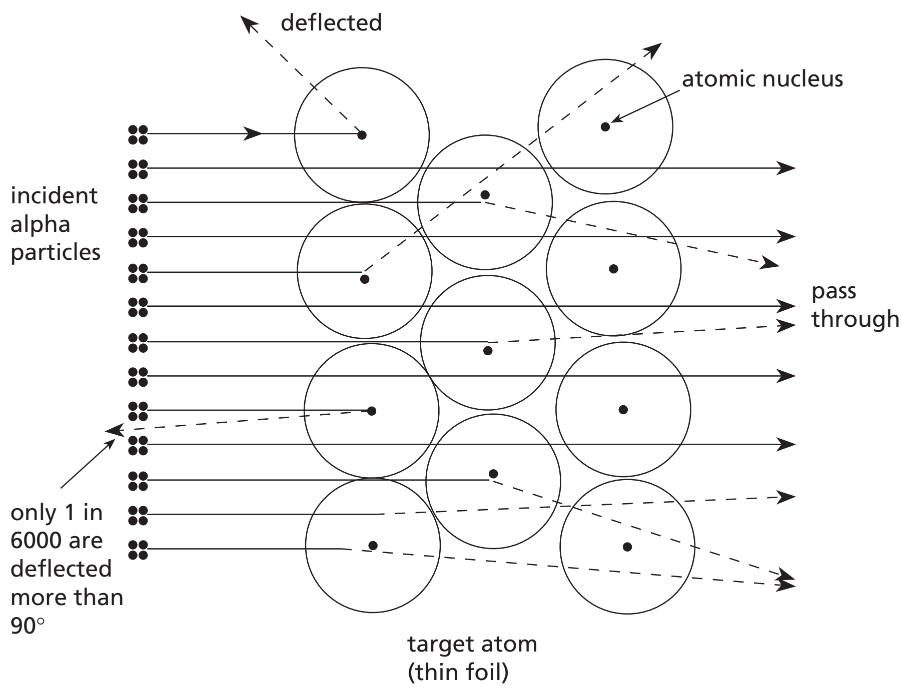


Figure 27.13
Alpha particles in a thin foil scattered by the atomic nucleus.

Geiger and Marsden used a beam of alpha particles from a radon source fired at gold foil a few atoms thick (about 10^{-7} m). Most of the alpha particles went straight through as though the metal was empty space and some scattered through small angles, but a few bounced straight back, deflected through 180°. Rutherford stated his nuclear model in 1911. The positive charge is concentrated in the massive centre of the atom with electrons revolving in orbits around it like planets around the sun (see Figure 27.14). Rutherford called this massive centre the **nucleus** (from an Indo-European language *knu* = 'nut, kernel'). He calculated that it had a diameter of about 10^{-14} m, about 1 ten-thousandth the diameter of the atom (10^{-10} m).

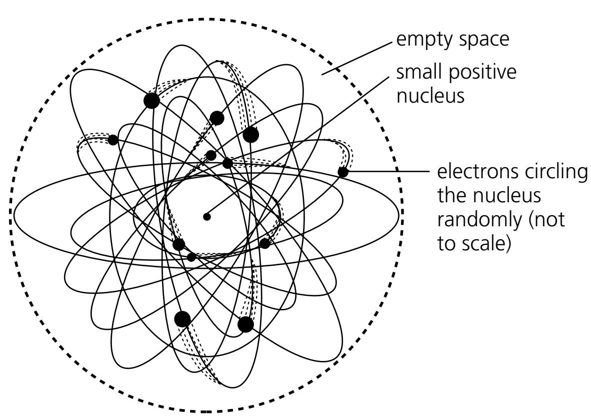


Figure 27.14
Rutherford's model.

Discovery of the proton

Once physicists had discovered alpha and beta particles, they developed a model for the structure of the nucleus. However, the nucleus of a hydrogen atom is too light to contain even a single alpha particle. In 1919, Rutherford bombarded some nitrogen atoms with α particles and succeeded in ejecting some positive particles. These particles are now known to be **protons** (Greek *protos* = 'first' or 'fundamental'), the nucleus of a hydrogen atom. These protons were soon found to have a positive charge, equal in magnitude but opposite in sign to the charge on an electron.

The work by Rutherford and other scientists provided evidence that protons were constituents of all atomic nuclei and he was able to propose a model for the nucleus as consisting of protons and much lighter electrons. Rutherford's laboratory is still radioactive to this day.

PHYSICS FACT

Famous artist Pablo Picasso said his paintings were inspired by Rutherford's theory that you could break up an atom into individual pieces and reconstruct it in different ways.

Have you seen any of Picasso's portraits in which body parts have been swapped around?

Discovery of the neutron

In the following year, Rutherford delivered a lecture with far-reaching consequences. He proposed that it was possible that a proton in the nucleus might combine with an electron to form a neutral particle, which he called a **neutron** (Greek *ne* = 'not', *uter* = 'one or the other'; hence not positive or negative). This could be the particle to explain why the hydrogen nucleus has a charge of 1+ and a mass of 1 unit whereas helium has a charge of 2+ but a mass of 4 units. Over the next 12 years researchers looked in vain but their means of detecting particles was by deflection in magnetic and electric fields. As the neutron was predicted to be uncharged, it was unlikely to be detected.

In 1930, two German physicists, W. G. Bothe and Hans Becker, bombarded boron and beryllium with α particles and found a strange radiation — one that could penetrate lead better than gamma rays (and that's highly penetrating!). In France, Irene Curie (daughter of Marie and Pierre Curie) and her husband Frederick Joliot directed α particles onto a block of paraffin wax, a material rich in H atoms (see Figure 27.15). (You would recognise paraffin wax as surfboard wax — perhaps it was an overcast day and Irene didn't go surfing.) As the radiation struck the wax, 'knock-on' protons were formed on the other side. This seemed as though a new type of radiation was coming from the beryllium.

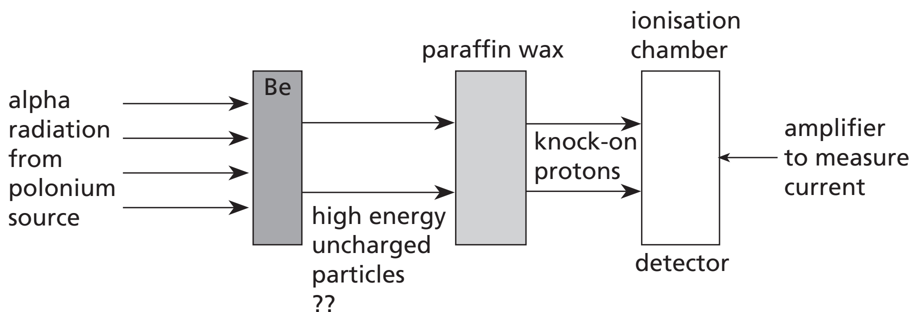


Figure 27.15
The apparatus used by Chadwick to identify the neutron.

In 1932 **James Chadwick**, while working at the Cavendish laboratory in Cambridge under Rutherford, used an ionisation chamber to measure the energies of the knock-on protons. Chadwick showed that the energy of the radiation was much higher than previously measured for γ rays. The radiation could not be γ rays but was instead due to uncharged particles of mass similar to that of the proton. These uncharged particles were neutrons. Chadwick was awarded the Nobel prize in 1935 for his discovery. During the war he headed Britain's team of nuclear physicists who worked on the atomic bomb. He was knighted in 1945.

Questions

9 Comment on the truth of this statement: 'In 1919, Rutherford proposed a model of the atom comprising a small positive nucleus of protons and neutrons with electrons circling it.'

DIFFICULTIES WITH THE RUTHERFORD ATOM 27.8

The model of the atom as proposed by Rutherford provided a great step forward, but it also raised some difficulties. It is known that when a charged particle changes velocity (either speed or direction) it will radiate energy. If the electrons are in orbit around the positively charged nucleus then they should radiate energy because they are continually changing direction. If this were true, the atom would continuously lose energy and spiral into the nucleus. But it doesn't — no energy is radiated and it doesn't spiral inward. Something was wrong with Rutherford's model.

Some sort of modification was needed, and in 1913 it was the Danish physicist **Niels Bohr** who provided it by adding an essential idea — the quantum theory (from the Latin *quantus* = 'how much'). This raised a storm of controversy among scientists and philosophers. Bohr said that electrons could only exist in certain fixed orbits around the nucleus (Figure 27.16). He postulated that an electron in each orbit would move without losing energy or emitting radiation but when an electron jumped from one orbit to a lower one a single photon (a quantum) of light was emitted. Chapter 29 looks at the Bohr model and these spectral emissions in more detail.

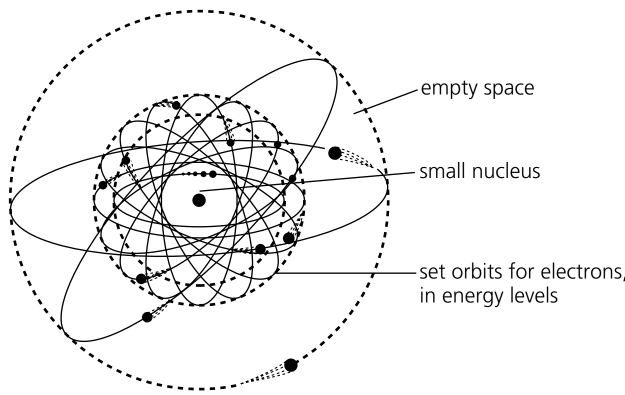


Figure 27.16
Bohr's model.

Questions

10 American physicist Douglas Giancoli remarked that students often say 'Don't tell us about the history and all the theories that were wrong. Tell us the facts as they are known today'. Critically evaluate this statement using examples from history.

11 If the Thompson model had been correct, how would the results of the Rutherford scattering experiment have been different?

THE MODERN STRUCTURE OF THE ATOM 27.9

We now know that an atom is made up of a dense nucleus composed of protons and neutrons and is surrounded by a cloud of electrons. About 99% of the mass is concentrated in the nucleus. If you had a nucleus the size of a strawberry it would bore its way through the ground. In Chapter 29 you will see that protons and neutrons have an internal structure — they are made up of particles called 'quarks'.

Chemical characteristics

The number of protons in an atom determines its name. Hydrogen has one proton, helium has two, uranium has ninety-two and so on. The number of protons is called the atom's **atomic number**, which has the symbol *Z*. Particles residing in the nucleus (protons and neutrons) are called **nucleons**. The different types of nuclei are referred to as **nuclides**; for example, there is a hydrogen nuclide, a helium nuclide and so on.

Mass and size

A proton or neutron has a mass 1836 times that of the electron, so in a hydrogen atom, for example, 99.95% of the mass is contained in the nucleus. However, the nucleus occupies only a tiny volume — about 1 trillionth the volume of the atom. If the nucleus were an orange seed, the outer region of the electron cloud would be 100 m away. How big is the electron? Well, the electron is considered to be a 'point' charge, so is not considered to have any size at all. Table 27.1 shows the masses and charges of some atomic particles.

Table 27.1 THE MASS AND CHARGE OF THE ATOMIC PARTICLES

PARTICLE	MASS (kg)	CHARGE (coulomb)
Proton	1.6726×10^{-27}	$+1.6 \times 10^{-19}$
Neutron	1.6750×10^{-27}	0
Electron	9.1095×10^{-31}	-1.6×10^{-19}

The simplest atom is that of hydrogen (Greek *hydro* = 'water', *genes* = 'to form'). It has one proton and no neutrons so its atomic number *Z* (number of protons) is 1 and its atomic mass *A* (p + n) is also 1. As it is an atom it is electrically neutral so it must have the same number of positive and negative charges; hence it has one electron. The number of protons determines the name of the atom; for example, all atoms with eight protons are called oxygen. If oxygen has eight electrons it is a neutral atom; if it has more electrons (e.g. 10) it is a negative ion (O^{2-}); fewer electrons (e.g. 6) makes it a positive ion (O^{2+}).

Isotopes

As more radioactive elements were discovered between lead (atomic number 82) and uranium (92), it became increasingly difficult to see how they could all fit into the ten atomic number spaces. Cases began to appear in which two radioactive elements, distinctly different in the type of radiation they emitted, were otherwise chemically identical. For example, in 1913 Frederick Soddy found that there were two forms of lead; one was the ordinary sort with a mass of 207 units and the other, taken from a uranium deposit in Norway, had a mass of 206 units. Soddy concluded that such inseparable elements must occupy the same place in the periodic table and gave them the same **isotopes** (Greek *iso* = 'same', *topos* = 'place'). Isotopes of an element have the same number of protons but different numbers of neutrons. In the example above, both isotopes have 82 protons but one has 124 neutrons (206 – 82) and the other has 125 neutrons (207 – 82).

The chemical properties of isotopes are identical but the physical properties are different because of the different masses. For example, hydrogen has three isotopes called hydrogen (H), deuterium (D) and tritium (T) as shown in Figure 27.17. Their composition is shown in Table 27.2. Water made with ordinary hydrogen (H_2O) has a density of 1.000 g/mL and boils at 100.00°C, whereas 'heavy water' made from deuterium oxide (D_2O) has a density of 1.1079 g/mL and a boiling point of 101.42°C.

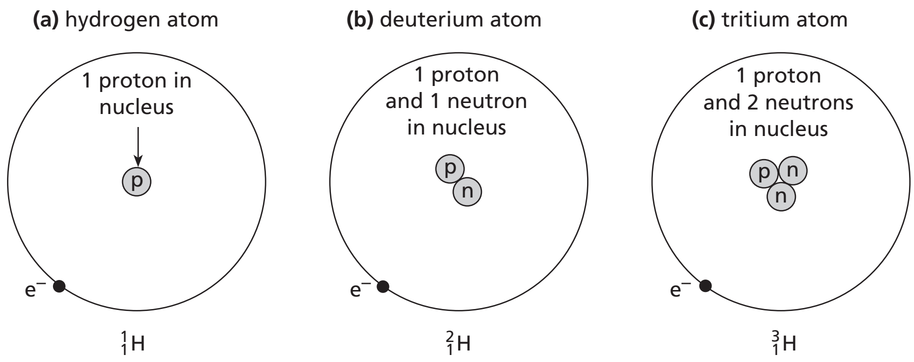


Figure 27.17
The isotopes of hydrogen: **(a)** hydrogen; **(b)** deuterium; **(c)** tritium.

Table 27.2 THE ISOTOPES OF HYDROGEN

ISOTOPE (NUCLIDE)	SYMBOL	PROTONS (ATOMIC NUMBER)	NEUTRONS	PROTONS + NEUTRONS (ATOMIC MASS)	ELECTRONS (= PROTONS)
Hydrogen	^1_1H	1	0	1	1
Deuterium	^2_1H	1	1	2	1
Tritium	^3_1H	1	2	3	1

Each isotope is referred to as a nuclide. The shorthand way of describing a nuclide is shown in Figure 27.18. For example, the isotopes of hydrogen are written as ^1_1H, ^2_1H and ^3_1H. The lead isotopes mentioned earlier would be written as $^{206}_{82}Pb$ and $^{207}_{82}Pb$ but can also be written as Pb-206 and Pb-207. (The atomic number 82 is not needed as all isotopes of lead have 82 protons.)

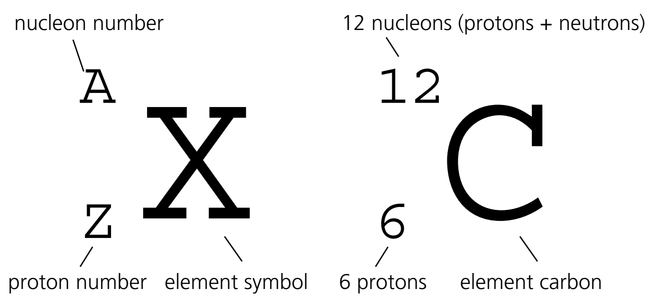


Figure 27.18
Notation of a nuclide. Can also be written as carbon-12 or just C-12.

Students often get confused when trying to work out symbols for nuclides. The most important thing to remember is that the atomic number Z (the number of protons) and the symbol go hand-in-hand. All isotopes of lead have an atomic number of 82 and all atoms with an atomic number of 82 are lead. There can be other nuclides with an atomic mass of 206 or 207 besides lead, for example $^{206}_{84}Po$ or $^{207}_{83}Bi$. The symbol and the atomic mass can be found by consulting the periodic table shown in Appendix 6 or the table in Appendix 7. Note that the atomic mass of an element is the actual mass of the atoms of an element as it is found in nature. Since most elements exist as isotopes, this mass is really a weighted average of the masses of the isotopes present. For example, ordinary carbon consists of 98.89% carbon-12, 1.11% of carbon-13 and a trace (1×10^{-8}%) of carbon-14. The average atomic mass is 98.89% of 12 plus 1.11% of 13 which equals 12.0111 units. This is called the **relative atomic mass** (A_r) of carbon. Only carbon-12 (the most abundant) has a relative atomic mass of exactly 12.0000.

Example

Magnesium is found naturally occurring on Earth as three different isotopes. The abundance of each isotope is as follows: $^{24}_{12}Mg$, 78.70%; $^{25}_{12}Mg$, 10.13%; $^{26}_{12}Mg$, 11.17%. Calculate the relative atomic mass of magnesium.

Solution

$$24 \times \frac{78.70}{100} = 18.888$$

$$25 \times \frac{10.13}{100} = 2.533$$

$$26 \times \frac{11.17}{100} = 2.904$$

Total (sum) = 24.32 units.

Most elements exist as a number of isotopes. Uranium contains three isotopes of atomic mass 234, 235 and 238. The particle composition of these isotopes is shown in Table 27.3. (The exact mass of a nuclide to four or more decimal places is in Appendix 8.)

Table 27.3 THE ISOTOPES OF URANIUM

ISOTOPE	ATOMIC NUMBER	PROTON NUMBER	ELECTRON NUMBER	NEUTRON NUMBER	ATOMIC MASS	EXACT MASS (u)
	(Z)	(Z)	(Z)	(A – Z)	(A)	
U-234	92	92	92	142	234	234.04098
U-235	92	92	92	143	235	235.04394
U-238	92	92	92	146	238	238.05082

Questions

You may need to refer to the periodic table to answer the following questions.

12 Carbon has two common isotopes of atomic mass 12 and 14. Write their symbols.

13 Helium exists in nature as two isotopes, of atomic mass 3 and 4. Prepare a table showing the particle composition of these isotopes.

14 Neutrons are bigger than electrons but have higher penetrating power. Explain why.

15 Write down the numbers of protons and neutrons in each of the nuclides represented by **(a)** $^{2}_{1}H$; **(b)** $^{12}_{6}C$; **(c)** $^{17}_{8}O$; **(d)** $^{23}_{11}Na$; **(e)** $^{32}_{16}S$; **(f)** $^{107}_{47}Ag$; **(g)** $^{127}_{53}I$; **(h)** $^{238}_{92}U$.

16 What is the average atomic mass of silicon that consists of 92.22% Si-28, 4.70% Si-29 and 3.08% Si-30?

27.10 MASS DEFECT AND BINDING ENERGY

When accurate measurements of the masses of nucleons and nuclei are made, a significant discrepancy emerges. **The mass of a nucleus is always less than the combined individual masses of its constituent nucleons**. For example, the mass of a deuterium nucleus is 3.34364×10^{-27} kg but the sum of the masses of the individual proton and neutron is 3.34760×10^{-27} kg.

Atomic mass units

The masses of neutrons, protons and electrons have been accurately measured using a mass spectrometer, in which the radius of curvature of a particle through known magnetic and electrostatic fields is determined (see Figure 25.30 in Chapter 25). The masses are expressed as 'unified **atomic mass units**', abbreviated *'amu'* or just *'u'*. **One unified atomic mass unit is one-twelfth the mass of an atom of the carbon isotope of atomic mass 12.0000 (that is, ^{12}C). The unified mass unit includes the masses of the six electrons of the carbon atom.** In kilograms, 1 u = 1.6606×10^{-27} kg.

The masses of atomic particles are as follows:

- proton (m_p) – 1.007 276 u
- neutron (m_n) – 1.008 665 u
- electron (m_e) – 0.000 549 u

A carbon-12 atom is made up of six protons, six neutrons and six electrons. The sum of these masses is as follows:

6 protons	6.043 656 u
6 neutrons	6.051 990 u
6 electrons	0.003 294 u
Total mass	12.098 940 u

However, the mass of a $^{12}_{6}C$ atom is 12.0000 u, which is less than the sum of its constituent particles. Where has the missing 0.098 940 u gone — into thin air?

The loss of mass was defined by scientist F. W. Aston as **mass defect** (Δm) and it represents the mass that has been converted to **binding energy**, energy that binds or holds the nuclear particles together. Holding six protons together in a nucleus is a difficult task as they are repelling each other. The binding energy is needed to overcome these huge electrostatic forces of repulsion. When we form a carbon nucleus we convert a mass of 0.098 94 u into binding energy. Einstein showed that mass and energy are equivalent and one can be converted into the other. Using his equation showing the relationship between mass and energy $E = mc^2$, we can calculate the amount of energy involved. The constant of proportionality is c^2 where c is the speed of light in a vacuum (3×10^8 m s^{-1}). The binding energy can be calculated:

$$E = 0.098\ 94\ \text{u} \times 1.6606 \times 10^{-27}\ \text{kg} \times (3.00 \times 10^8)^2 = 1.48 \times 10^{-11}\ \text{J}$$

Alternatively we could express the binding energy in units called **megaelectron volt** (MeV) where 1 u of mass defect is equivalent to 931 MeV. In the above case of carbon, a mass defect (Δm) of 0.098 94 u equals 0.098 94 u $\times$ 931 MeV/u = 92.1 MeV.

To break up the nucleus, you would have to expend the same amount of energy in the process.

The binding energy of a nucleus is the energy converted from mass when a nucleus is formed from its constituent nuclear particles, all initially in their free state.

Example

For the nuclide $^{4}_{2}He$, calculate **(a)** the mass defect; **(b)** binding energy in MeV per nucleon.

Solution

- Mass of $^{4}_{2}He$ (nucleus and electrons) = 4.002 603 u.
- Components: 2 p = $2 \times 1.007\ 276$ u = 2.014 552 u
  2 n = $2 \times 1.008\ 665$ u = 2.017 330 u
  2 e = $2 \times 0.000\ 549$ u = 4.032 980 u.
- Δm (mass defect) = 4.002 603 – 4.032 980 = 0.030 377 u.
- 1 unit of mass (1 u) = 931 MeV (megaelectronvolt) of energy.
  Hence: $^{4}_{2}He$ binding energy = $0.030\ 377 \times 931$ MeV
  = 28.28 MeV

Obviously, the more protons present, the greater the repulsive forces and hence the greater the binding energy necessary. For example, C-12 has a binding energy of 92.1 MeV whereas U-235 requires 1781 MeV. However, expressing the required energy as the **binding energy per nucleon** gives a measure of how stable the nucleus is. The greater the binding energy per nucleon, the more stable it is.

For example, C-12 has 12 nucleons so 92.1 MeV ÷ 12 = 7.67 MeV/nucleon. U-235 has only 1781 ÷ 235 = 7.58 MeV per nucleon. The carbon nucleus is more stable.

Variation in binding energy

If we plot a graph of the binding energy per nucleon against the atomic mass (see Figure 27.19), the most stable nuclides are apparent — those with atomic masses around 50–60 u, for example Fe-56. Those heavy nuclides with atomic masses to the right of this region are most likely to undergo nuclear fission and break into smaller nuclides to become more stable. Light nuclei, to the left of Fe-56, are more likely to undergo nuclear fusion and join together to form heavier nuclides. Fission and fusion are described in the next chapter.

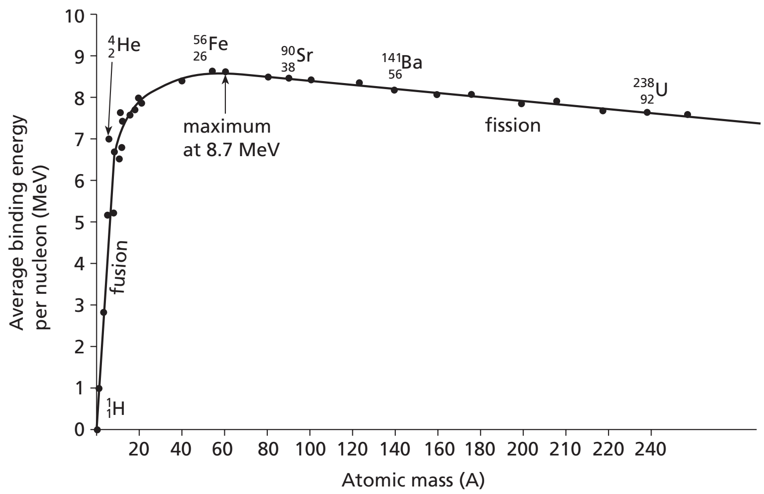


Figure 27.19
Graph showing the binding energy per nucleon as atomic mass changes.

Question

17 Use the table of exact atomic masses given in Appendix 8 to calculate the binding energy in **(i)** MeV and **(ii)** MeV per nucleon of **(a)** a tritium atom $^{3}_{1}$H, **(b)** helium-3 ($^{3}_{2}$He), **(c)** N-14, **(d)** O-16.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***18** In a Millikan oil drop experiment, an oil drop of mass 1.077×10^{-15} kg is held stationary between a pair of electric plates held 2.0 cm apart. The voltage across the plates is 110 V. Assume $\boldsymbol{g}$ = 9.8 m s^{-2}.
(a) What is the magnitude of the electric field between the plates?
(b) What is the charge on the oil drop? How many elementary charges is this equivalent to?

***19** Write down the number of protons and neutrons in **(a)** $^{207}_{82}$Pb; **(b)** $^{35}_{17}$Cl; **(c)** $^{15}_{7}$N; **(d)** At-215; **(e)** Bi-216.

***20** **(a)** Distinguish between *nucleon*, *nucleus*, *nuclide* and *neutron*. **(b)** Do *neutron* and *nucleus* come from the same Greek roots? If not, what is the difference?

***21** Calcium has four isotopes with atomic masses of 40, 42, 43 and 45. **(a)** Write their symbols. **(b)** How many **(i)** neutrons and **(ii)** nucleons does each have?

Ch27-Walding 4th 25/8/04 1:33 PM Page 615

22 Calculate **(a) the mass defect and **(b)** binding energy (MeV/nucleon) for **(i)** ${}^{35}_{16}S$; **(ii)** ${}^{235}_{92}U$; **(iii)** Cd-113; **(iv)** Li-7. See Appendix 8 for masses.

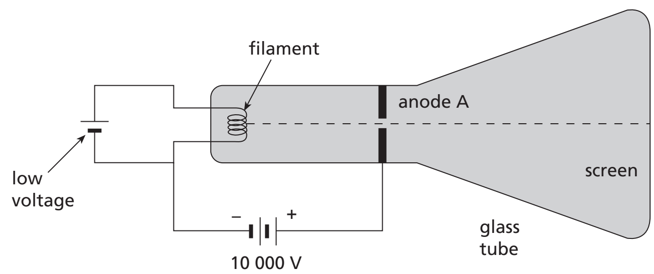
***23 Discuss the concept that progress in science depends on the development of the necessary equipment and technology to perform key experiments. Cite examples from history to illustrate your arguments.

**24 Figure 27.20 shows a simple arrangement of a cathode ray tube.

(a) Briefly explain why **(i)** the filament is heated; **(ii)** the electrons accelerate between the filament and the anode A.

(b) What effect, if any, is there on the number of electrons to the right of the anode if **(i)** the low voltage supply is increased; **(ii)** the high voltage supply is increased?

Figure 27.20
For question 25.



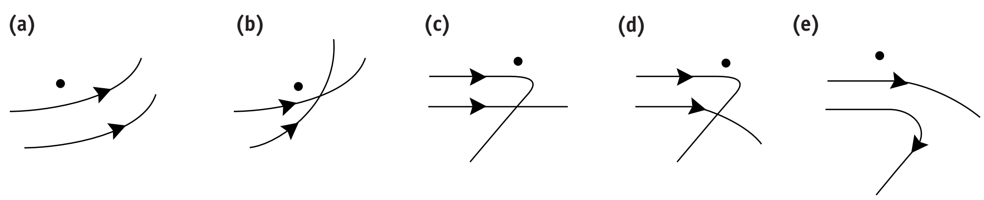
Extension — complex, challenging and novel

***25 In a Millikan oil drop experiment, the potential difference between two large horizontal plates, 2.4 cm apart, is 1200 V. A tiny plastic sphere with a deficiency of three electrons is introduced into the electric field between the plates and it remains stationary. What is the mass of the sphere?

***26 Charged latex spheres can be used in Millikan experiments. One particular negatively charged sphere of mass 3.52×10^{-14} kg is held stationary between two charged horizontal plates by an electric field of strength 9.8×10^{4} V m^{-1}. How many excess electrons are there on the sphere?

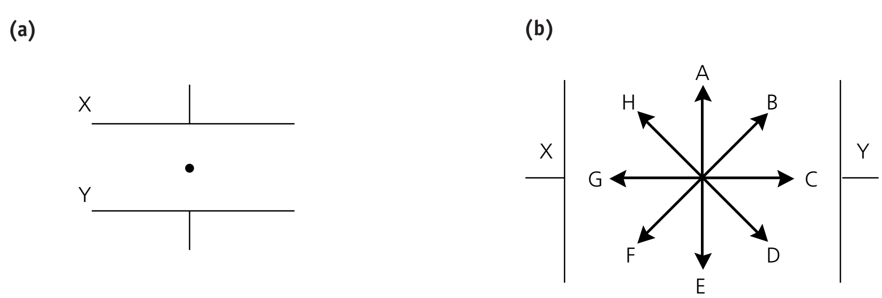
27 Figures 27.21(a)** to **(e)** show paths of two α particles being deflected by a heavy nucleus. Which of the diagrams correctly represent(s) the deflection of two α particles of the same energy?

Figure 27.21
For question 28.



***28 An oil drop is held stationary between a pair of horizontal plates X and Y as shown in Figure 27.22**(a)**. The oil drop has a charge of $+6.4 \times 10^{-19}$ C. If the plates are turned through 90° as shown in Figure 27.22**(b)**, in which direction will the droplet now move?

Figure 27.22
For question 29.



Ch28-Walding 4th 25/8/04 1:37 PM Page 616

CHAPTER 28

Nuclear Physics

28.1 THE ATOMIC AGE

Radioactivity and nuclear energy have become some of the most important concerns facing society since the Second World War. The benefits to society are immense but so too are the problems they bring. In this chapter we will be examining some aspects of nuclear physics that will help to answer questions such as these:

- When you irradiate food with gamma rays, does the food become radioactive?
- Can you accurately tell the age of bones millions of years old?
- If gamma radiation can go straight through the body, how can it kill cancer tumours?
- Why do you need a fission bomb to start a fusion bomb?
- Why does an airline pilot get exposed to more radiation than an air traveller?
- I thought electrons were negatively charged; how can you get a positive one?

28.2 PROPERTIES OF NUCLEAR RADIATION

In the previous chapter, the history and properties of nuclear radiation were described. Let's expand on these:

- Any radiation that can remove an electron from an atom and create a heavy positive ion and free electron is termed **ionising radiation**. Ionising radiations include electromagnetic radiation (gamma rays, X-rays, and ultraviolet radiation) as well as energetic particles such as alpha and beta particles. Gamma rays are said to be **nuclear radiation** because they are created within the nucleus; X-rays come from the electron cloud around the nucleus.
- By the early 1900s, the properties of alpha, beta and gamma radiation had been measured, allowing physicists to better understand the process of radioactivity. The deflection of the particles in a magnetic field is shown in Figure 28.1**(a)** and how they pass through matter is shown in Figure 28.1**(b)**.

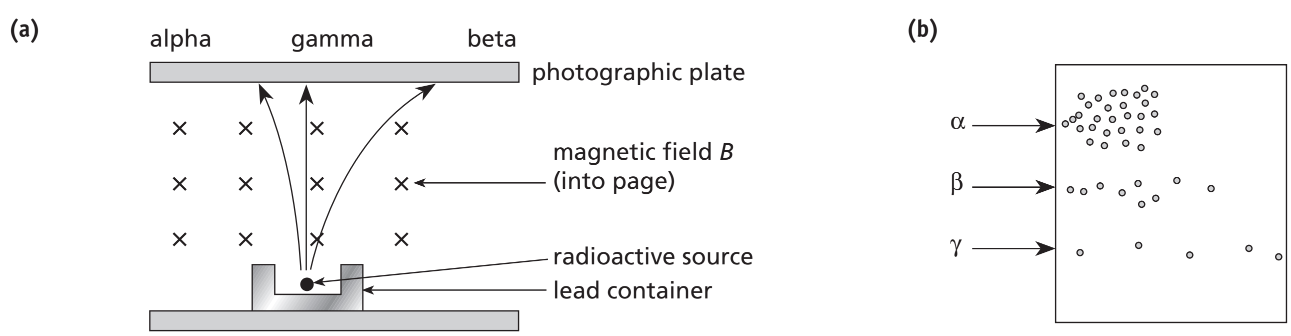


Figure 28.1
(a) Deflection of alpha, beta and gamma rays by a magnetic field.
(b) Ionising ability.

Alpha (α) particles ${}^{4}_{2}He$

As they collide with matter, alpha particles slow down, transferring their kinetic energy to the other molecules, shaking many of them apart and leaving a trail of positive and negative ions in their wake.

Beta (β) particles ${}^{0}_{-1}e$

Beta particles are electrons moving at high speed ranging from 0.3 to 0.99 times the speed of light (3×10^8 m s^{-1}). Because of their speed and smallness, they are more penetrating than alpha particles and can travel about 1 m in air before slowing down to become just like the surrounding electrons.

Gamma (γ) rays

Gamma radiation differs from alpha and beta radiation in that it is not made up of charged particles and is not deflected in electric or magnetic fields. Instead, gamma rays are electromagnetic radiation of extremely short wavelength (about 10^{-13} m). Since they have no charge they have tremendous penetrating power because they interact with the absorbing material only via a direct head-on collision with an electron or nucleus. Materials such as lead are good absorbers of gamma radiation mainly because of their high electron density. Even so, gamma rays can still penetrate up to 10 cm of lead.

DETECTING NUCLEAR RADIATION 28.3

One of the most common means of detecting radiation is by Geiger–Müller counters but also used are photographic plates, electroscopes, spark chambers and cloud chambers, and by fluorescence.

Fluorescence

When ionising radiation strikes certain substances such as ZnS, diamond or barium platinocyanide, a large number of individual flashes or scintillations can be seen under a microscope. Tedious counting of flashes over a set period was used by Rutherford and his co-workers in the early 1900s. These **scintillation detectors** slowly lost favour until 1947, when **photomultiplier tubes** were developed to count the scintillations electrically. Today, semiconductor detectors are used.

Photographic plates

Becquerel's discovery of radioactive emissions was based on the fogging of photographic plates. As the ionising particles strike the silver chloride or bromide grains in the gelatin emulsion on the plate they change them into silver atoms. On development, the silver is 'fixed' and the unaffected salts are removed. This leaves a permanent photographic record of the particles' tracks.

EI Activity 28.1 PHOTO DETECTIVE

If you print your own photos or know someone who does, you may like to test the effect of nuclear radiation on an unexposed sheet of photographic paper.

1 Place a sheet of the paper in one of the black plastic bags it normally comes in and tape it closed.

2 Under teacher supervision, place three radioactive specimens, one each of α, β and γ-type radiation, on top of the plastic for an hour and record their positions.

3 Develop the paper and see if the results agree with both the penetrating properties mentioned above and the ability of the radiation to fog the paper.

4 Was the image clear? How could you test it without the plastic bag?

Electroscope

Marie Curie measured the activity of fluorescent salts using an electrometer, invented by her husband Pierre and his brother Jacques. It is based on the principle of the **electroscope**, which should be familiar to you (Figure 28.2). When the air surrounding a negatively charged electroscope is ionised, the positive ions will be attracted to the electroscope and cause the leaves to collapse. The rate of collapse will be proportional to the ionisation produced in the air above. The **electrometer** is similar but has a pointer instead of leaves.

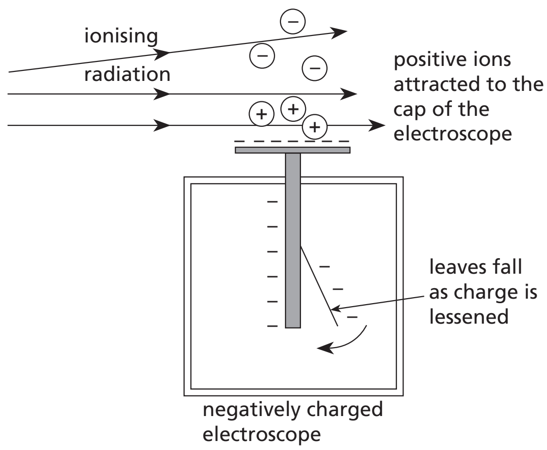


Figure 28.2
A leaf electroscope.

Geiger–Müller counters

Figure 28.3 shows a **Geiger–Müller tube**, commonly known as a **Geiger counter**. It consists of a thin positively charged central wire surrounded by a negatively charged tube filled with a low pressure inert gas. When a radioactive particle enters the tube through the window, it ionises a few atoms. The resulting free electrons are drawn to the positive wire. However, the electric field is so strong that these electrons gain sufficient energy to ionise more atoms of gas. More free electrons are created and the process is repeated many times. This avalanche of electrons is collected by the central wire, creating a signal used to record the passage of the original particle of radiation.

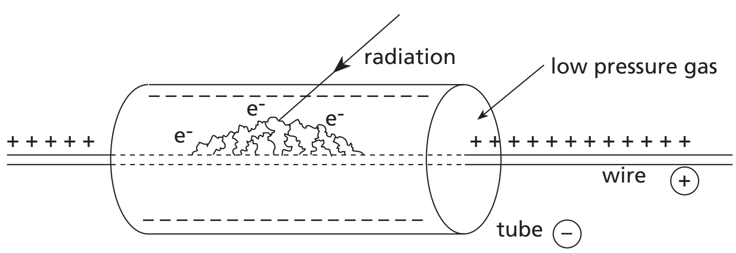


Figure 28.3
A Geiger–Müller tube.

Cloud chambers and bubble chambers

The **cloud chamber** was invented by Scotsman Charles Wilson (1869–1959) and is based on the tendancy of drops of moisture to condense on gaseous ions. When an ionising particle passes through air that is supersaturated with water or other vapour, the liquid droplets form a visible 'cloud track' indicating the path of the particle. Cloud chambers consist of a small container in which a small amount of alcohol is added and the resulting vapour cooled with dry ice (–40°C) to cause condensation. Vapour trails behind jet aircraft are from a similar process of condensation. The tracks of alpha particles can be readily seen as white lines against a black background. They are short, wide tracks. Beta particles produce longer thinner tracks.

Figure 28.4
A cloud chamber.

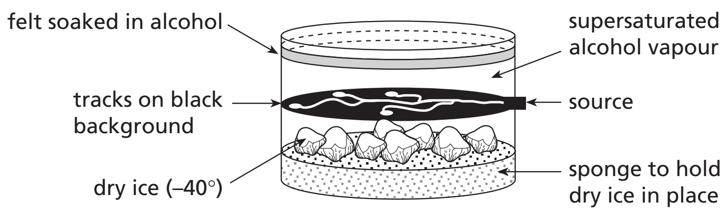


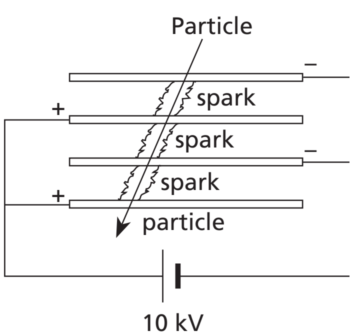
Photo 28.1
A bubble chamber.

In 1952 Glasser invented the **bubble chamber**, which uses a superheated liquid (e.g. liquid H_2 or isopentane) instead of a supercooled vapour. Radioactive particles ionise the liquid and the resulting positive ions provide sites for the formation of bubbles from the boiling liquid. The bubbles show the path of the radiation. The advantage of the bubble chamber is that it can show tracks of very short-lived high-energy particles such as the type produced by particle accelerators. The low gas density in the cloud chamber is insufficient to cause interactions with these short-lived particles.

Spark chambers

A spark chamber consists of a set of parallel plates spaced closely together. Alternate plates are grounded and the ones in between are kept at a very high voltage (about 10 kV). When a charged particle passes them, the ions produced in the gas leave a trail of ions and electrons between the plates, providing a conducting path for a spark to jump. (See Figure 28.5.)

Figure 28.5
Path of an ionising particle in a spark chamber.



SR Activity 28.2 INTERFACING THE GEIGER COUNTER

After the Chernobyl nuclear power plant explosion in the Ukraine in 1986, Russian authorities tried to withhold news of the leak for as long as possible. Meanwhile, in London, several senior physics students had interfaced their school's Geiger counter to a computer and left it running over the weekend to monitor background radiation in London. They were completely unaware of the accident in Russia. When they examined the data on the Monday morning, they were astonished to see that they had logged incredibly high radiation levels. They thought it was a mistake but they soon realised that they were the first people in England to detect that the radioactive plume had crossed the English Channel. They beat all the Government's sophisticated monitoring equipment. What a buzz!

Most school Geiger counters have sockets for output to external devices. Can you design a simple interface that will enable you to connect the Geiger counter to one of the ports of a computer? *Hint:* you could use the parallel port and write a program that reads when the appropriate pin goes 'high' and increments a counter. This will be no easy task! You will need expert help but you could probably sell it to your school.

Questions

1 Each of the devices mentioned has different strengths and weaknesses. Make a list to compare each device.

2 **(a)** List radiation detectors that make use of **(i)** ionising ability; **(ii)** the photoelectric effect. **(b)** Both the cloud chamber and the bubble chamber show the tracks of radioactive emissions. Explain the basic difference in the way they work.

Ch28-Walding 4th 25/8/04 1:37 PM Page 620

28.4 NUCLEAR STABILITY AND RADIOACTIVE DECAY

Transmutations

For centuries, medieval researchers sought in vain for a material that could turn ordinary metals such as lead into the precious metal gold — a process they called **transmutation** (Latin *trans* = 'across', *mutare* = 'change'). These people were called *alchemists* (Greek *chyma* = 'to fuse a metal'). They developed numerous techniques that are still used in laboratories today, such as distillation, crystallisation, sublimation and fusion. They derived many useful compounds in the process, such as caustic soda, red lead, tin oxide and various alloys; but they used many substances not commonly found in laboratories today: hair, skull, brains, bile, blood, milk, urine and horn. No wonder they didn't turn lead into gold. But it wasn't the chemicals that caused them to fail; it was their underlying theory of matter. The alchemists believed in Aristotle's four-element theory (earth, wind, fire and water). Until this was overthrown, science couldn't progress.

Little did they know that in nature many atoms transmute from one form to another in the normal course of events. When atoms transmute they are said to be radioactive and particles are emitted by the nucleus. These particles can be alpha, beta or a neutron. In many cases a gamma ray is also emitted. The question that bothered physicists was why some nuclei were radioactive (e.g. uranium) and why some were stable (lead) and stayed unchanged indefinitely. The answer lies in the way the nucleus is structured.

Nuclear forces

The nucleus is made up of protons and neutrons. However, the protons are positively charged and because they are almost touching, the electrostatic repulsive force is enormous. But why don't they fly apart? The answer lies in the role of the neutrons. The neutrons serve two purposes — first, to add some distance between the protons to reduce the repulsive force and, second, to act as a nuclear 'glue'. This gluing force is called the '**strong force**'. As the number of protons increases, the number of neutrons must also increase. This 'strong force' binds adjacent nucleons together. It is a very short-ranged force because, unlike the electrostatic force that decreases as the inverse square of the distance, the 'strong force' decreases rapidly. When nucleons are just a few diameters apart, the 'strong force' is nearly zero.

For the smaller nuclei, the number of neutrons required for stability is about the same as the number of protons present (that is, a n/p ratio of 1:1). For example, the most stable isotope of oxygen is $^{16}_{8}O$. It has eight protons and eight neutrons. Similarly, ordinary carbon is $^{12}_{6}C$ (6p, 6n). But as the number of protons increases, the number of neutrons required for stability increases more — the n/p ratio becomes greater than 1:1. For example, stable zinc is $^{66}_{30}Zn$, which has thirty protons and thirty-six neutrons so the n/p ratio is 1.17. Table 28.1 lists some common examples. Note that the n/p ratio starts at 1.00 and has increased to 1.59 for uranium. Figure 28.6 shows a plot of neutron number against proton number. There are no completely stable nuclides above a proton number (Z) of 82.

Table 28.1 SOME N/P RATIOS

STABLE NUCLIDE	PROTONS (Z)	NEUTRONS (N)	n/p RATIO
$^{16}_{8}O$	8	8	1.00
$^{65}_{30}Zn$	30	35	1.17
$^{207}_{82}Pb$	82	125	1.52
$^{238}_{92}U$	92	146	1.59

Ch28-Walding 4th 25/8/04 1:37 PM Page 621

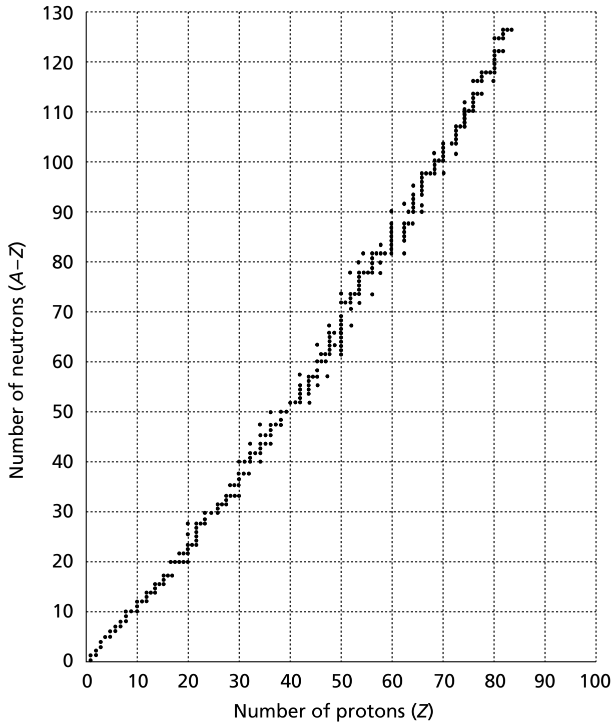


Figure 28.6
Graph of a proton (or atomic) number *Z*, versus a neutron number *N*, for naturally occurring isotopes.

The particle 'zoo'

The 1930s were to witness a new burst of radioactivity research even greater than the two previous bursts in 1895 (Curie, Thompson, Becquerel) and 1912 (Rutherford, Bohr). In 1932 Chadwick discovered the neutron and, very soon afterwards, American physicist Carl Anderson discovered another fundamental particle, the positive electron or **positron**. The existence of the positron had been predicted by Paul Dirac several years earlier. This was a major development in physics. Further discussion of the positron and other fundamental particles is reserved for the following chapter. The important particles discussed so far are shown in Table 28.2.

Table 28.2 SYMBOLS OF ATOMIC PARTICLES

PARTICLE	SYMBOL
Alpha particle (α)	${}^4_2He$
Proton	${}^1_1H$ or ${}^1_1p$
Neutron	${}^1_0n$
Electron (β particle)	${}^{\ 0}_{-1}e$
Positron	${}^{\ 0}_{+1}e$
Gamma ray	γ

Radioactive decay

The world is made up of stable nuclei — all atoms don't just disintegrate in front of us. But there are some atoms that are radioactive and decay or disintegrate into other types of atoms. Some nuclei, such as Pa-221, may only last for 6 microseconds whereas lead-206 will last for billions of years and is said to be infinitely stable. Between these two extremes are nuclei that may exist for seconds, hours, days or years before decaying. Lead-214 will, on average, last for 27 minutes before giving off β and γ particles.

When the original unstable 'parent' nucleus decays it produces a daughter nucleus and at least one other particle. The reaction can be written like the normal chemical equation:

$$\text{Parent nucleus} \rightarrow \text{daughter nucleus} + \text{particle(s)}$$
$$^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\text{He}$$

In all cases it can be seen that:

- **The sum of the atomic masses on the left of the equation must be the same as the sum of the atomic masses on the right**.
  In the case above the 234 + 4 = 238, so this rule is obeyed.
- **The total charge on the left-hand side of the equation must equal the total charge on the right-hand side**.
  Charge refers to the nuclear charge and that of its emitted particles. A proton has a charge of +1; an electron (β particle) has a charge of −1. In the case above, the left-hand side shows 92 protons, so the charge is +92 or simply 92. The sum of the nuclear charges on the right is 90 + 2 = 92; so the rule is obeyed.
  Another important rule is:
- **The number of protons (atomic number) determines the name and symbol of the element**.
  A mistake students often make when trying to work out the symbol for the daughter nucleus is to use the atomic *mass* and not the atomic *number*.

Example

Balance the following radioactive decay:

$$^{226}_{88}\text{Ra} \rightarrow {}^{4}_{2}\text{He} + ?$$

Solution

Step 1: The atomic masses (the top numbers) have to be equal to 226 on both sides; the daughter nucleus must have an atomic mass of 226 − 4 = 222.
Step 2: The nuclear charges (the bottom numbers) have to be equal to 88: the daughter must have a charge of 88 − 2 = 86.
Step 3: The nuclear charge (atomic number) determines the name of the element; $Z = 86$ refers to *radon*.
Note: the top number (222) is *not* used to find the name of the atom.
Step 4: Write equation: $^{226}_{88}\text{Ra} \rightarrow {}^{4}_{2}\text{He} + {}^{222}_{86}\text{Rn}$.

Questions

3 State the number of neutrons and protons in the following: **(a)** $^{226}_{88}\text{Ra}$; **(b)** $^{1}_{1}\text{H}$; **(c)** $^{239}_{93}\text{Np}$.

4 What element is represented by X in each of the following: **(a)** $^{226}_{88}\text{X}$; **(b)** $^{1}_{1}\text{X}$; **(c)** $^{247}_{97}\text{X}$; **(d)** $^{82}_{38}\text{X}$?

5 Balance the following equations:

(a) $^{214}_{83}\text{Bi} \rightarrow {}^{0}_{-1}\text{e} + ?$

(b) $^{239}_{93}\text{Np} \rightarrow {}^{0}_{+1}\text{e} + ?$

(c) $^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + ?$

(d) $^{45}_{20}\text{Ca} \rightarrow {}^{0}_{-1}\text{e} + ?$

(e) $^{58}_{29}\text{Cu} \rightarrow {}^{0}_{+1}\text{e} + ?$

(f) $^{234}_{94}\text{Pu} \rightarrow {}^{4}_{2}\text{He} + ?$

TYPES OF DECAY 28.5

There are three main ways that nuclei decay naturally. They are alpha decay, beta decay and positron decay.

These types of decay are associated with three unstable states of a nuclide:

- too many neutrons (beta decay)
- too many protons (positron decay)
- too many protons and neutrons — too much mass (alpha decay).

Alpha decay

Atoms heavier than uranium-238 do not occur naturally. We can produce them artificially but they have too many neutrons and protons to be stable; in other words, they have too much mass for the nuclear 'glue' to work. Such atoms decay by alpha emission and the parent nucleus loses two protons and two neutrons as a fast moving, energetic alpha particle. The alpha particles emitted have discrete kinetic energies, usually up to 10 MeV.

$$^{A}_{Z}X \longrightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}He + \text{energy } (\gamma)$$

$$^{236}_{91}Pa \longrightarrow {}^{232}_{89}Ac + {}^{4}_{2}He + \gamma$$

Alpha decay occurs because the 'strong nuclear force' is unable to hold large nuclei together. Because the strong nuclear force is a short-range force, it acts only between neighbouring nucleons. But the repulsive electrostatic force can act across the whole nucleus. For very large nuclei, the large number of protons means that the total repulsive force is great compared with the attractive strong nuclear force, which cannot hold the nucleus together. The only way to achieve stability is to shed some protons and neutrons. This occurs in packets of 2 p and 2 n, that is, the helium nucleus $^{4}_{2}He$, known as an alpha (α) particle.

Beta decay

There are two types of beta particles: beta minus (the electron) and beta plus (the positron). Each type is associated with a different type of instability in a nuclide. From now on, *beta decay* will refer to the electron (β–) and *positron decay* will be used for β+.

Beta decay occurs when there is a surplus of neutrons.

A beta particle is an electron that has come from the nucleus. The symbol $^{0}_{-1}e$ stands for an electron whose charge is –1 (i.e. Z = –1) and negligible atomic mass (A = 0). When a parent nucleus emits a beta particle, the daughter nuclide produced has the same mass number as the parent:

$$^{A}_{Z}X \longrightarrow {}_{Z+1}^{A}Y + {}^{0}_{-1}e + \text{energy } (\gamma)$$

Two examples of beta decay are:

$$^{239}_{92}U \longrightarrow {}^{239}_{93}Np + {}^{0}_{-1}e$$

$$^{234}_{91}Pa \longrightarrow {}^{234}_{92}U + {}^{0}_{-1}e$$

It can be seen that the number of nucleons has not changed but the daughter has one more proton than the parent. It is as if one of the neutrons has changed into a proton and in the process (to conserve charge) has given off an electron. In fact, neutrons actually do decay in this manner. Outside the nucleus, neutrons only last for about 11 seconds before this reaction occurs.

$$^{1}_{0}n \rightarrow {}^{1}_{1}p + {}^{0}_{-1}e$$

In the 1920s, accurate measurements of the masses of reactants and products in beta decay showed that the masses were not equal. Some mass appeared to be lost. Physicists were troubled at the prospect of the law of conservation of mass being violated. In 1930, Wolfgang Pauli proposed an alternative solution: the missing mass was being carried off by a particle of zero charge and negligible mass that made it difficult to detect. The Italian physicist Enrico Fermi (1901–54) suggested the name **neutrino** meaning 'little neutral one'. The symbol for the neutrino is the Greek letter *nu* (ν). A bar is place over the symbol ($\bar{\nu}$) to indicate that in beta decay an antineutrino is formed. An **antineutrino** is an **antiparticle** to the normal neutrino — more on that in Chapter 29. For the moment, you do not have to include the neutrino in your equations (unless your teacher says so). Neutrinos are flooding through your head right now but nothing much stops them. In fact, it has been estimated that you would need a lead block 90 light years thick to stop about 50% of them.

Positron decay

When a nuclide has a surplus of protons it undergoes positron decay:

$$^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + {}^{0}_{+1}e + \text{energy } (\gamma)$$
$$^{239}_{93}Np \rightarrow {}^{239}_{92}U + {}^{0}_{+1}e$$

A positron is a positive electron ${}^{0}_{+1}e$, or e^+. It has the same mass as an electron ($A = 0$) but it has the opposite charge, so its atomic number is said to be $Z = +1$, the same as its charge. It is sometimes called a 'beta plus' ($\beta+$) to distinguish it from ordinary beta radiation, which is then called beta minus (β^-). In this book we will use the terms *beta radiation* meaning an electron and *positron radiation* as its opposite or antiparticle. In positron decay, the number of nucleons (the top number) does not change. But as the number of protons decreases, it appears that a proton has been changed into a neutron and a positron ejected. When a positron and an electron collide they annihilate each other and give out a burst of gamma rays, which travel in opposite directions, a process made use of in positron emission tomography (PET). This is discussed in Chapter 33 on Medical Physics.

English physicist Paul Dirac predicted the existence of the positron in the late 1920s and American physicist Carl Anderson was awarded the Nobel prize in 1936 for his discovery of it in a cosmic ray shower. In fact, it is now believed that all particles have antiparticles, protons and antiprotons for example.

Predicting the type of decay

The graph of proton number versus neutron number is repeated in more detail in Figure 28.7. From it you can see what type of decay will occur in unstable nuclides.

- Beta decay will occur in those nuclides that are *above* the line of stability.
- Positron decay will occur in those nuclides that are *below* the line of stability.

Beta decay is typical in nuclides with a higher proportion of neutrons than in the stable isotopes of the same element. The opposite is true for positron decay, so that nuclides that undergo these two forms of decay lie on either side of the band of stable nuclides.

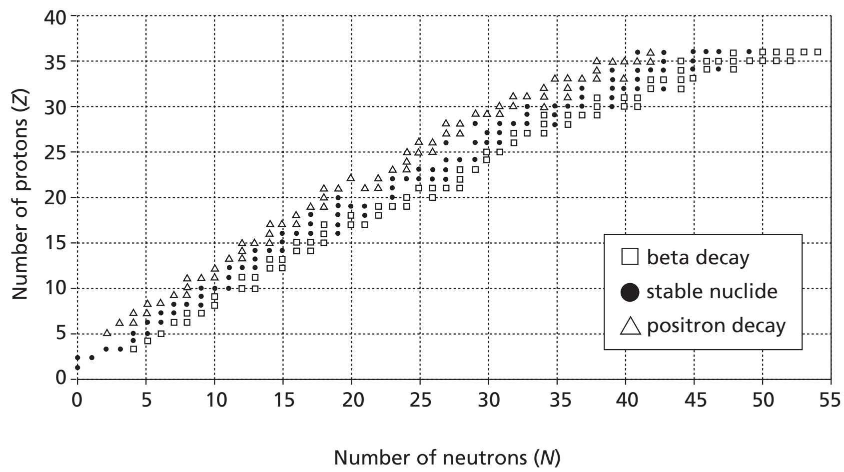


Figure 28.7
Predicting the type of decay from the stability graph. The band of stable nuclei is shown by the black dots. Note that in this graph the number of neutrons is on the x-axis.

Gamma decay

In pure gamma decay there is no change to the numbers of protons or neutrons and therefore no transmutation. It is a release of energy in the form of electromagnetic radiation of a particular wavelength and frequency.

Remember, a gamma ray is an electromagnetic radiation that originates from within the nucleus. An X-ray, on the other hand, is an electromagnetic radiation originating from the electron cloud surrounding the nucleus.

Questions

6 Complete the following equations:

(a) ${}^{32}_{15}P \rightarrow {}^{0}_{-1}e + ?$

(b) ${}^{234}_{90}Th \rightarrow ? + {}^{234}_{91}Pa$

(c) ${}^{22}_{11}Na \rightarrow ? + {}^{0}_{+1}e$

(d) ${}^{1}_{0}n \rightarrow ? + {}^{1}_{1}H$

(e) ${}^{1}_{1}H + {}^{0}_{-1}e \rightarrow ?$

7 What are the parent nuclei for each of the following daughter nuclei produced by alpha decay: **(a)** ${}^{206}_{81}Tl$; **(b)** thallium-210; **(c)** Po-218; **(d)** Pb-206?

8 Write equations for the beta decay of **(a)** C-14; **(b)** Na-24; **(c)** P-32.

9 Write equations for the positron decay of **(a)** ${}^{22}Na$; **(b)** ${}^{18}F$; **(c)** ${}^{19}Ne$; **(d)** ${}^{199}Pb$.

HALF-LIFE

28.6

Radioactive decay is a random event, just as a car accident is a statistically random event. There is no means of predicting whether a particular driver will be involved in a crash. However, it is possible to say statistically how many crashes will happen in Australia each year. The more cars, the more accidents.

In a similar way, it is quite impossible to predict when a particular nucleus will decay but it is possible to predict the number of nuclei that will decay in a given time from a particular source. Likewise, if you throw 100 coins into the air, you can reliably predict that about 50 will come down heads and 50 tails; but you can't say what a particular coin will do.

For example, N-13 decays to C-12 by positron emission:

$$ {}^{13}_{7}N \rightarrow {}^{12}_{6}C + {}^{0}_{+1}e $$

Ch28-Walding 4th 25/8/04 1:37 PM Page 626

If we had 10 000 N-13 atoms, about 12 of them would decay in 1 second (12 s^{-1}). We say the **decay rate** or **activity** (A) is 12 disintegrations per second (dps) or A = 12 s^{-1}. So after 1 second there would be 9988 left. After another second about 12 more N-13 atoms would decay and there would be 9976 left. But as the number of N-13 atoms decreases so too does the decay rate. After 10 minutes, there would be about 5000 N-13 atoms left (i.e. half the starting number) and the decay rate would be 6 disintegrations per second. When another 10 minutes had elapsed, there would be about 2500 atoms of N-13 left and the decay rate would be down to 3 dps. The period of 10 minutes in which it takes half the N-13 atoms to decay is called the **half-life** ($t_{1/2}$). It also represents the time taken for the decay rate to fall to half its original rate. The SI unit for activity is **Becquerel** (Bq), so 1 Bq = 1 disintegration per second (1 dps or 1 s^{-1}). In less active substances, the activity may be expressed as disintegrations per minute (dpm).

Table 28.3 summarises the data. The decay is shown graphically in Figure 28.8.

Table 28.3 DECAY OF NITROGEN-13 ISOTOPE

TIME ELAPSED (t) (MINUTES)	NUMBER OF N-13 ATOMS REMAINING (N)	DECAY RATE (A) (Bq)
0	10 000	12
10	5000	6
20	2500	3
30	1250	1.5
40	625	0.75

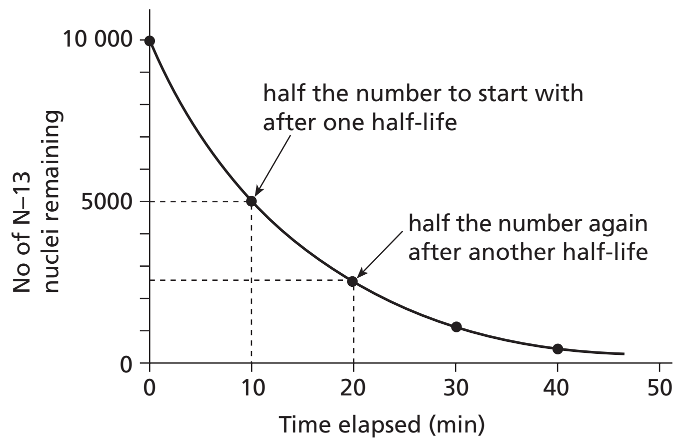


Figure 28.8
Graph of nitrogen-13 decay.

Half-lives can be very short or very long (Table 28.4).

Table 28.4 HALF-LIFE OF SOME ISOTOPES

NUCLIDE	DECAY	HALF-LIFE
^{206}Pb	stable	infinite
^{234}U	α	4.51 billion years
^{234}Th	β, γ	24.1 days
^{222}Rn	α	3.82 days
^{218}Po	α	3.05 min
^{214}Po	α	1.64×10^{-4} s
^{1}H proton	γ	10^{37} years

NOVEL CHALLENGE

Other things in nature have a half-life. For example, $t_{1/2}$ for human protein is 80 days, for rat muscle it is 21 days, and for rat blood 6 days. In non-mammals half-life values are much lower because they have lower temperatures. How is this different from nuclear half-lives?

The half-life of a radioactive isotope is the time taken for half the radioactive atoms in a sample to decay.

If N_0 = number of parent nuclei at start, and N = number of atoms at end of the time period:

- after 1 half-life: $N = N_0 \times \frac{1}{2}$
- after 'n' half-lives: $N = N_0 \left(\frac{1}{2}\right)^n$.

Example:

Iodine-131 has a half-life of 8 days and undergoes beta decay according to the equation: ${}^{131}_{53}\text{I} \rightarrow {}^{131}_{54}\text{Xe} + {}^{0}_{-1}\text{e}$. If a milk sample contains 3×10^{18} atoms at a particular time, calculate **(a)** the number remaining after 60 days; **(b)** the time that would have to elapse for there to be 1 million atoms left.

Solution

(a) Number of half-lives elapsed (n) = $\dfrac{60}{8}$ = 7.5.

$$N = N_0\left(\frac{1}{2}\right)^n = 3 \times 10^{18}\left(\frac{1}{2}\right)^{7.5} = 1.6 \times 10^{16} \text{ atoms}$$

Note: one way to do this on your calculator is to enter: 3 EXP 1 8 × . 5 x^y 7 . 5 =

(b)

$$N = N_0\left(\frac{1}{2}\right)^n$$

$$1 \times 10^6 = 3 \times 10^{18}\left(\frac{1}{2}\right)^n$$

$$\frac{1 \times 10^6}{3 \times 10^{18}} = \left(\frac{1}{2}\right)^n$$

$$3.333 \times 10^{-13} = \left(\frac{1}{2}\right)^n$$

$$\log 3.333 \times 10^{-13} = \log\left(\frac{1}{2}\right)^n = n \log\left(\frac{1}{2}\right)$$

$$-12.48 = n \times -0.301$$

$$n = \frac{-12.48}{-0.301} = 41.46 \text{ half-lives}$$

$$= 41.46 \times 8 \text{ days}$$

$$= 332 \text{ days}$$

Note: instead of specifying N and N_0 as meaning the number of atoms, it could also mean the mass of the nuclide in a specimen. The formula still works.

Decay series

Sometimes a radioactive isotope decays into another isotope that is also radioactive. Sometimes the daughter decays into another isotope, which decays further and so on. This successive chain of decays is called a decay series. A good example is shown in Figure 28.9 in which U-238 decays by alpha emission to Th-234, which in turn decays by beta emission and so on until the stable isotope Pb-206 is reached. In an alpha decay ${}^{4}_{2}\text{He}$ is lost so the atomic mass decreases by 4 and the atomic number decreases by 2. This appears as a diagonal arrow. Beta emission (in which n → p) has no effect on atomic mass so it appears as an arrow to the right. Positron emission appears as an arrow to the left.

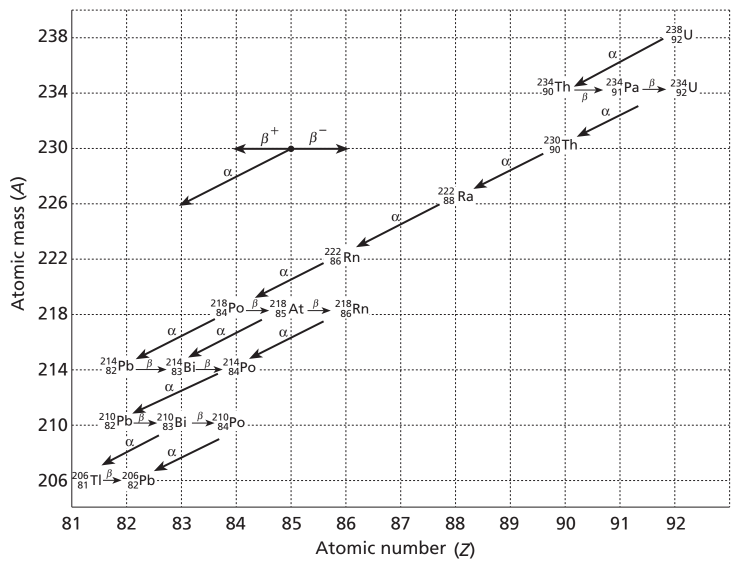


Figure 28.9

In the uranium–lead decay series, note that there are several paths the sequence can take. For example, Po-218 can decay by alpha emission to Pb-204 or by beta emission to At-218 and so on. The series ends at the stable lead isotope Pb-206. Other radioactive series also exist.

28.7 LAWS OF RADIOACTIVE DECAY

Several laws and mathematical relationships have been developed for radioactivity. It was Rutherford in 1919 who first suggested that radioactive decay was an exponential process. His work underpins most of the laws we have today.

Activity

The rate at which a radioactive nuclide decays is called its **activity** (*A*). The SI unit for activity is the **becquerel** (Bq), which represents one disintegration per second (dis s^{-1} or dps). For example, a sample in which 12 atoms decay in one second is said to have an activity of 12 disintegrations per second (dps); that is A = 12 Bq. If you use a Geiger counter to measure activity in the classroom, you would most likely measure the number of disintegrations over a longer period of time, say, 1 minute, and calculate the activity by dividing the 'count' (disintegrations) by the time elapsed in seconds; this would give activity in Bq. You could also express activity as the number of disintegrations per minute (dis min^{-1} or dpm).

The activity law

In an earlier example, the rate of decay of a nitrogen-13 atom was discussed:

$$^{13}_{7}N \rightarrow {}^{12}_{6}C + {}^{0}_{+1}e$$

At the start of the experiment, the sodium was undergoing 12 disintegrations per second (12 Bq); after 10 minutes its activity had dropped to 6 Bq and after a further 10 minutes it had dropped to just 3 Bq (Table 28.5).

Table 28.5 DECAY OF N-13 NUCLIDES

TIME ELAPSED (MINUTES)	NUMBER OF RADIOACTIVE ATOMS REMAINING, *N*	ACTIVITY (Bq)
0	10 000	12
10	5 000	6
20	2 500	3

Inspection of the data shows that the activity is directly proportional to the number of radioactive atoms remaining in the sample:

$$A \propto N$$

or, by replacing the proportional sign with an equal sign and a constant, $A = \lambda N$. The constant, λ, is called the **disintegration constant**. It has the same unit as the unit of the activity. If activity is in Bq (i.e. s^{-1}), then the disintegration constant will also have the same unit (s^{-1}).

Example 1

Calculate the disintegration constant for the above data.

Solution

$$A = \lambda N$$

$$\lambda = \frac{A}{N} = \frac{12\ s^{-1}}{10\ 000} = 1.2 \times 10^{-3}\ s^{-1}$$

Example 2

The disintegration constant of Ra-226 is $4.3 \times 10^{-4}\ y^{-1}$. Calculate the number of atoms in a sample of radium-226 that has an activity of 3 kBq.

Solution

The disintegration constant is given in y^{-1} and has to be converted to s^{-1} because activity is measured in s^{-1}:

$$\lambda = \frac{4.3 \times 10^{-4}}{1\ \text{year}} = \frac{4.3 \times 10^{-4}}{365 \times 24 \times 60 \times 60\ \text{seconds}} = 1.36 \times 10^{-11}\ s^{-1}$$

$$N = \frac{A}{\lambda} = \frac{3 \times 10^{3}}{1.36 \times 10^{-11}} = 2.2 \times 10^{14}\ \text{atoms}$$

Exponential decay law

In maths, you may have learnt the general exponential decay law: $N = N_0\, e^{-kt}$, where e is the base of the natural logarithm, k is the rate constant, and t is time elapsed. In radioactive decay, the formula becomes:

$$N = N_0\, e^{-\lambda t} \text{ or } A = A_0\, e^{-\lambda t}$$

where the Greek letter *lambda* (λ) is the disintegration constant. The relationship between the disintegration constant and half-life is derived in the following manner:

$$N = N_0\, e^{-\lambda t} \quad \text{or} \quad \frac{N}{N_0} = e^{-\lambda t}$$

Take the natural logarithm ($\log_e$ or ln) of both sides:

$$\ln\left(\frac{N}{N_0}\right) = -\lambda t$$

Half-life is defined as the time that one-half of the radioactive atoms in a sample will decay; that is, when:

$$N = \tfrac{1}{2}N_0 \text{ and } t \text{ is replaced by } t_{\frac{1}{2}},$$

hence

$$\ln\left(\frac{\frac{1}{2}N_0}{N_0}\right) = -\lambda t_{\frac{1}{2}} \quad \text{or} \quad \ln \tfrac{1}{2} = -\lambda t_{\frac{1}{2}}$$

$$\ln 2 = \lambda t_{\frac{1}{2}}$$

$$t_{\frac{1}{2}} = \frac{\ln 2}{\lambda}$$

$$t_{\frac{1}{2}} = \frac{0.693}{\lambda}$$

$$\text{or disintegration constant } (\lambda) = \frac{0.693}{t_{\frac{1}{2}}}.$$

Example

The half-life of Ra-226 is 1620 years. Calculate its disintegration constant, λ.

Solution

The disintegration constant for Ra-226 is calculated by:

$$\lambda = \frac{0.693}{1620}$$

$$= 4.3 \times 10^{-4}\ \text{y}^{-1}$$

Questions

10 $^{124}_{55}$Cs has a half-life of 31 s. **(a)** Calculate its decay constant in s^{-1} and min^{-1}; **(b)** if there was 20.0 g of Cs-124 to start with, state how much will be left (in grams) after **(i)** 62 s, **(ii)** 124 s, **(iii)** 10 minutes.

11 **(a)** $^{68}_{32}$Ge has a half-life of 9.0 minutes. How many minutes will it take for the germanium in a 1.00 g sample to decay to 1.00 mg?
(b) If its activity at a particular time is 3.55 kBq, how many minutes will elapse before its activity is 250 Bq?

12 A sample of $^{13}_{7}$N ($t_{\frac{1}{2}}$ = 10.0 minutes) contains 6.90×10^{16} atoms of N-13. **(a)** Calculate its decay constant; **(b)** state its initial activity; **(c)** state its activity after **(i)** 1 hour, **(ii)** 6 hours; **(d)** state after how long will its activity be 1 Bq.

Activity 28.3 'ACTIVITIES' ACTIVITY

1 *Spreadsheet decay* If you have access to a computer and spreadsheet, see if the following challenges you.

- Imagine you had a Tc-99m isotope ($t_{\frac{1}{2}}$ = 6 hours) with an initial activity, A_0, of 1000 Bq. Develop cell formulas that will calculate the activity, A, of the sample every hour for 24 hours. *Hint:* appropriate formulas are: $A = A_0(\frac{1}{2})^n$ or $A = A_0e^{-\lambda t}$.
- Set up a column that will calculate the number of nuclei remaining, N. *Hint:* use $A = \lambda N$.
- Set up a column that will calculate the natural logarithm of activity, ln A.
- Plot A vs t, N vs t and ln A vs t. Are they what you would expect?
- Print out a page of the results, including the graphs.

2 *Random programming* If you are sufficiently familiar with computer programming, write a program that will simulate the decay of a collection of radioactive nuclei. Use whatever language you like (e.g. BASIC, Pascal, C++).

Ch28-Walding 4th 25/8/04 1:37 PM Page 631

- Start with 1000 nuclei (N_0).
- Use a random number generator to determine if any nuclei will decay. *Hint:* generate a random number between 0 and 10; if number >5 = will decay; <5 = will not decay.
- Subtract 1 if 'decay'; determine number of nuclei remaining, N.
- Draw a graph of N vs t.
- Hand in a printout of the code and the output.

ANALYSIS OF EXPERIMENTAL DATA 28.8

The meticulous and often dangerous collection of data on rates of decay helped scientists to uncover the structure of the nucleus and develop an understanding of radioactivity. This goes on today, even in school laboratories.

The exponential decay law is written:

$$N = N_0\, e^{-\lambda t}$$

Because A is proportional to N, we can rewrite this as:

$$A = A_0\, e^{-\lambda t}$$

If we take natural logs of both sides:

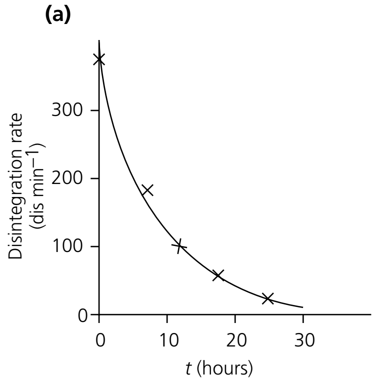
$$\ln A = \ln A_0 - \lambda t$$

This can be arranged to give:

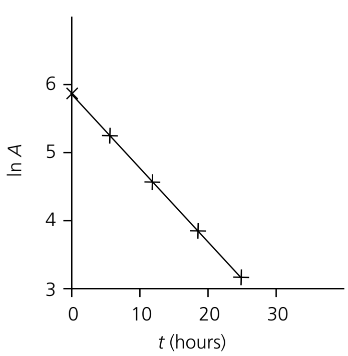
$$\frac{\ln A - \ln A_0}{t} = -\lambda$$

Figure 28.10
A–t time curve is exponential **(a)** whereas ln *A–t* is a straight line.

(a)



(b)



Graphing

As the rate of radioactive decay is exponential, a graph of activity vs time should show an exponential curve (Figure 28.10**(a)**), whereas a graph of the log of activity (ln A) vs time should be a straight line (Figure 28.10**(b)**).

The slope of the ln A vs t graph is equal to $\frac{\ln A - \ln A_0}{t}$, which is equal to $-\lambda$ according to the previous equation.

Example

A sample of radioactive actinium Ac-288 has an initial activity of 363 disintegrations per minute and its activity is measured in a laboratory every six hours. The results are shown in Table 28.6.

Table 28.6 DECAY OF Ac-288 NUCLIDES

TIME ELAPSED (h)	ACTIVITY (DISINTEGRATIONS/MIN)	ln A
0	363.0	5.89
6	184.0	5.21
12	93.2	4.53
18	47.3	3.85
24	24.0	3.18

Plot a graph of ln A vs t and calculate the half-life.

Solution

Refer to Figure 28.10**(b)**.

$$\text{Slope} = \frac{3.18 - 5.89}{24}$$
$$= -0.113\ \text{h}^{-1}$$
$$\lambda = -\text{slope} = +0.113\ \text{h}^{-1}$$
$$t_{\frac{1}{2}} = \frac{0.693}{\lambda} = \frac{0.693}{0.113\ \text{h}^{-1}} = 6.13\ \text{hours}$$

Notes:

- The unit of time for the 'disintegration constant' will automatically be the same as the unit for 'time elapsed'. In this case *hour* was the unit of time.
- The unit of time used for 'activity' does *not* have to be the same as the unit for 'time elapsed'. You may recall that in the exponential decay law formula, 'activity' appeared on both sides of the equation and the time unit would cancel out. *Minutes* were chosen in this experiment to get more accurate results.
- You may ask why all the data collection was necessary when only a starting and final activity are necessary in the formula. In experimental determination of half-life, many data are collected and plotted to improve accuracy. It is just good scientific practice.

Questions

13 Phosphorus-32 is a positron emitter that gives the stable isotope Si-30. The activity of a sample was measured every minute in order to measure its half-life. The results are shown in Table 28.7.

Table 28.7

t (min)	0	1	2	3	4	5	6	7	8	10	12
A (Bq)	5000	3840	2915	2220	1723	1293	977	745	617	361	209

Plot ln *A* against time and determine its half-life.

14 The radioactive isotope Kr-88 undergoes gamma decay. The activity of the isotope was measured every 30 minutes as shown in Table 28.8.

Table 28.8

t (min)	0	30	60	90	120
A (Bq)	88	78	69	61	52

(a) Plot ln *A* vs *t*.
(b) Determine the disintegration constant (s^{-1}).
(c) Determine the half-life.

Radioactive dating

The universe is believed to be somewhere between 8 billion and 15 billion years old. Our Earth has been around for over 4 billion (4×10^9) years of that. Grains of the mineral zircon from gneiss rocks have been measured as being 3.962 billion years old, with a three million year margin of error. The estimate is based on the belief that small amounts of radioactive uranium-238 were trapped in the zircon at the time of crystallisation. Since then the uranium has gradually been decaying, eventually into stable lead. The age calculation is done by measuring the amounts of lead and uranium trapped in the rock and by knowing the half-life of U-238, the age *t* can be calculated.

The age of any object made from previously living material such as wood can be estimated using radiocarbon dating. All living plants take in and excrete carbon dioxide. The vast majority of the C atoms is present as the isotope C-12 but a very small fraction (about 1.3×10^{-10}%) is the radioactive isotope C-14. The ratio has remained constant for thousands of years because even though C-14 decays (with a half-life of 5730 years), more is being made by the cosmic radiation from space bombarding nitrogen in the atmosphere:

$$^{14}_{7}\text{N} + ^{1}_{0}\text{n} \rightarrow ^{14}_{6}\text{C} + ^{1}_{1}\text{H}$$

When plants and animals die they stop exchanging C-14 with the atmosphere so the C-14 decays without being replaced. So the C-14:C-12 ratio drops. After 5730 years, it would be half the normal ratio. If the ratio can be determined, then the time that has elapsed since death can be established.

Questions

15 A piece of wood from a giant redwood tree has a C-14 ratio only one-quarter of that found in living tissue. How old is the wood?

16 For hundreds of years the Shroud of Turin has been claimed to be the burial garment of Jesus. Recently, a sample was analysed and found to have a C-14 count of 92% of that found in living tissue. **(a)** How old is the shroud? **(b)** What is controversial about the answer? **(c)** If it really was from Jesus's time, what would be the amount of C-14 as a percentage of the C-14 in living tissue?

TRANSMUTATIONS BY NUCLEAR REACTION 28.9

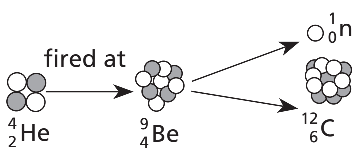
A nuclear reaction happens when there is a change in the structure of the nucleus. Nuclear reactions can be classified as either:

- natural radioactive decay, or
- artificial nuclear reactions.

The radioactive decay processes already discussed are examples of spontaneous nuclear reactions. They occur naturally. No external influence is required.

Other nuclear reactions can be induced artificially by bombarding a nucleus with a projectile, such as a proton, an electron, an alpha particle or other heavy nuclei such as ^{14}N, ^{16}O, etc. (Figure 28.11).

Figure 28.11
Bombardment of a nucleus by alpha particle 'bullets'.



Alpha bombardment

In 1919, after the proton had been discovered, Rutherford designed a series of experiments to probe further into the production of the proton. He bombarded nitrogen gas with 'bullets' of alpha particles:

$$^{4}_{2}\text{He} + ^{14}_{7}\text{N} \rightarrow ^{17}_{8}\text{O} + ^{1}_{1}\text{p}$$

The equation shows that the **transmutation** of nitrogen into oxygen had occurred. When fast, high energy alpha particles are used, neutrons are sometimes formed. For example, if we bombard beryllium nuclei with alpha particles we produce carbon nuclei and fast moving neutrons:

$$^{4}_{2}\text{He} + ^{9}_{4}\text{Be} \rightarrow ^{12}_{6}\text{C} + ^{1}_{0}\text{n} \quad \text{See Figure 28.11.}$$

Deuteron bombardment: turning lead into gold?

Although it is almost impossible to turn lead into gold, **deuterons** ($^{2}_{1}H$) can be used to make gold from mercury or platinum:

$$^{2}_{1}H + ^{199}_{80}Hg \rightarrow ^{197}_{79}Au + ^{4}_{2}He$$
$$^{2}_{1}H + ^{196}_{78}Pt \rightarrow ^{197}_{79}Au + ^{1}_{0}n$$

For both of these reactions, the deuterons must be given high enough speed by a suitable accelerator so that they can enter the nuclei of the target atoms. It costs a lot to use an accelerator, so it is cheaper to go and mine the gold. Bad luck!

Neutron bombardment

The bombarding 'bullets' can also be whole nuclei or single particles, such as neutrons, electrons or protons.

One of the common ways of producing transmutation is by using neutrons. Because they carry no charge, neutrons can enter the nuclei of the target atoms more easily than charged particles can. The target is said to have 'captured' the neutron. By the same token, free neutrons are a serious health hazard because of their penetrating power. The good news is that free neutrons have a short half-life — about 11 minutes before they decay into a proton and an electron.

Some examples of transmutations caused by neutron bombardment and capture:

$$^{10}_{5}B + ^{1}_{0}n \rightarrow ^{7}_{3}Li + ^{4}_{2}He$$
$$^{238}_{92}U + ^{1}_{0}n \rightarrow ^{239}_{92}U$$
$$^{59}_{27}Co + ^{1}_{0}n \rightarrow ^{60}_{27}Co$$

The Co-60 produced in the last reaction decays spontaneously by β and γ emission. Cobalt-60 is used as the gamma source for school radioactive specimens. By completely encasing the radioactive sample in plastic, the β particles are absorbed whereas the γ rays can penetrate through.

28.10 NUCLEAR FISSION AND FUSION

Artificial nuclear reactions can also be classified as either nuclear **fusion** or nuclear **fission**.

In nuclear fusion, the colliding particle unites with the parent and fuses (Latin *fusus* = 'melted', 'unite') into a single nucleus with a higher mass. Sometimes other small particles (n, p) are given off.

In nuclear fission, the lighter colliding particle makes the heavy parent more unstable and it fragments or fissions (Latin *fissus* = 'cleaved', 'split') into smaller nuclei and other particles.

Nuclear fusion

The alpha bombardment of nitrogen carried out by Rutherford in 1919 can be classed as a fusion reaction (Figure 28.12**(a)**):

$$^{4}_{2}He + ^{14}_{7}N \rightarrow ^{17}_{8}O + ^{1}_{1}p$$

Another example of fusion is that occurring in the upper atmosphere and forms the basis of C-14 dating, as described earlier (Figure 28.12**(b)**):

$$^{14}_{7}N + ^{1}_{0}n \rightarrow ^{14}_{6}C + ^{1}_{1}H$$

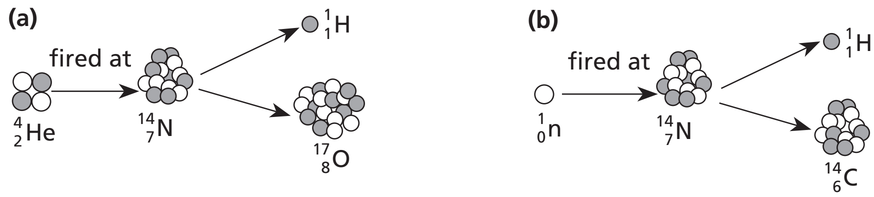


Figure 28.12
Nuclear fusion reactions.

In 1932 J. D. Cockcroft artificially accelerated protons and bombarded a lithium target. The following nuclear fusion reaction took place:

$$^{1}_{1}H + ^{7}_{3}Li \rightarrow ^{8}_{4}Be$$

Nuclear fission

The German scientists Otto Hahn and Fritz Strassman made an amazing discovery in 1938. When U-235 was bombarded with neutrons, sometimes smaller nuclei were produced, which were approximately half the size of the original, for example $^{136}_{56}Ba$ and $^{84}_{36}Kr$. They were baffled by this but Lisa Meitner and Otto Frisch (two Jewish physicists who escaped Nazi Germany in 1938 and were working in Scandanavia) quickly realised what happened. The U-235 nucleus absorbed the neutron to form a U-236 nucleus. Then, like a drop of water, it split into two roughly equal pieces (see Figure 28.13). They called it 'nuclear fission' because it reminded them of biological fission (cell division).

A tremendous amount of energy is released because the mass of U-235 is considerably greater than that of the fission fragments. In early 1940 when Germany was already at war, Hitler banned the sale of uranium from the Czech mines he had overrun. American physicists were alarmed that the Germans might be developing a bomb so the Allies began their own research in the USA (the 'Manhattan Project'), which culminated in the nuclear destruction of the two Japanese cities Hiroshima and Nagasaki, thus ending the Second World War.

To make this fission work, more neutrons must be released than are consumed so as to produce a **chain reaction**. The released neutrons go on to react with other U nuclei and so on. Figure 28.14 shows a four generation chain reaction.

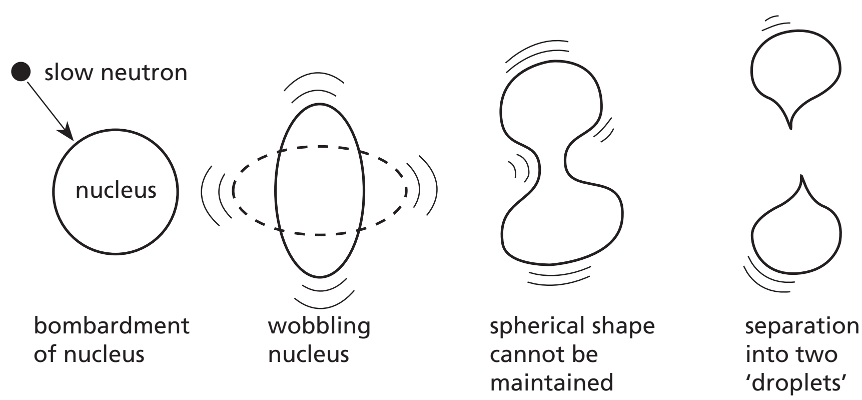


Figure 28.13
Liquid drop model of nuclear fission.

Ch28-Walding 4th 25/8/04 1:37 PM Page 636

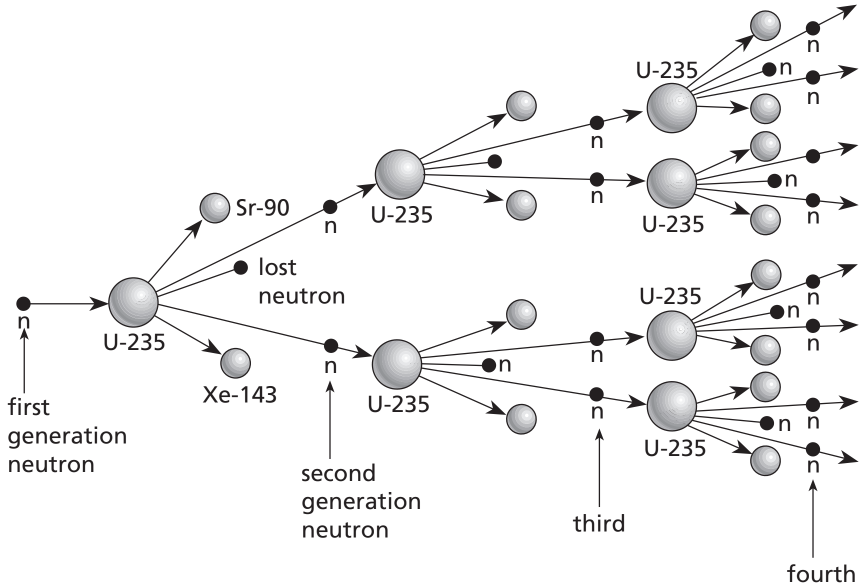


Figure 28.14
A chain reaction.

In each generation the number of fissioning nuclei increases even though some neutrons do not go on to strike another U-235 atom. These are said to be 'lost'. In a nuclear reaction, a chain reaction is kept under control by absorbing excess neutrons with 'control rods' of substances such as cadmium.

PHYSICS FACT

If you held a 10 kg lump of U-238 in your hand it would feel slightly warm. But if you found two separate 10 kg lumps of U-238 and brought them together you'd be blown apart and a crater 50 m deep would form.

Mass and energy in nuclear reactions

In 1905, long before the discovery of nuclear reactions, **Albert Einstein** (1879–1955), a German scientist who moved to the USA in 1933, published his now-famous special theory of relativity. This theory (discussed in Chapter 30) proposed that mass and energy are not separate quantities; rather, they are different forms of one another. The equation relating the two is:

$$E = mc^2$$

where:
m = change in mass or **mass defect** (kg) in a reaction
E = energy released or absorbed (J)
c = speed of light (3×10^8 m s^{-1}).

Example 1

If 1.0 g of a substance is converted completely into energy, how much energy is produced?

Solution

$$\begin{aligned} E &= mc^2 \\ E &= 1.0 \times 10^{-3}\ \text{kg} \times (3 \times 10^8)^2 \\ &= 9 \times 10^{13}\ \text{J} \end{aligned}$$

NOVEL CHALLENGE

Example 1 shows that, when 1 g of a substance is converted completely to energy, 9×10^{13} J is released. By comparison, the chemical energy released when 1 g of the explosive TNT is reacted is 4000 J, and when 1 g of petrol is burnt 30 000 J is released.

Comment on the following: '1 kg of uranium fuel releases 9×10^{16} J'.

Example 2

For the following U-235 fission reaction, calculate the energy released **(a)** per fission; **(b)** per kilogram of U-235 reacted.

$$^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{94}_{38}\text{Sr} + {}^{139}_{54}\text{Xe} + 3\,{}^{1}_{0}\text{n} + \text{energy}$$

NOVEL CHALLENGE

The time it takes to hear a nuclear explosion was worked out by the Rand Corporation in 1968 (in the USA). The formula is: $t = 5.8 \times 10^{-19} R^{10} W^{-3}$ where t is the time in seconds to hear the explosion, W = the equivalent amount of TNT explosive in megatonnes, and R = the distance from the explosion in metres. Calculate how much time would elapse before you could hear a 1000 megatonne bomb at 500 m.

Solution (see Appendix 8 for masses.)

(a)

• Mass of reactants, m_r	U-235	235.043 94 u
	neutron	1.008 665 u
	total m_r	236.052 61 u
• Mass of products, m_p	Sr-94	93.9154 u
	Xe-139	138.9184 u
	3 neutrons	$3 \times 1.008\,665$ u
	total m_p	235.859 8 u

- Mass defect, $\Delta m = |m_p - m_r| = 235.859\,795 \text{ u} - 236.052\,61 \text{ u} = 0.192\,81 \text{ u}$
- 1 u of mass = 1.66×10^{-27} kg
- Mass defect = $0.192\,81 \times 1.66 \times 10^{-27} = 3.2 \times 10^{-28}$ kg

$$\begin{aligned} E &= mc^2 \\ &= 3.2 \times 10^{-28} \times (3 \times 10^8)^2 \\ &= 2.88 \times 10^{-11}\text{ J} \end{aligned}$$

(b) Mass of one U-235 atom = $235 \text{ u} \times 1.66 \times 10^{-27} \text{ kg/u} = 3.90 \times 10^{-25}$ kg.

Energy released by 1 kg = $2.88 \times 10^{-11} \text{ J} \times \dfrac{1 \text{ kg}}{3.90 \times 10^{-25} \text{ kg}} = 7.38 \times 10^{13}$ J.

Questions

17 Calculate the energy released in the following fission reaction in **(a)** J per fission; **(b)** J per kg of U-235: ${}^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{90}_{38}\text{Sr} + {}^{135}_{54}\text{Xe} + 11{}^{1}_{0}\text{n}$ + energy.

18 Which one of the following fusion reactions releases the most energy per fusion event?

(a) ${}^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{0}\text{n}$.
(b) ${}^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + {}^{1}_{0}\text{n}$.
(c) ${}^{2}_{1}\text{H} + {}^{3}_{2}\text{He} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{1}\text{p}$.

In equation **(c)**, be careful to use the mass of the proton (${}^{1}_{1}\text{p}$), not the mass of a hydrogen atom (${}^{1}_{1}\text{H}$).

THE NUCLEUS AS A SOURCE OF POWER 28.11

One issue that has created more debate and social upheaval than any other is nuclear power. People want cheap electricity but they don't want nuclear power stations; they want good export income but they don't want to mine and sell uranium; and they want to be safe from foreign invaders but they don't want the atomic bomb. Common fears are:

- that a reactor might blow up like a uranium bomb
- that a reactor will suffer **meltdown**, the melting of the fuel core because of the heat generated by the fission process
- that radioactive gases will escape into the atmosphere, as did occur at Chernobyl
- that radioactive wastes have to be stored for millions of years
- that terrorists will steal uranium to make a bomb.

These are all real fears. They could all happen but to be able to debate these issues properly, you need to understand how the power of a nucleus can be tapped.

Chain reactions

A chain reaction is one in which the products of the reaction initiate further reactions. There are two types:

- controlled chain reactions — nuclear fission reactors
- uncontrolled chain reactions — the nuclear bomb.

Either way, the underlying nuclear reaction is similar to the one you saw before:

$$ {}^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{90}_{38}\text{Sr} + {}^{143}_{54}\text{Xe} + 3\,{}^{1}_{0}\text{n} + \text{energy} $$

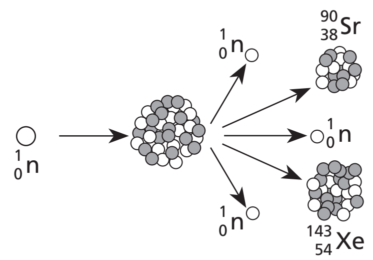


Figure 28.15
The nuclear fission of uranium-235.

Not only are Sr-90 and Xe-143 produced (Figure 28.15), but many different pairs of nuclei are produced, one usually bigger than the other. Examples of pairs are: ${}^{92}_{36}$Kr and ${}^{141}_{56}$Ba; ${}^{85}_{35}$Br and ${}^{148}_{57}$La; ${}^{94}_{38}$Sr and ${}^{139}_{54}$Xe. There are hundreds of different combinations but you will notice that the sum of atomic numbers (the lower numbers) add up to 92 for each pair. But when the mass of the products is less than the mass of the reactants you know that the missing mass has been converted to energy. With the right technology this energy can be harnessed.

> **PHYSICS FACT**
>
> In 1903 Rutherford knew that energy was being slowly released by nuclear decay, and said that if you could speed it up you'd have a bomb. In 1914, writer H. G. Wells wrote the book *The World Set Free* where he described a world 42 years into the future (1956) where a nuclear bomb destroyed cities. He was most perceptive and way ahead of anyone else with his thinking.

28.12 NUCLEAR FISSION REACTORS

The first attempt to harness nuclear energy was in 1942 following a complex program of research involving the coordinated efforts of almost 100 000 scientists and technicians, headed by Enrico Fermi. It was a fission reactor built from 60 tonnes of uranium and about 400 tonnes of graphite blocks in a squash court under the stands of the football stadium at the University of Chicago.

There are basically two types of nuclear fission power reactors — the **thermal reactor** and the more modern but uncommon and far more complex **fast-breeder** reactor. Do we have either in Australia? You bet! Read on.

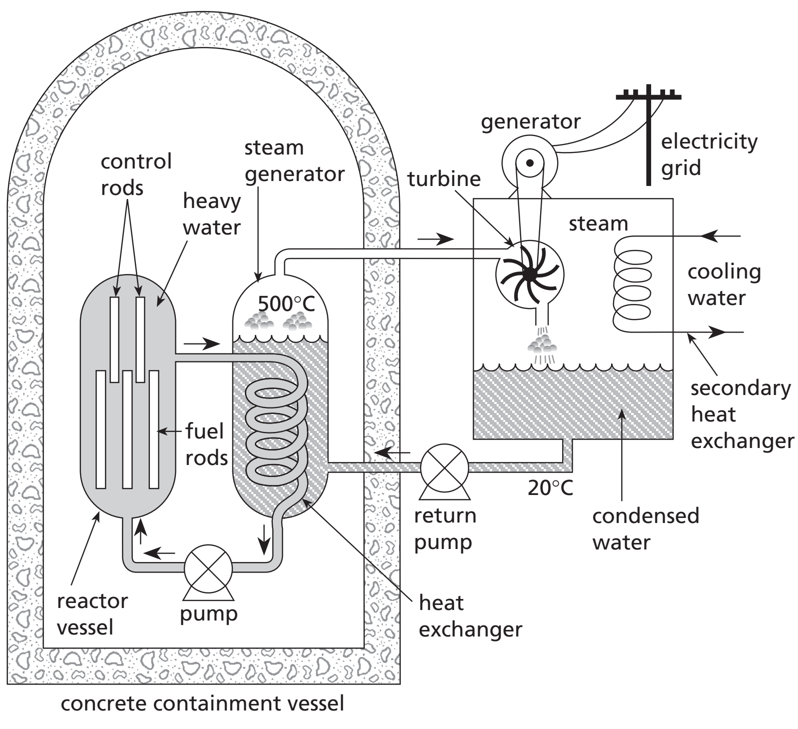


Figure 28.16
Schematic diagram of a typical pressurised water reactor in a thermal nuclear power station.

> **PHYSICS FACT**
>
> In 1950 Ford Motor Co. in the US proposed a nuclear-powered car called the 'Nucleon'. It was never built. Imagine your day trip to the beach. Stop off at the hospital on your way home for a bone marrow transplant.

The Australian Nuclear Science and Technology Organisation (ANSTO) has had a reactor working at its base in Lucas Heights, Sydney, since 1960 and is being replaced. This reactor is called HIFAR — 'High Flux Atomic Reactor'. The reactor is not used to generate electricity; rather, it is used to produce radioisotopes for medicine and industry and for nuclear research. There are no power reactors in Australia but there are hundreds in other countries throughout the world. One of the best sources for further information about the nuclear industry is the Uranium Information Centre's website at the address http://www.uic.com.au where dozens of briefing papers can be found.

Fuel

The fuel has to be a nuclide that undergoes nuclear fission, can sustain a chain reaction and release energy. U-235 is such a fuel. Uranium is mined as a low-grade ore (about 0.3% uranium) and after crushing, chemical treatment and concentration, it appears as uranium oxide (U_2O_3) or 'yellow cake'. This concentrate contains two isotopes, U-238 (99.3%) and U-235 (0.7%). To have a self-sustaining chain reaction, this nuclear fuel usually is **enriched** so that it has about 5% of the U-235 isotope. At Lucas Heights, they enrich theirs to a very high 50% U-235.

The rods are spread to allow the fuel to be cooled and the neutrons to be slowed down to thermal energies for maximum effect in the fission process. But some neutrons will leak away and be lost from the reactor core so the amount of uranium must be sufficiently large to compensate for the loss of neutrons. The minimum mass of uranium needed is called the **critical mass**. It is in the order of a few kilograms.

Moderator

To have a self-sustaining chain reaction, on average at least *one* neutron produced in each fission 'event' must go on to produce another fission. The average number that go on to produce further fissions is called the **multiplication factor**, f. If $f < 1$, the reaction will die out and is called **subcritical**. If, on the other hand, $f > 1$, the reaction is called **supercritical** and an uncontrolled 'run-away' explosion will take place, that is, a nuclear bomb. The aim is to have $f = 1$. Even though 2.4 neutrons are produced on average per fission event in a power reactor, 1.4 of them will be lost, leaving the one neutron to continue the reaction.

The neutrons released in fission of ^{235}U nucleus are very energetic; they have about 2 MeV of energy and are therefore fast-moving. Fast-moving neutrons are unlikely to interact and combine with a nucleus to initiate fission. Slower-moving neutrons are much more likely to be captured by a nucleus. It is important to slow down the fast neutrons from about 20 000 km s^{-1} to 2 km s^{-1}. These slow neutrons are called **thermal neutrons**. This reduction in speed is achieved by placing a **moderator** around the reactor core. To be effective, the moderator must have a mass very similar to the mass of a neutron, so that a neutron can lose maximum energy, in a single collision. The moderator must not absorb neutrons. Carbon (graphite), water, heavy water (D_2O), or liquid sodium are a few suitable moderators. In the Chernobyl accident, the graphite moderator actually caught fire.

Control rods

A reaction is started by placing a neutron source inside the core alongside the fuel rods. After the reaction starts, control is achieved by use of control rods containing neutron-absorbing material such as cadmium or boron steel. These substances have large nuclei so they can easily absorb the neutrons. For example, at Lucas Heights in Sydney, cadmium rods enclosed in a stainless steel coat are used. If the core is not reactive enough, the control rods are raised, which increases the neutron flux, thus accelerating the reaction. If the core becomes too reactive, the control rods are lowered. In 1942, Enrico Fermi had an assistant standing beside the rope that supported the control rods. He was armed with an axe and had orders to chop the rope if the reactor started to run away. The axe was never used, but is now in a museum.

Coolant

The **coolant** is a liquid that circulates through the reactor core to remove excess heat energy and stop it from overheating. Water is a good coolant but heavy water (deuterium oxide, D_2O) is normally used so that it can also act as a moderator.

As the coolant comes in contact with the fuel rods it too becomes radioactive so it cannot be allowed to escape to the atmosphere or down the drain. Instead the hot radioactive coolant is passed through a heat exchanger, where the heat is transferred to a secondary loop. Here, water is converted to steam to drive turbines and produce electricity.

Other forms of reactor use gas (for example, helium) as a coolant. In some cases where a coolant that does not slow down neutrons is required, molten sodium is used to cool the reactor core.

In the worst case scenario of a meltdown accident, the core would get so hot it would melt its way through the floor of the building and into the ground. It was often said that it would 'melt its way through to China', hence the title of the movie *The China Syndrome*. But as soon as the core hit the underground water table it would erupt like a volcano.

Reactor shielding

When in operation, reactor cores emit high levels of radiation harmful to humans. Consequently, the core must be isolated from other areas of the reactor to ensure human safety. This is achieved by use of a shield made from high density materials such as concrete and steel. To prevent any release into the atmosphere the buildings are usually airtight. However, the building at Chernobyl and at others of this design throughout the old Russian republic did not have a containment building. That's why the gases escaped so easily.

Reactor output

The power output of a reactor is measured in megawatts (MW), that is, 1 million joules per second (1 MJ s^{-1}). However, fission reactors have an efficiency of about 32% so they must generate 3000 MW of thermal energy to produce 1000 MW of electrical energy. Research reactors like the type at Lucas Heights only produce about 10 MW.

Fast breeder reactors: FBRs

Fast reactors depend on the fission of ^{239}Pu by fast neutrons. The fuel used is a 10% Pu-239 and 90% U-238 mixture usually extracted as a by-product from a normal thermal fission reactor. Because fast neutrons are used, no moderator is required and the core can be much more compact and operate at a higher temperature, so usually liquid sodium is used as a coolant. Fast neutrons strike U-238, causing a series of beta decays, which result in the production of Pu-239. Thus this reactor produces more of the fissionable Pu-239 material than it started with, hence the name 'breeder'. The problem with FBRs is that the Pu-239 is extremely toxic and has a very long half-life. Secondly, liquid sodium reacts violently with water or air so the engineers must be very careful with the design specifications and control systems.

Uncontrolled fission: the atomic bomb

Fission weapons are relatively uncomplicated. Scientists working on the uranium bomb that was dropped on Hiroshima didn't even test it beforehand. The fission or *atomic* bomb consists of a cannon-like tube with a chemical explosive at one end and two subcritical masses of U-235. To allow safe transportation, the two pieces of fuel must be below critical mass and when detonation is to take place, the two subcritical pieces of almost pure U-235 are fired together. In the bomb dropped on Nagasaki, a ball of subcritical Pu-239 lumps formed the

PHYSICS FACT

Before the two bombs were dropped on Japan at the end of the Second World War, 'Trinity' tests were undertaken.
A plutonium bomb was exploded in the New Mexico desert (USA) on 16 July 1945. A crater 350 m diameter was formed and windows 300 km away were shattered. Three weeks later similar bombs were ready for deployment in Japan. 'Little Boy' was dropped on Hiroshima on 6 August 1945. It was 3 m long, 75 cm diameter, weighed 4.1 tonnes and had an explosive power equivalent to 12 700 tonnes of TNT. It was released at an altitude of 9500 m and at 2100 m the explosive charge detonated, making the uranium go 'critical'. By 580 m, 43 seconds had elapsed and the whole thing exploded in a few millionths of a second. 'Fat Man' was dropped on Nagasaki on 9 August 1945. It was 3.5 m long, 1.5 m diameter and weighed 4.5 tonnes. It was equivalent to 22 350 tonnes of TNT. Japan surrendered the next day.

core. Around the outside of this ball was a layer of chemical explosive, which compressed the plutonium and made it go critical (Figure 28.17).

A nuclear power plant can't explode like a nuclear bomb because the concentration of U-235 is too low. In a reactor, the concentration is only about 3% whereas in a bomb it is enriched to about 90%.

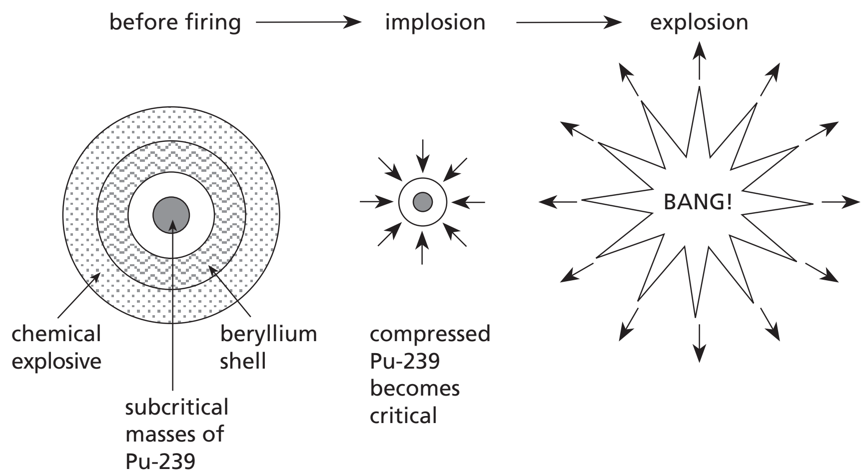


Figure 28.17
An atomic explosion.

Activity 28.4 CRITICAL EVALUATION

SR

1 Enrico Fermi found out in 1939 that uranium could be split into two fragments to release huge amounts of energy. He said it was lucky that nuclear fission had not been discovered five years earlier. What did he mean by this and what possible consequences could have arisen?

2 Students often say that moral questions about the use of physics for weapons should not be discussed in class. Develop some points for and against this comment.

NUCLEAR WASTE PRODUCTS 28.13

The biggest problem for supporters of nuclear power generation is what to do with the wastes and how to convince the public that the wastes can be disposed of safely.

Types of waste

Radioactive waste may be gas, liquid or solid. It is usual to classify it into one of three levels, depending on how radioactive it is:

Low-level waste This waste consists primarily of protective clothing, used wrappers, worn out or damaged plant, and water from showers where protective clothing is washed. It is usually just above the limit that regulations define as radioactive.

Medium-level waste This waste consists of irradiated fuel containers, reactor components and old sealed sources returned from hospitals and industry.

High-level waste This waste comes from fuel reprocessing and contains most of the fission products from the spent fuel and remains quite hot. It contains small amounts of U-235, U-238, Pu-239 and Sr-90.

Disposal

Disposal involves the confinement, isolation and sometimes cooling of the wastes so that they are harmless to the environment. Unlike coal- or oil-fired power plants, which spread their gaseous waste and residues over large areas, nuclear waste is small in volume. The unfortunate part is that the waste is so dangerous.

Unlike many other wastes, nuclear waste will decay in time to become harmless, depending on its half-life. Sr-90 has a half-life of 28 years; Pu-239, 24 400 years. After about 1000 years, however, the wastes have decayed to about the same activity as the original ore body — about 100 gigabecquerel (10^{11} Bq). Some people say 'Why not put it in a rocket and send it to the Sun?' The amount of energy and money required makes this commercially non-viable.

Storage

There are several thousand cubic metres of waste being stored at about 50 different sites throughout Australia. Half is low-level waste. Throughout the world, research continues into safe disposal sites, such as a long way underground, under the sea bed or fusing it with clay to make ceramic beads (e.g. Synroc). At present, ANSTO has accumulated 1600 highly radioactive, highly enriched spent fuel rods at Lucas Heights and is storing them underground in concrete wells until they can convince the United States to take them back for reprocessing. Because only 5% of the original uranium-235 is used up, reprocessing will recover the remaining 95% for reuse. ANSTO has only been able to get rid of some spent fuel rods to Scotland in 1963 and 450 to South Carolina in 1987. Greenpeace accuses ANSTO of not having a decent disposal option for this waste, which is growing at the rate of 36 rods per year and even more when (if?) the new reactor gets built. The idea of burying it is a problem, says Greenpeace, as one cost estimate of a permanent repository is US$17 billion.

NOVEL CHALLENGE

A letter-writer to the *Courier Mail* suggested that we could deflect an approaching asteroid with a nuclear bomb. Propose several reasons why this would not be appropriate.

PHYSICS FACT

An old engineering proverb goes: 'Faster, better, cheaper. Choose two of the above'.
Is this only funny to engineers?

NEI Activity 28.5 NOTHING IS FOOLPROOF FROM FOOLS

The International Atomic Energy Agency (IAEA) argues that nuclear materials, properly handled, are very safe. They suggest that only 17 nuclear accidents have occurred worldwide since 1945 with a total of only 59 deaths (by comparison with about 60 road deaths per week Australia-wide).

Nuclear accidents can be classified as follows:

Radiation accidents (not involving a nuclear reactor) The most publicised have been in Morocco (1984), where eight people died after one person took some radiography isotope home for his mantlepiece; and in Goiania, South America (1986), where people took glowing Cs-137 home in their pockets from a radiotherapy unit — 4 dead, 249 contaminated.

Nuclear power reactor accidents Examples are at Three Mile Island, USA (1979) — partial melting of the core; and at Chernobyl (1986) — 100 000 people evacuated after the graphite moderator caught fire.

Either

1. Write a two-page report about a nuclear accident (including one of the above, if you like).
   (a) Describe what went wrong from a nuclear physics viewpoint.
   (b) Describe the long-term environmental effects.
   (c) Discuss whether the benefits of nuclear technology outweigh the environmental effects.

or

2. Design a public survey to discover what people in your local community know about the issues involved in nuclear energy. Conduct a small survey and analyse the results. Submit a full report: introduction, methodology, results, analysis, conclusion.

NOVEL CHALLENGE

Poet W. H. Auden wrote in his poem *Marginalia*:

No tyrant ever fears
His geologists or his engineers.

What do you suspect he meant by this? What evidence would he need to substantiate this claim?

NUCLEAR FUSION 28.14

Nuclear fusion is the process in which two small nuclei join together to produce a larger nucleus and a release of energy.

The dream of producing a sustainable and controlled nuclear fusion on Earth is one that scientists have had for fifty years. Nuclear fusion powers the Sun and it powers the hydrogen bomb but trying to get it working in a controlled fashion here on Earth is a difficult task. For two small nuclei to fuse, we have to overcome their massive electrostatic repulsions. It normally takes more energy to do this than could ever be released, so very careful selection of nuclei is needed.

The first release of energy from a fusion reaction was in 1932, when Cockcroft and Watson demonstrated the following reaction:

$$^{2}_{1}\text{H} + ^{3}_{1}\text{H} \rightarrow ^{4}_{2}\text{He} + ^{1}_{0}\text{n} + 17.6 \text{ MeV}$$

The energy comes from the high binding energy per nucleon of the very stable helium nucleus compared with the smaller binding energy per nucleon of deuterium $^{2}_{1}\text{H}$ and tritium $^{3}_{1}\text{H}$. Other fusion reactions that form helium nuclei include:

- $^{2}_{1}\text{H} + ^{2}_{1}\text{H} \rightarrow ^{3}_{2}\text{He} + ^{1}_{0}\text{n} + 3.3 \text{ MeV}$
- $^{2}_{1}\text{H} + ^{3}_{2}\text{He} \rightarrow ^{4}_{2}\text{He} + ^{1}_{1}\text{p} + 18.3 \text{ MeV}$

Since some mass has had to be converted to energy, this will be accompanied by an equivalent mass decrease. The energy is normally given off in the form of gamma rays.

High temperature fusion

The British physicist Sir Arthur Eddington (1882–1944) first suggested in 1920 that the Sun might produce all of its energy by nuclear fusion. In 1930 Dr Hans Bethe, an American scientist, gave a clear indication of how this process could work.

At the core of the Sun the temperature is about 100 million °C. The temperature and pressure are high enough to enable hydrogen atoms to fuse together to form helium atoms. As the gamma rays that are produced move outward through the Sun they heat the surrounding gas. At the Sun's surface the temperature is about 6000°C. Although the reaction occurs in a number of steps, the process can be represented as:

$$4\,^{1}_{1}\text{H} \rightarrow ^{4}_{2}\text{He} + 2\,^{0}_{+1}\text{e} + \gamma + \text{neutrino } (\nu)$$

The energy produced is 26.72 MeV when the four hydrogen atoms react. About 26.2 MeV is carried off by the gamma rays and about 0.5 MeV by the neutrinos. Every second, 4 million tonnes of hydrogen is converted to helium in the Sun, providing the energy that makes life on Earth possible.

NOVEL CHALLENGE

Two new heavy nuclides have been prepared recently. Work out what they are from these incomplete equations: $^{249}_{98}\text{Cf} + ^{12}_{6}\text{C} \rightarrow 4\,^{1}_{0}\text{n} + ?$; and $^{242}_{94}\text{Pu} + ^{22}_{10}\text{Ne} \rightarrow 4\,^{1}_{0}\text{n} + ?$

Fusion reactors

High temperatures and pressures like those found inside the Sun are needed to sustain nuclear fusion. There are two major problems associated with this:

- How do we reach this high temperature and maintain it?
- How do we contain the fuel, since most containers would melt well before 100 million degrees C is reached?

For the construction of fusion power reactors the reactions mentioned earlier involving deuterium and tritium seem to be promising. There is a vast supply of deuterium available in ordinary water, particularly in sea water. Therefore, there is no scarcity of fuel for fusion reactors.

Ch28-Walding 4th 25/8/04 1:37 PM Page 644

In order for the particles to come close enough to fuse together, the deuterium or tritium atoms must have very high kinetic energy (high speed), sufficient to overcome the Coulomb repulsion. Such high speed can be achieved by heating the gas to a temperature of 100 million °C. At such high temperatures, gas contains electrons and positive ions and is called ‘plasma’ (Figure 28.18).

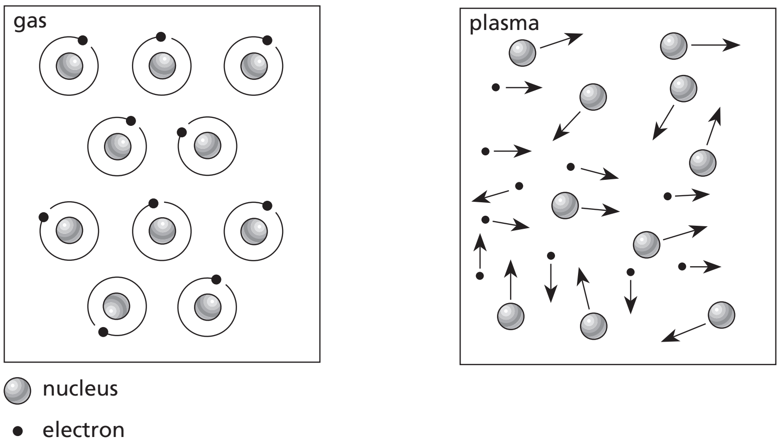


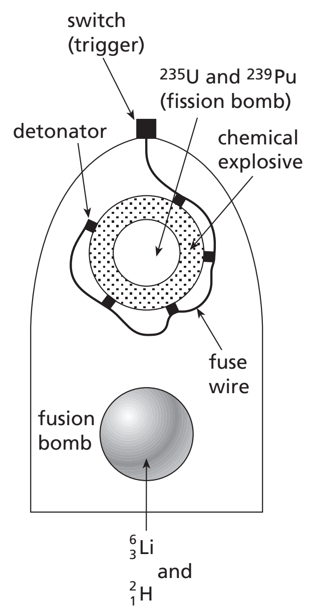
Figure 28.18
In gas, the electrons orbit the nuclei. In plasma, the electrons are separated from the nuclei.

Uncontrolled fusion

The energy of fusion has been harnessed for destructive purposes in the form of the **hydrogen bomb** (Figure 28.19). This type of bomb is technically more advanced than the fission bomb and potentially more devastating.

To obtain the high temperatures and pressures required for fusion, the H-bomb begins with a fission explosion. The fission neutrons combine with Li-6 to form tritium, and when the temperature is hot enough, the tritium atoms fuse with the deuterium atoms in an uncontrolled fusion reaction. The H-bomb has never been detonated in anger. The first successful detonation was in 1952 in the United States when it was said that the explosion was ‘brighter than a thousand suns’. (See Photo 28.2.)

Figure 28.19
The H-bomb is ignited by a fission bomb.



Activity 28.6 NUCLEAR UNCLEAR

NEI

1 Prepare a table showing the similarities and differences of a thermal fission reactor, a research reactor (e.g. ANSTO’s HIFAR), an FBR, a fusion reactor and a coal-fired power station. Some headings you could use are: fuel, moderator, control rods, power output, wastes, but there are others.

2 Use an encyclopedia to find the meaning of: yellowcake, Synroc, Candu reactor, trefoil symbol, tokamak.

Questions

19 Greenpeace has said that having ships take spent fuel rods back to the United States through the waters of the South Pacific would be like ‘floating Chernobyls’. What do they mean?

20 Australia mines uranium for sale overseas but won’t allow nuclear power reactors to be built here. It seems hypocritical — but is it? The government says that by being a part of the nuclear community we have control over what happens to the uranium. Prepare a short statement with your argument for or against this policy.

Photo 28.2
The first fusion bomb (known as ‘Mike’) was exploded on 31 October 1952. During Operation Ivy, Mike yielded 10.4 megatonnes (TNT equivalent) and was reported to be ‘brighter than a thousand suns’.

BIOLOGICAL EFFECTS OF RADIATION 28.15

Nuclear radiation is everywhere. It has been a part of the natural environment since the Earth was formed some 4 billion years ago. We cannot feel it but in high doses its effects can be devastating. When ionising radiation (α, β, γ, X-rays and neutrons) passes through tissue, it can interfere with the DNA, causing it to split up. When the DNA replicates (makes copies of itself) it can get 'confused' and produce defective cells. All people who deal with ionising radiation should be aware of the consequences.

However, not all radiation effects are the same. Different types of radiation have different ionising and penetrating powers. Hence, the effects on human tissue will also be different. You could ask: would you rather be hit by a truck moving at 10 km/h, a motorbike going at 100 km/h or a bullet going at 1000 km/h? All three do damage but in different ways. Alpha radiation is made up of heavy, relatively slow moving particles but is more intensely ionising over a very short distance than beta, gamma or X-rays. Even though it cannot penetrate the outer layers of the skin, it can be inhaled as a dust or get into open cuts in the skin. Then it is very dangerous as it can move around in the blood and migrate to the brain and lungs — a deadly result.

Beta particles have a lower risk as they can be stopped by a few millimetres of tissue and are not very strongly ionising. Nevertheless they can cause skin burns. Because gamma and X-rays are highly penetrating, they can pass deep into the body and cause great damage. For example, when you have an X-ray, the radiation passes right through you to the photographic film behind. Gamma is like the bullet.

In summary, radiation may result in:

- the death of the cell
- prevention of cell division
- permanent modification to the cell.

Questions

21 Find out how cobalt-60 actually kills cells. Why doesn't irradiation of the patient make the patient radioactive?

22 Prepare a case for or against the use of radioactive sources in school physics laboratories.

MEASURING RADIATION — DOSIMETRY 28.16

Because there are so many different factors affecting the radiation dose, there are several different ways of measuring it.

Absorbed dose (D)

When ionising radiation interacts with matter, some of its energy is transferred to the absorbing material, such as the tissues of the body. Radiation that deposits one joule of energy per kilogram of tissue is called the **absorbed dose**. It has the units J kg^{-1} or 1 **gray** (Gy).

For example, if a 50 kg person absorbed 100 J of radiation energy, this absorbed dose would be 100 J/50 kg = 2 J kg^{-1} or 2 Gy. This could make you very sick but would probably not be fatal. If a 20 kg child absorbed the same energy, the dose would be 100 J/20 kg = 5 Gy, which would probably be fatal.

Dose equivalent (H)

The amount of damage 1 Gy of absorbed dose can do depends on the nature of the radiation. For example, alpha particles are 20 times more damaging to tissue than X-rays, gamma rays or beta particles. To quantify the potential damage of radiation, physicists use weighting to reflect the biological impact. These weightings are called **quality factors** (QF) and are shown in Table 28.9.

Table 28.9

RADIATION	QUALITY FACTOR (QF)
heavy nuclei	20
fusion fragments	20
alpha	20
neutrons <10 keV (slow)	5
10 keV – 100 keV	10
100 keV – 2 MeV (fast)	20
2 MeV – 20 MeV	10
>20 MeV	5
protons	5
beta	1
gamma photons	1
X-ray photons	1

A measure of the radiaion dose that combines the amount of radiation (in J/kg) with the quality factor (QF) is called the 'dose equivalent' and is measured in a unit called **sievert** (Sv).

$$\text{Dose equivalent} = \text{absorbed dose} \times \text{quality factor}$$
$$\text{or } H = D \times QF$$

For example, if an absorbed dose of 2 Gy was from an alpha source, the dose equivalent would be 2 Gy times a QF of 20 = 40 Sv. This is a big dose and would probably be fatal (>6.5 Sv is lethal). It is common to use millisievert (mSv) or microsievert (μSv). In Australia, the average annual background radiation dose is about 2 mSv.

28.17 RADIATION RISKS TO YOUR HEALTH

Radiation is all around us. It comes from the ground, from space (cosmic radiation) and can be from artificial sources such as medical X-rays and nuclear bomb testing such as the French did in the Pacific Ocean. Non!

The effective 'whole body' *dose limits* established by the International Commission for Radiological Protection for artificial sources are:

- radiation workers: 20 mSv (100 mSv averaged over 5 years and maximum of 50 mSv in any one year)
- members of the public: 1 mSv annual average over lifetime; maximum of 5 mSv in any one year; pregnant woman (abdomen) 13 mSv/3 months; foetus 1 mSv/9 months.

The effects of ionising radiation on the body are summarised in Table 28.10.

INVESTIGATING

No new nuclear reactors have been built in the USA since 1972 and the number operating has dropped from 146 to 143 over recent years.

How does this compare with the UK, France and Indonesia?

Table 28.10 EFFECTS OF IONISING RADIATION ON THE BODY

EXPOSURE (Sv)	EFFECT
High (3–6 Sv)	severe sickness with up to 100% deaths for 4.5 Sv and over
Medium (1–3 Sv)	slight to moderate sickness; recovery complete within 3 months; delayed effects may shorten life by a few percent
Low (0–1 Sv)	No sickness; person unaware of any biological changes

SR

Activity 28.7 WHAT'S YOUR POISON?

You can work out your total annual radiation dose by summing the various sources. What does your annual dose come to? Is this dangerous? See Table 28.11.

Table 28.11

SOURCE	AVERAGE ANNUAL DOSE (μSv)	LOCAL VARIATIONS
Cosmic radiation	300	+200 μSv per 100 m of altitude
		+ 20 μSv per 10° of latitude
		+ 20 μSv per hour of flying time
		+1000 μSv per day in a space station
Rocks	600	−150 μSv if you live in a wooden house
Food and drink	300	Nil
Air (breathing)	700	Nil
Manufactured source	100	+ 30 μSv for nuclear testing
		+ 20 μSv if you watch TV and use the computer for about 15 hrs/week
Medical	0	See medical treatment, Table 28.12
Your total	2000	

Table 28.12 MEDICAL TREATMENT

PROCEDURE		DOSE (μSv)
X-rays:	Chest	30
	Leg	20
	Dental	140
	Head	70
	Intestine	3 000
	Mammography	400
CAT scan:	Head	1 800
	Abdomen	7 200
Radiopharmaceutical drugs:		
	Bone scan	5 000
	Thyroid scan	2 000
	Lung scan	800
	Brain scan	1 200
	Heart scan	17 000
	Tumour therapy	22 000
	Thyroid therapy	8 000 000
Your total dose due to medical treatment:		

1 Calculate the annual dose for an airline pilot who lives in a wooden house in Brisbane (latitude 28°) at an altitude of 100 m and makes 200 return flights to Sydney in a year. He had a chest X-ray after swallowing a fish hook and a mouth X-ray for an impacted wisdom tooth.

2 Why is altitude a factor in determining dose? Is this related to airline flights being listed as a hazard?

For further information see Chapter 33, 'Medical Physics'.

INVESTIGATING

The international symbol for nuclear disarmament made up of the semaphore (flag) signals for N and D is shown in the following figure.

What part of the symbol is N and which part is D?

PHYSICS FACT

Radon gas (from bricks and other earthy materials) causes between 7000 and 30 000 of the 130 000 lung cancer deaths in the USA per year (Environmental Protection Agency estimate, 2001).

NOVEL CHALLENGE

A person receives a dose of 0.3 mGy of slow neutrons, 6 mGy of gamma rays and 0.1 mGy of fast neutrons. How many millisieverts is this?

28.18 APPLICATIONS OF NUCLEAR TECHNOLOGY

Uses of nuclear technology include more than just nuclear power and radioactive dating. Here are some other uses:

Food and medical equipment irradiation Within a year of Roentgen's discovery of X-rays in 1895, physicists were proposing that food be irradiated to kill off microbes. By 1921 patents had been given for irradiation of meat to destroy parasites. Later, in 1931, it was being used to kill off bacteria and preserve food indefinitely.

Gamma radiation and high energy X-rays can be used to prolong the shelf-life of foods by slowing the ripening process and stopping the sprouting of vegetables like potatoes and onions. Although the food does not become radioactive, irradiation can cause physical and chemical changes in the food. When molecules are split by nuclear radiation, the fragments are called **radiolytic** products (Greek *lysis* = 'to split'). Although 50 years of research has shown irradiation to be useful and generally safe, there is concern that these radiolytic substances could be carcinogenic (cancer-forming).

To date, in Australia and New Zealand, only herbs and spices and some tropical fruits have been approved to be irradiated (for sanitary reasons) and to a maximum dose of 1 kGy. The tropical fruits include breadfruit, carambola (star fruit), custard apple, lychee, longan (dragon's eye lychee), mango, mangosteen, papaya and rambutan. Food that has been irradiated must be labelled with a statement that says it has been treated with ionising radiation.

Medical equipment has been sterilised with gamma rays for years. If you look on a syringe packet it will probably say that it has been gamma irradiated and sealed with the gas ethylene oxide inside.

PHYSICS FACT

ANSTO scientists measured the U/Pb ratio in Gold Coast beach sand. They found it was Precambrian sand from the Antarctic 600 million years old.

INVESTIGATING

In the nuclear power plant accident at Tokaimura, one worker — Mr Hisashi Ouchi — received 18 Sv of radiation and spent 3 months in hospital. Normally, anything greater than 5 Sv is fatal.
So how come he survived?

Questions

23 **(a)** List the arguments for and against food irradiation.
(b) Do you think Australians should be allowed to eat irradiated foods? Give reasons.
(c) How can we easily check that irradiated foods do not have any residual radioactivity?
(d) List the arguments for and against the irradiation of medical products.
(e) Do you think medical supplies should be irradiated by the manufacturers before they are placed on the market? Won't they become radioactive?
(f) How is this different from the food irradiation argument?

Industrial radiography Gamma rays can be used to examine the interior of solid objects such as the welds in natural gas pipelines (Figure 28.20).

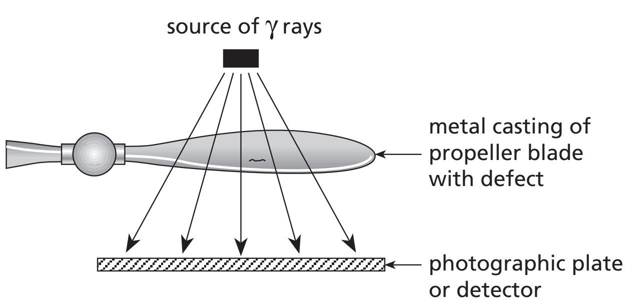
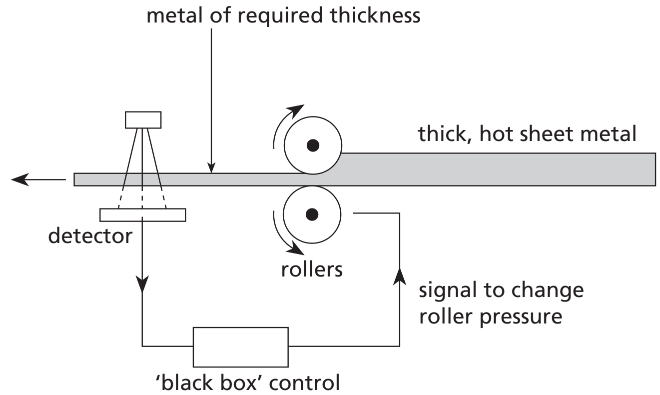


Figure 28.20 Detection of flaws inside metal parts.

Neutron radiography Slow neutrons are fired at an object and, depending on the presence of elements like hydrogen, cadmium and boron, a clear image of the internal structure can be obtained in much the same way that X-ray images are taken. These atoms are strong absorbers of slow neutrons and are remarkable in the detail they can show. Common uses are to detect flaws in gas turbine blades, corrosion of aircraft components and the presence of explosives in luggage.

Gauging Radiation has its intensity reduced by matter placed between the source and a detector. By measuring the radiation through a plastic film, its thickness can be monitored as it comes out of the rollers and adjusted to maintain a thickness of incredible uniformity (Figure 28.21). In Australia, the thickness of paper, felt, steel and glass are areas where radioactive gauging is also used.

Figure 28.21
Use of radiation to monitor the thickness of sheet metal.



PHYSICS FACT

British nuclear physicist James Chadwick was a slum boy from Manchester. He did his postdoctoral research under Rutherford and later went to Berlin. When the Second World War broke out, he was taken prisoner by the Germans. While in the prisoner of war camp, he ordered special toothpaste from the Berlin Auer Company which had radioactive thorium in it to make teeth glow. He extracted the thorium and used it to continue his nuclear experiments while in the POW camp. In 1932 Chadwick discovered the neutron. That's what you call persistence.

Neutron activation analysis (NAA) This is used in forensic science to match soil and hair from crime scenes to suspects. Australia has become a world leader in this field.
Neutron transmutation doped silicon (NTDS) This is a process used to produce the 'doped' silicon for silicon chips in the computer industry. Bombardment of silicon with neutrons produces phosphorus:

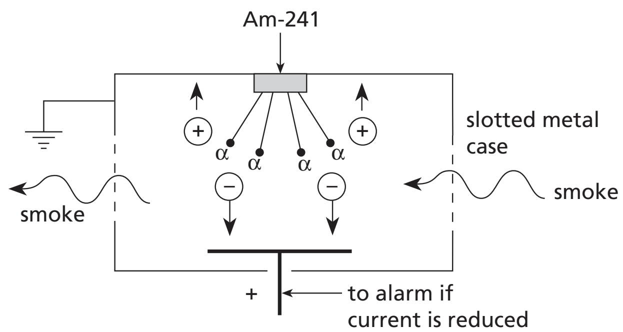
$$^{28}_{14}\text{Si} + 3\,^{1}_{0}\text{n} \rightarrow \,^{31}_{15}\text{P} + \,^{0}_{-1}\text{e}$$

By controlling the intensity of the radiation, a transmutation of 1 in every 10^8 atoms can be achieved — just the right amount for N-type semiconductors. This is a big income earner for ANSTO and for Australia.
Medical applications These are among the most important uses of radiation and radioactivity. They are discussed in Chapter 33.

EI Activity 28.8 SMOKE DETECTORS

Smoke alarms in homes operate on the principle of ionisation. A radioactive source of americium-241, with a half-life of over 7000 years, emits alpha particles that ionise air particles in a chamber. When smoke enters the chamber, the ionisation changes and an alarm sounds.

Figure 28.22
A smoke detector.



1 Examine a smoke detector and, by unscrewing the cover, look at the components. Can you identify the ionisation chamber? Don't open it! Blow some smoke from a candle into it — did the alarm sound? How long did it take?

2 Hold a Geiger counter near it. Can you detect radiation? If not, why?

3 Is there any warning about the Am-241? Is it clear to a non-physics student that it is dangerous? Are there sufficient instructions about how to dispose of it once it stops working?

4 If you replaced the battery regularly, would the detector be good for thousands of years? Explain.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***24** What element is represented by X in each of the following: **(a)** $^{233}_{92}X$; **(b)** $^{2}_{1}X$; **(c)** $^{226}_{88}X$; **(d)** $^{32}_{15}X$?

***25** What elements are formed by the radioactive decay as shown in each of the following?
(a) $^{24}_{11}Na(\beta^-)$. **(b)** $^{22}_{11}Na(\beta^+)$. **(c)** $^{210}_{84}Po(\alpha)$. **(d)** $^{32}_{15}P(\beta^-)$.

***26** Complete the following nuclear reactions:

(a) $^{234}_{91}Pa \rightarrow ^{234}_{92}U + ?$

(b) $^{?}_{?}Rn \rightarrow ^{218}_{84}Po + \alpha$.

***27** A certain nuclide, represented by $^{b}_{a}X$, ejects an α-particle followed by an emission of a β-particle. Use Y and Z as daughter symbols and write two nuclear equations to represent the process.

***28** Complete the following nuclear equations:
(a) $^{11}_{5}B + ^{4}_{2}He \rightarrow ^{14}_{7}N + ?$
(b) $^{23}_{11}Na + ^{1}_{0}n \rightarrow ?$
(c) $^{27}_{13}Al + ^{2}_{1}H \rightarrow ^{4}_{2}He + ?$
(d) $? + ^{4}_{2}He \rightarrow ^{42}_{20}Ca + ^{1}_{1}H$
(e) $? + ^{1}_{0}n \rightarrow ^{27}_{12}Mg + ^{1}_{1}H$
(f) $? + ^{4}_{2}He \rightarrow ^{12}_{6}C + ^{1}_{0}n$

***29** The fission fragment Sr-96 undergoes four successive β^- emissions before a stable nucleus is formed. What is the stable nucleus formed?

***30** For the isotope carbon-14:
(a) write the equation for its beta-minus decay;
(b) determine the mass loss per atom of carbon-14 in kg;
(c) determine the energy change for this process.

****31** Indium-116 decays by beta-minus decay to tin-116. The atomic masses are 115.905 53 u for indium-116 and 115.901 79 u for tin-116. **(a)** Write the equation for the beta-minus decay of indium-116. **(b)** Determine the mass loss in kg per atom of indium-116. **(c)** Determine the energy change for this process.

****32** Can C-14 dating be used to measure the age of stone walls and tablets of ancient civilisations?

****33** Carbon-14 was used to date a medieval linen sample. Calculate how old it was if it had a C-14 : C-12 ratio that was only 1.56% of the expected living tissue ratio.

****34** Some rocks in your neighbourhood show that their percentage of uranium-236 is 67% of what you expected. If $t_{\frac{1}{2}}$ for U-236 is 2.39×10^7 years, calculate the age of the rocks.

****35** The half-life of a cobalt-60 source used for food irradiation is 5.26 years.
(a) If the original activity of a working sample is 500 GBq, what is the activity after 2.63 years?
(b) If it is not safe to dispose of it until its activity is less than one-thousandth of its original activity, calculate how many half-lives this is.

36 Iridium-192 is used in the treatment of early cancer of the breast. It has a half-life of 74 days. If initially there is 3.6 mg of the isotope present, what time will elapse before it has been reduced to **(a) 0.90 mg; **(b)** 0.25 mg?

37 Iodine-131, used for destroying malignant tumours of the thyroid, has a half-life of 8.07 days. **(a) What is its disintegration constant? **(b)** If the activity of I-131 is 5×10^{10} Bq, how many iodine atoms are present? **(c)** How many days will elapse before its activity is 1 MBq?

**38 Immediately after a $^{238}_{92}U$ nucleus decays to $^{234}_{90}Th + ^{4}_{2}He$, the daughter thorium nucleus still has 92 electrons circling it. Since thorium normally holds only 90 electrons, what do you suppose happens to the two extra ones?

**39 Design an experiment to measure how much liquid is in a can without opening it. *Hint:* alpha radiation would not be suitable for this experiment. Why not?

**40 The activity of a sample of a beta-emitting phosphorus nuclide was measured and the count was corrected for background radiation. The results were as shown in Table 28.13.

Table 28.13 A RADIOACTIVITY COUNT

TIME ELAPSED (h)	ACTIVITY (min^{-1})
0.0	36 506
0.5	31 501
0.75	29 268
1.0	27 106
2.0	20 244
5.0	8 256
10.0	1 913
13.0	800
18.0	181

Plot ln *A* vs *t* for these data and determine the half-life of the phosphorus.

**41 One of the long-term effects from a nuclear explosion is the radioactive fallout. If all 3.0×10^6 radioactive fission nuclei were spread evenly among the human population (5 billion), how many radioactive nuclei would each human breathe in? Why is this not a realistic calculation?

**42 The following is a problem regularly faced by scientists and engineers in the nuclear industry. You have to dispose of some low-level nuclear waste and you must choose from three types of storage containers (Table 28.14).

Table 28.14 NUCLEAR WASTE MANAGEMENT

CONTAINER	COST OF ONE BOX ($)	VOLUME OF ONE BOX (m^3)	LIFE OF ONE BOX (y)
Steel and lead box	500	2	10
Steel and plastic box	1000	2	100
Carbon fibre box	2400	2	1000

You have to dispose of 30 m^3 of waste by burying it in boxes until its activity reaches the low level of 600 Bq/kg of waste. Initially the 'hot' waste has an activity of 10 500 Bq/kg and has a half-life of 12 years.

(a) How many years have to pass before the activity falls to 600 Bq/kg?
(b) How many containers do you need?
(c) Which container would you choose? Why?
(d) What is the total cost of your containers?
(e) Would the operation have any other costs?
(f) If the containers leaked, who would you blame?

Extension — complex, challenging and novel

****43** For the reaction: $4{}^{1}_{1}H \rightarrow {}^{4}_{2}He + 2{}^{0}_{+1}e$:

(a) calculate the energy released when four ${}^{1}_{1}H$ nuclei fuse;

(b) calculate the energy released when 1 kg of ${}^{1}_{1}H$ fuses.

****44** **(a)** Which one of the following produces the most energy per kilogram of reactant?

(i) ${}^{239}_{94}Pu + {}^{1}_{0}n \rightarrow {}^{137}_{52}Te + {}^{100}_{42}Mo + 3{}^{1}_{0}n$ (See Appendix 7 for masses.)

(ii) $2{}^{2}_{1}H \rightarrow {}^{4}_{2}He$

(b) State whether the above reactions are fusion or fission.

*****45** In the carbon cycle that occurs on the Sun, He-4 is built from four protons and C-12. First, C-12 absorbs a proton to form a nucleus, X_1. Then X_1 decays by positron emission to X_2, which then absorbs a proton to become X_3, which itself absorbs a proton to become X_4. X_4 then decays to X_5 by positron decay and X_5 reacts via: $X_5(p,\alpha)X_6$.

(a) Determine the formulas of X_1 to X_6 by writing out complete balanced nuclear equations.

(b) Sum the six reactions and write a balanced net reaction.

*****46** The Sun radiates 3.9×10^{23} J of energy into space every second.

(a) Calculate how much mass is lost per second on the Sun.

(b) If the Sun has a mass of 2×10^{30} kg, calculate how many years will elapse before the Sun has lost 50% of its mass.

*****47** Nuclear fission of U-235 releases about 3.5×10^{-11} J per fission event. Calculate this as J per kg of U-235 reacted and calculate how many times greater it is than the combustion of methane, which releases 50 MJ per kg.

*****48** The Earth receives 1.8×10^{14} kJ per second of solar energy. **(a)** What mass of solar material is converted to energy over a 24 hour period to provide the daily amount of solar energy to the Earth? **(b)** If coal releases 32 kJ of energy per gram, what mass of coal would have to be burnt to provide this same amount of energy?

*****49** An unstable isotope disintegrates by beta decay to form a stable product. In an experiment to determine its half-life, the following data were collected at the same time each school day (Table 28.15). *Note:* better results are obtained by measuring the time taken for a certain count. In this case 2000 counts were made. Background radiation = 13 counts per minute.

Table 28.15 DECAY DATA FOR QUESTION 49

DAY	TIME FOR 2000 COUNTS
Mon	1 min 16 s
Tue	1 min 39 s
Wed	2 min 10 s
Thurs	2 min 51 s
Fri	3 min 40 s
Mon	8 min 28 s
Tue	10 min 15 s
Wed	13 min 48 s
Thurs	17 min 33 s
Fri	22 min 59 s

(a) Calculate the observed activity and then the actual activities by subtracting background activity.

(b) Calculate the half-life by a graphical method. *Hint:* watch the 'time elapsed'; there are some days when no data were taken (who'd go to school at the weekend?).

*****50** Radioactive gold-198, which has a half-life of 2.7 days, is routinely used by ANSTO for detecting movement of substances through the environment. It was once suspected that the high concentration of aluminium in the brains of Alzheimer's patients was from using aluminium saucepans for cooking. Devise a procedure using Au-198 to measure the movement of saucepan metal into the brains of Alzheimer's patients.

*****51** A 70 kg patient receives a 10 second dose of gamma radiation from a 3.7×10^{13} Bq cobalt-60 source. The gamma rays have an energy of 1.25 MeV (1 MeV = 1.6×10^{-13} J). Only 2% of the radiation reaches the patient and only 50% of that is absorbed; the rest just passes through. Show that the effective dose is 0.011 Gy.

*****52** In a fast breeder reactor, U-238 is bombarded with a nuclear particle and turns into U-239, which promptly decays into Pu-239 by several steps. Deduce the full set of reactions for the FBR.

Ch29-Walding 4th 25/8/04 1:40 PM Page 654

CHAPTER 29

Quantum Physics and Fundamental Particles

29.1 INTRODUCTION

Until the end of the nineteenth century, classical physics, based on the laws of Newton, had sufficed to explain all our natural surroundings of matter, space and time. It was at the turn of the century that the experimental observations by physicists and subsequent theoretical explorations began to question the validity of the Newtonian laws, especially at very small distances, very high speeds and within the world of the emerging atom. For example, lines had been noticed in the spectra of light emitted by heated gases or gas discharges. The Rutherford atomic model would not have predicted these lines. Light itself was difficult to explain as it seemed to have both a particle nature and a wave nature, and the field of thermodynamics did not seem to be related to molecules and atoms at all.

The original hypotheses and theories evolved over the twentieth century into two great pillars of theoretical thinking and analysis. Today, these pillars of physics are called 'quantum theory' or **quantum mechanics**, and **general relativity**. Along the way, it has taken the profound thoughts of dozens of brilliant minds in physics to bring these theories to their present stage of development. In this chapter, we will take a short glimpse at some of this historical work. Sometimes the path is highly intertwined, but it is never boring. Both of these theories have given us a picture of our surroundings, from the infinitesimally small subnuclear domain within the atom to the vast reaches of space and the nature of the universe itself. The two great theories are independent:

- General relativity successfully describes the motion and behaviour of bulk matter and its gravitational interaction by the radiation of gravity waves.
- Quantum theory successfully explains the behaviour of subatomic matter in terms of constituent particles and their force interactions, which has culminated in the **standard model** of particle physics.

Standard quantum theory

Standard quantum theory today gives us three fundamental forces. These act between, and within, individual atoms of matter that are made up of twelve basic particles. These forces are:

- the **electromagnetic force**, which holds the electrons within the atom
- the **strong nuclear force**, which binds the nucleus together
- the **weak nuclear force**, responsible for radioactive decay and the interactions of nature's most amazing particle, the neutrino.

A fourth fundamental force is called the **gravitational force**. Gravity acts over huge distances and holds the universe together. It is in the realm of general relativity and space–time. Gravity, surprisingly enough, is the least well understood, despite the efforts of Newton and Einstein. This force is still the odd one out in a grand unified theory of everything, or '**TOE**' as physicists call it. Physics will need in the future to develop a concise TOE if it is to answer the big questions of — Who we are? What we are? and Where we are? Famous physicists, such as Stephen Hawking and colleagues, are working to combine quantum theory and general relativity, but it is complex theoretical work. Let's go back to the start!

Ch29-Walding 4th 25/8/04 1:40 PM Page 655

QUANTUM THEORY EFFECTS 29.2

Planck's black body radiation

At the beginning of the last century, interesting experiments were being performed on the nature of the radiation emitted by a **black body**. A black metallic object will not reflect any light shining onto it, so as it is heated, any light radiation that it emits is solely coming from within itself. A good example is the electric hotplate of a stove, which begins to glow red, then orange and even white if it is allowed to become hot enough. The distribution of intensity versus frequency of light emitted for this type of hot body is given in Figure 29.1.

Figure 29.1
Black body radiation.

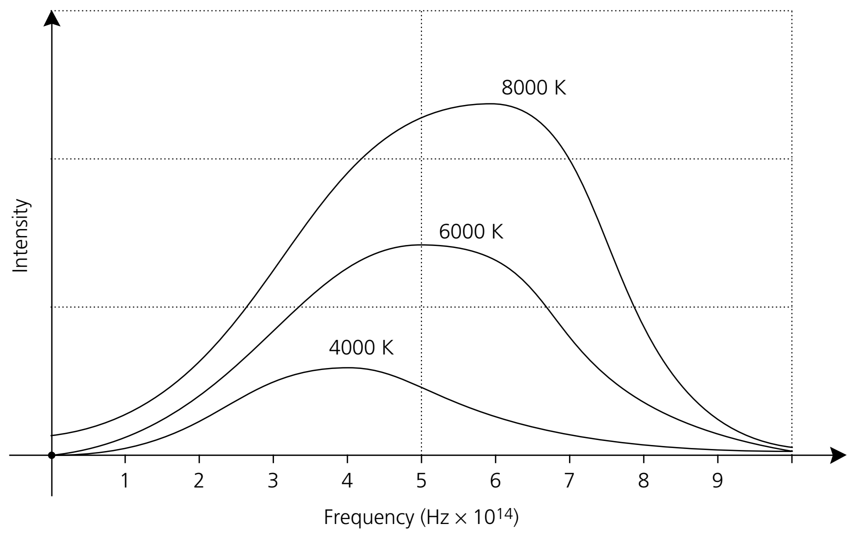
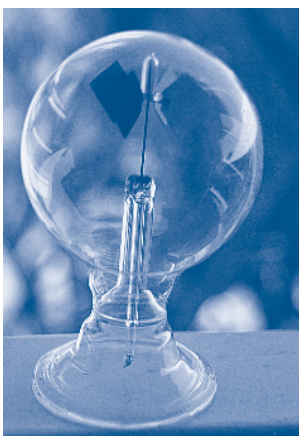


Photo 29.1
Crookes' radiometer with the black and white vanes clearly visible.



The shapes of these graphs at different temperatures went against all theoretical predictions based on James Clerk Maxwell's electromagnetic theories. Questions such as, 'Why weren't ultraviolet, X-rays or gamma rays produced?' or 'Why was there more red frequency radiation than blue?' could not be satisfactorily answered. Physicists such as Robert Kirchhoff and Nobel laureates John Rayleigh and Wilhelm Wien had produced equations that described only parts of these distribution curves, but none could satisfactorily describe the whole range.

A German physicist, **Max Planck** (1858–1947), finally produced the equation that did describe the black body distribution, and in doing so, he proposed a revolutionary theory of subatomic matter. Planck proposed that the energy released by a black body was, in fact, emitted by atoms, and that these atoms could only vibrate at certain frequencies that were multiples of some smallest value. He had to assume that the energy released by the atoms was not given off continuously, but in small energy packets that he called **quanta** (singular quantum), from the Latin *quantus*, meaning 'how much'. Each frequency, f, of light emitted by the atoms is proportional to the change in energy of the atom, so that, for example, since violet light is twice the frequency of red light, the energy quanta of violet light are twice the size of those of red light. Mathematically, the quanta energy is given by $E = hf$, where the constant h is called the **Planck constant** and has a value of 6.63×10^{-34} J s. Since the Planck constant is extremely small in magnitude, energy quanta are not noticeable in most everyday circumstances. A typical light source such as an incandescent bulb releases millions of quanta per second, which lead to the amount of light energy that we are familiar with.

With this idea in place, Planck was able to describe the reason for the absence of high energy emissions from black bodies. The vibrating atoms were simply not large enough to

NOVEL CHALLENGE

A Crookes' radiometer consists of four paddles suspended on a needle point in a low pressure glass container. One side of the paddle is painted black, the other side white. When placed in the Sun it turns around. *Explain whether the black side moves away from the sun or towards it (and why).* It goes the opposite way near a block of dry ice (–44°C). Most people (even scientists) gave the wrong explanation. Check our web page and all will be revealed.

provide the necessary energy changes. Also, certain states of vibration of the atoms were more likely and this accounted for the peak in the frequency distribution curves. As we will see later, Planck's idea that the whole atom vibrates is, in fact, not quite correct. Energy emissions are due to electron movements (transitions) within the atom. Quantum theory today shows that electrons in atoms can only move between defined energy levels within the atom. Planck himself did not have any evidence for energy quanta, but it was an excellent idea that perfectly described solutions to several problems in physics at the time. The quantum theory has provided the basis for all modern physics since 1900 and for his work, Max Planck received the 1918 Nobel prize for physics. It now remained for the quantum idea to be applied to both light and matter.

Light itself can be assumed to come in small packets called **photons**, which give light radiation a reason for behaving like particle systems, under certain conditions. If light radiation is governed by the wave equation for velocity, c, frequency and wavelength, namely $c = f\lambda$, then light photons will have energy given by:

$$E = hf = \frac{hc}{\lambda}$$

Note: c is the velocity of electromagnetic radiation (light) and equals 3.0×10^8 m s^{-1}.

NOVEL CHALLENGE

The **photon** was named by US physicist Gilbert Lewis in 1926 using the Greek *phos*, meaning 'light'. See if you can prove these statements wrong (we doubt you can):

A All words beginning with the prefix *phos* are related to the concept of light.

B The planet Venus used to be called Phosphor when appearing as the morning star.

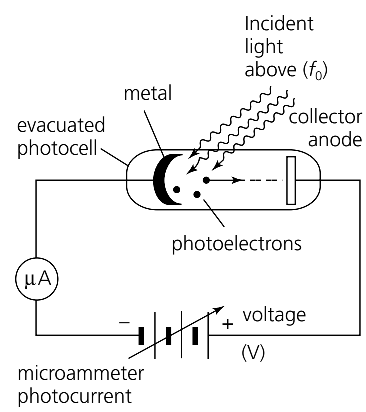
The photoelectric effect

Further proof of the quantum idea came when **Albert Einstein** (1879–1955) applied the theory to explain the **photoelectric effect**. When a metal surface is illuminated by a high-frequency light source, electrons may be ejected from the metal as a **photocurrent** with definite characteristics (Figure 29.2). Experiments on this effect, by physicists from as early as 1887, had confirmed that electrons were ejected from the metal only if the frequency of the incident light exceeded a minimum value called the **threshold frequency**, f_0, which was different for various metals. Even very intense light, if the frequency was below the threshold value, would not eject electrons and cause the flow of the photocurrent. Two other important characteristics of the photoelectric effect are:

- Once a photocurrent is registered, increasing the incident light intensity increases the amount of photocurrent flowing.
- Light of a higher frequency than that required to produce a photocurrent increases the kinetic energy of the ejected electrons.

The electron kinetic energy is measured by a negative potential applied to the collector plate, which repels the ejected electrons and eventually becomes large enough to stop the photocurrent. This reverse **cut-off voltage**, V_c, applied to the collecting plates in the electron tube is also called the **stopping potential**. This is the opposite to an electron gun.

Figure 29.2
Photoelectric apparatus.



Each of these experimental observations was impossible to explain using conventional wave theories of light. Einstein applied the newly developed quantum theory to this effect in 1905, and provided the perfect explanation. His explanation revived the light particle model, and for this effort he was later awarded the Nobel prize for physics in 1921.

Einstein assumed that the light quanta, called photons, interacted with the surface electrons in the metal so that a single photon ejects a single electron. The photon will give either all of its energy to the electron or none of it. Each electron can only absorb the energy of one photon and the collision interactions between photons and electrons in the metal are totally elastic and obey the law of conservation of energy. Einstein defined three forms of energy in the system, namely:

- photon energy, $E = hf$, which is frequency-dependent
- work function or energy of binding of the electron to the metal, which is measured as $W = hf_0$, where f_0 is the threshold frequency
- maximum kinetic energy, $E_{K(max)}$, of the ejected electrons from the metal surface.

NOVEL CHALLENGE

At night, turn all the lights off and turn the oven element ON. Feel the heat before it glows. Watch the element glow and see how it changes from red to red/orange. In furnaces, the elements change colour further and end up almost white. If the blue–violet end of the spectrum indicates higher energy, why doesn't the element go red → blue instead of red → white?

Einstein's photoelectric equation relates these energy values together such that conservation occurs, namely:

$$E_{K(max)} = hf - W$$

or

$$\left(\frac{1}{2}mv^2\right) = hf - hf_0 = qV_c$$

where V_c is the cut-off voltage necessary to reduce the flowing photocurrent to zero; v is the ejected electron velocity.

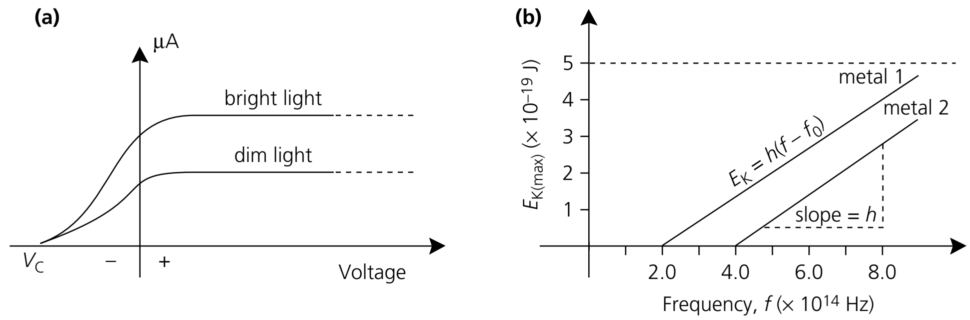


Figure 29.3 Results from photoelectric experiments: **(a)** I/V characteristics; **(b)** $E_{K\ MAX}$/frequency.

NOVEL CHALLENGE

In a dark room let your eyes become dark-adapted. Give your eyelid a sharp tap with your finger and you should see a flash. A single rod will detect a single photon, but your visual system only responds when between two and ten photons are absorbed by your rods within 0.1 seconds; you will then see the flash. Estimate how much power two visible photons will give to your eyes in 0.1 seconds.

Figure 29.3 represents a typical set of graphs obtained from photoelectric experiments carried out on various metals. Notice that the slope of the straight lines of graph **(b)** can be used to calculate the value of Planck's constant. This experimental determination of h was first performed by Robert Millikan in 1916.

The quantum theory was by now well and truly established both in theory and in experiment. The term photoelectric effect can be applied equally well to other phenomena such as photoionisation in gases, whereby light radiation can ionise gas atoms, or photoconduction where incident light photons are absorbed by various crystalline materials, giving their electrons enough energy to break free and become electrical conductors. Today, photovoltaic semiconductor materials are common, such as solar cells, photo diodes and transistors. In these materials, incident light photons of sufficient energy create electron-hole pairs in the crystal and increase electrical conduction. Refer back to Chapter 23.

Questions

1. What are the four fundamental forces in nature? On what do they each act? Which has the biggest range, and the smallest range?
2. Calculate the energy and wavelength of light of frequency 4.3×10^{14} Hz. What colour would it appear to our eye?
3. Which light has the more energetic photons, red or violet? Explain why.
4. If the threshold frequency for rubidium metal is 5.0×10^{14} Hz, calculate the value of the work function of the metal and the maximum velocity of photoejected electrons when the metal is illuminated by light photons of frequency 8.1×10^{14} Hz. The mass of the electron is 9.11×10^{-31} kg.

The Compton effect and light pressure

Further evidence for the particle nature of electromagnetic radiation came with the discovery of X-ray scattering, by **Arthur Holly Compton** (1892–1962). Compton used apparatus as shown in Figure 29.4, and showed that the X-ray photons behaved like particles with definite momentum characteristics. The X-rays collided with the electrons in the graphite target. The scattered X-ray photons, after collision, possessed reduced energy and longer wavelengths when compared with the unscattered photons. In a **Compton collision**, between an X-ray photon and an electron, the change in energy is not complete and the reduction in energy

Ch29-Walding 4th 25/8/04 1:40 PM Page 658

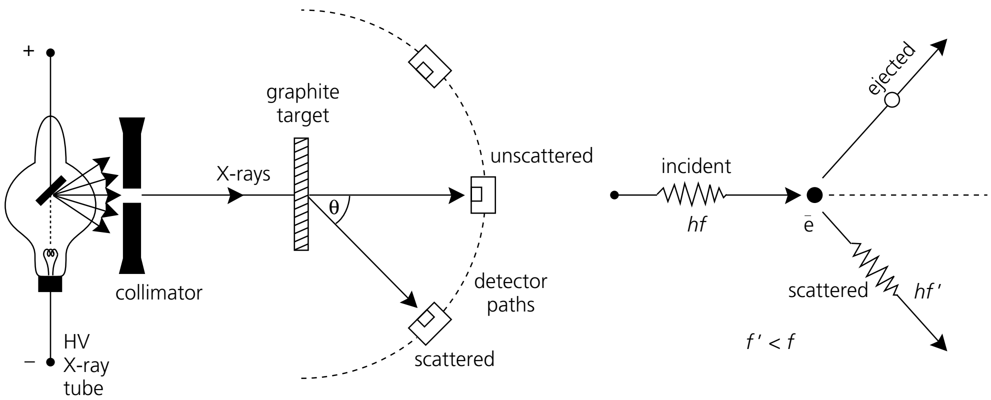


Figure 29.4
Compton scattering.

and wavelength is dependent on the angle of scattering. The electron involved is scattered or ejected from the graphite in such a way that both energy and momentum are conserved in the collision. Remember that momentum is a vector quantity. Compton's collision calculations correctly predicted the speed and direction of the recoil electrons. For this work, as well as further X-ray spectra analysis, Compton shared the 1927 Nobel prize for physics with British physicist Charles Wilson.

When considering the photon as a particle, we can derive a formula for the photon momentum. Einstein's mass–energy equivalence relationship $E = mc^2$ links the idea of physical mass to energy; however, the quantum idea states that photon energy is $E = hf$, thus the photon particle energy will be:

$$E = mc^2 = hf$$

or

$$mc = \frac{hf}{c}$$

But mc is the definition of photon momentum, p, hence:

$$p = \frac{hf}{c} = \frac{h}{\lambda} \text{ where } c = f\lambda$$

Since light or electromagnetic photons have momentum, $p = \frac{h}{\lambda}$, then, in collisions with surfaces, they should be able to exert a force and create **light pressure**. This is exactly what does occur in practice. The pressure exerted depends on the rate of change of momentum per unit area of illuminated surface. The pressure of light is extremely small at the Earth's surface. It is a factor of 2.5×10^{10} less than standard atmospheric pressure. As early as 1903, Edward Nichols and George Hull measured light pressure using mirrors and a sensitive suspended fibre torsion balance, achieving a result of 7.01×10^{-6} N m^{-2}.

The revolution in thinking caused by the quantum theory and its successful application to black body radiation, the photoelectric effect and X-ray Compton scattering, caused electromagnetic energy to be given a dual nature by physicists. The **wave–particle duality** concept for light and other forms of electromagnetic energy is our current explanation. If we are describing what light is (!) then we need to consider what we are explaining. In general, if light energy is interacting with other forms of light energy, then the wave behaviour model is the best explanation, as, for example, in optical effects such as interference and diffraction. If light is interacting with matter, then the particle behaviour model is the best explanation, as, for example in Compton collisions. The mathematical model bringing together the wave–particle duality concept used to describe matter and energy is called wave mechanics or quantum mechanics and will be discussed further in Section 29.4. First, before taking a look at these ideas, let's further investigate the models and theories applying to the atom and see how the quantum idea is vitally important here also.

THE BOHR ATOM AND ATOMIC SPECTRA 29.3

In 1911, the New Zealand-born British physicist Ernest Rutherford had established the existence of the atomic nucleus, and he made it possible to consider the simplest atom of hydrogen as a single positive charge with a single negative electron circling it in planetary fashion. This atomic model had a serious flaw in that, according to the electromagnetic equations of Maxwell, any electron revolving in circular fashion around a nucleus is under centripetal acceleration and should continuously radiate electromagnetic energy. This would allow the electron to continuously lose energy and cause it to spiral in toward the nucleus. Thus, the equations predicted that the Rutherford atom should be highly unstable and not exist for any length of time as it would quickly lose its energy and collapse. Clearly, this was not what actually happens.

In 1913, Danish physicist Niels Bohr applied the quantum concept to the problem and proposed a revolutionary hypothesis. His idea was that the electron would only radiate energy in exact quanta or definite amounts of energy. As it did so, it would move inward toward the nucleus in definite **quantum orbitals** or allowed orbits until a stable orbit was reached. This stable orbit is called a **stationary state**. Normal atoms exist with their electrons in stationary states, but if energy is added to any atom, such as by particle bombardment or sufficient heating, then the electrons are forced into higher energy states (orbitals) temporarily by absorbing definite quanta. This process of absorption produces an **excitation energy** state. As the atom restabilises, the electron jumps back down to a stationary state in a possible series of steps. Each orbital jump results in the emission of a photon of electromagnetic energy of definite predictable value.

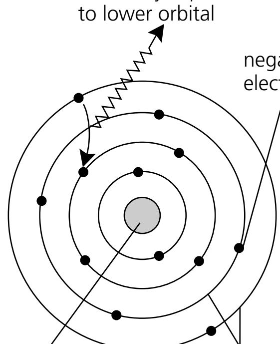


Figure 29.5
The Bohr atomic model.

Thus, every change in orbit by an electron corresponds to the absorption or emission of a quantum of electromagnetic radiation. (Refer to Figure 29.5.) If an atom absorbs too much energy then the outermost electron will be promoted completely away from the attraction of the nucleus and will be lost. This is called **ionisation energy** and for the simplest hydrogen atom is equal to 2.17×10^{-18} J. Bohr also proposed that within the atom only two electrons could occupy the same orbital at any one time. Further work on this idea resulted in the **Pauli exclusion principle** and it became possible to show that atoms are arranged in the periodic table as a result of electrons being arranged in definite patterns from the lowest energy orbitals outward.

The **spectrum** or range of emitted light released by hydrogen gas atoms had been known since 1885 when Johann Balmer, a Swiss physicist, had worked out a mathematical link between the wavelengths of the light colours emitted. You should recall that the component wavelengths of any light source can be examined by passing the light through a **prism spectrometer**, which splits the light into its spectral colour components. Niels Bohr could now choose orbits for the hydrogen electron that would yield exactly the required wavelengths for the emitted spectral lines of the hydrogen spectrum, according to the generalised Balmer equation:

$$\frac{1}{\lambda} = R_H\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

where R_H is the Rydberg constant (1.097×10^7 m^{-1}); n_f and n_i are the initial and final **principal quantum numbers**.

Figure 29.6 represents the **energy level diagram** for hydrogen that correlates the Bohr orbitals and their corresponding energies with the series of spectral lines present in the hydrogen spectrum. The spectral series are named after their discoverers.

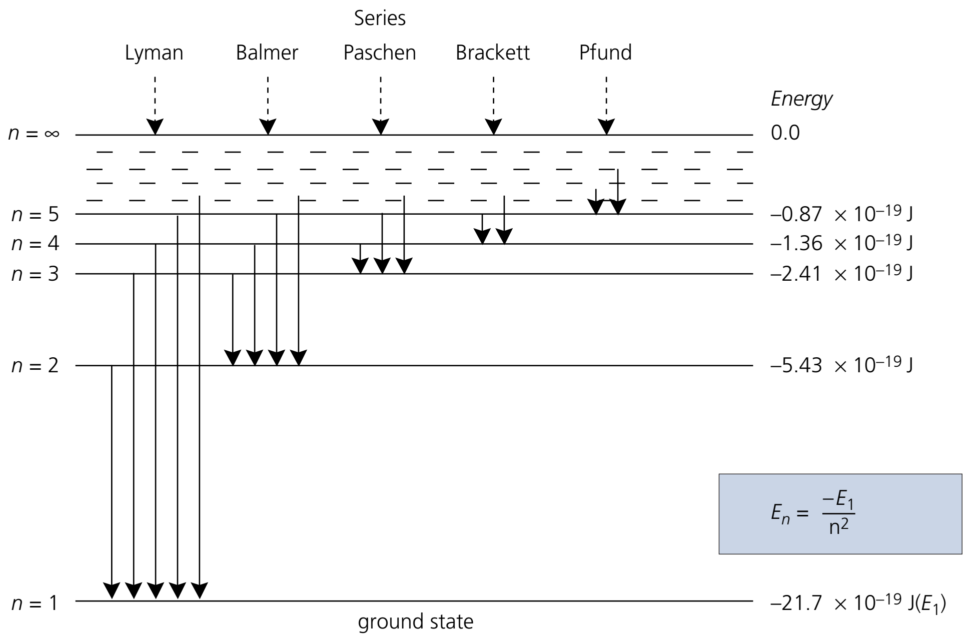


Figure 29.6
Energy level diagram for hydrogen.

Mathematically, Bohr's theory used a **quantum condition** that specified that the angular momentum of the electron was restricted to allowed orbits, given by:

$$mvr_n = \frac{nh}{2\pi}$$

where $n = 1, 2, 3 \ldots$ integers; r_n is the radius of the n^{th} orbit.

He was able to show that, at least for the simplest hydrogen atom, the radius of the electron for each quantum orbital possible is given by:

$$r_n = \frac{n^2h^2}{4\pi^2kmq^2}$$

where n is the principal number; k is the Coulomb constant; m, q are electron mass and charge.

Thus, for $n = 1$, the **Bohr radius** takes a value of 5.3×10^{-11} m, and represents the orbital radius for the electron in the lowest energy state or **ground state** of the hydrogen atom. Notice that in the energy level diagram of atomic hydrogen, the energy associated with the ground state is a negative value of -21.7×10^{-19} J and represents the ionisation energy for hydrogen. In order to remove an electron out to infinity, this amount of energy is required. Because we can say that the electron at infinity has zero energy, by definition, all possible energy states of the hydrogen atom can be regarded as negative. According to the Bohr theory, the energy of each quantum orbital is given by the series:

$$E_n = E_1\left(\frac{1}{n^2}\right) \text{ where } E_1 \text{ is the ground state energy}$$

$$E_1 = -2.17 \times 10^{-18} \text{ J or } -13.6 \text{ eV (electronvolts)}$$

Notice that other spectral series represented in the hydrogen spectrum are named after the people who discovered them. It is only the Balmer series of jumps back to $n = 2$ that represents visible light emissions. You might like to check this out, using an energy–wavelength calculation.

Bohr's theory was very good at predicting the spectral line series of the hydrogen atom, but could not correctly predict those for more complex atoms, nor could it predict other observable features such as spectral line intensity differences and fine splitting of the lines themselves within a magnetic field. Even the basic notion of why the electron oscillated only within defined quantum orbitals could not be explained. Nevertheless, his application of the quantum theory to atomic structure was very important and, for his work, Bohr gained the 1922 Nobel prize for physics. Today, the newly discovered artificial radioactive element of atomic number 107 is called **Nielsbohrium** (Ns). The isotope was discovered by a Soviet group at Dubna in 1976 by bombarding bismuth with chromium ions to form $^{261}_{107}$Ns.

NOVEL CHALLENGE

We read in a US science magazine that if you shone a laser beam onto glow-in-the-dark plastic it would go dark (yes, dark!) where the laser hit. This sounds like rubbish, but we tried it. What do you think happened?

Example

(a) Determine the energy of an electron in both the fourth and second quantum orbitals of the hydrogen atom.

(b) What is the frequency of the energy emitted when an electron jumps between these orbitals?

(c) Calculate the wavelength of this emitted light **(i)** in metres, **(ii)** in nanometres.

Solution

(a) Fourth level $n = 4$.

$$E_4 = \frac{E_1}{4^2} = \frac{-2.17 \times 10^{-18}}{16} = -1.36 \times 10^{-19}\ \text{J}$$

Second level $n = 2$.

$$E_2 = \frac{E_1}{2^2} = \frac{-2.17 \times 10^{-18}}{4} = -5.34 \times 10^{-19}\ \text{J}$$

Electron jump, energy released is:

$$\Delta E = E_i - E_f = E_4 - E_2$$
$$\Delta E = 4.07 \times 10^{-19}\ \text{J}$$

(b) Use the equation $\Delta E = hf$ or $f = \frac{\Delta E}{h}$

$$f = \frac{4.07 \times 10^{-19}}{6.63 \times 10^{-34}} = 6.14 \times 10^{14}\ \text{Hz}$$

which represents the Balmer series line of colour blue.

(c) Use the equation $v = f\lambda$ or $= \frac{v}{f}$

$$\lambda = \frac{3 \times 10^8}{6.14 \times 10^{14}} = 4.88 \times 10^{-7}\ \text{m}$$

Convert to nanometers: 4.88×10^{-7} m = 488×10^{-9} m = 488 nm

Franck–Hertz experiment

In 1914 two German physicists, James Franck and Gustav Hertz, the nephew of Heinrich Hertz, performed a very important experiment that supported Bohr's ideas on quantum atomic absorption and emission. Their apparatus is represented in Figure 29.7. Your school laboratory probably has a demonstration electronic valve apparatus that can be used to obtain similar results. The glass chamber contains mercury vapour at a low pressure of about

Ch29-Walding 4th 25/8/04 1:40 PM Page 662

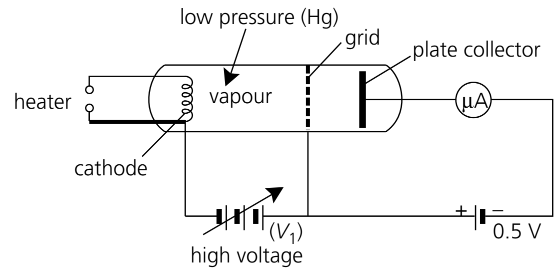


Figure 29.7
The Franck–Hertz apparatus.

1.0 mm Hg. A hot cathode emits electrons towards a mesh grid that is maintained at some variable potential with respect to the cathode. Beyond the grid mesh is a solid metal plate maintained at about negative 0.5 V with respect to the grid that collected the high energy electrons and allowed the measurement of tube current by the microammeter. The experiment involved gradually increasing potential V_1 and noting the tube output current. Typical results are shown graphically in Figure 29.8, which consists of a series of current peaks and troughs separated by an average value of 4.9 V.

Franck and Hertz explained these results in terms of quantum absorption. At voltages below 4.9 V, the electrons interact elastically with the mercury atoms. At 4.9 V, the electrons transfer most of their energy to the mercury atoms because the first excitation energy for mercury is 4.86 eV. They now do not have enough energy to reach the collecting plate and the current falls into a trough. If the voltage is increased again, the electrons gain enough energy to reach the plate again. At 9.8 V, the electrons can make two inelastic collisions with the mercury atoms and so the current falls again into a trough. If the spectrum of the mercury is examined, an ultraviolet line can be found at 253 nm, which corresponds to the emission from the atoms of photons of energy 4.9 eV as they return to their ground state. It is this wavelength that is produced in a fluorescent light tube and converted to white light by the phosphor coating on the inside of the glass tube itself.

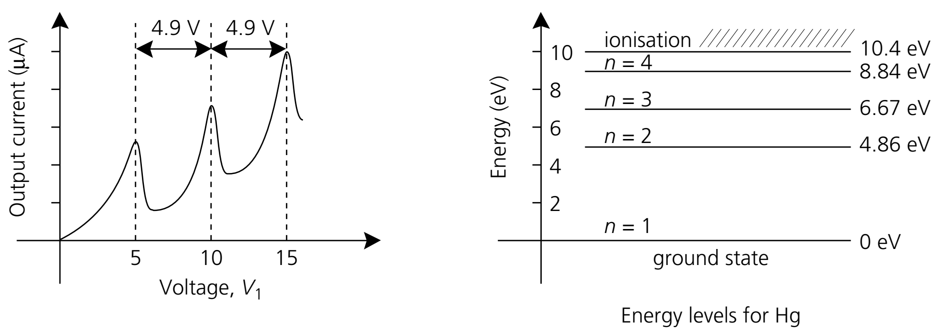


Figure 29.8
Graphical results of the Franck–Hertz experiment and the energy level diagram for mercury.

NOVEL CHALLENGE

In 1800, English astronomer William Herschel placed a thermometer in various parts of the spectrum of sunlight. He found that the highest temperature was beside the red where there was no colour. Explain that if you can.

The Franck–Hertz experiment verified that atoms contain discrete energy levels and cannot absorb random amounts of energy. The colliding electrons lose energy only in discrete quantum chunks corresponding to precise energy differences between the atom's energy states. This same energy quantum is reradiated as a precise single wavelength when the excited mercury atom returns to its ground state. Figure 29.8 illustrates the energy level diagram for mercury. It should be realised at this point that every atom has its own characteristic energy level diagram and thus the excitation spectra will be like an atomic fingerprint. (See the photo in the colour section.)

Spectroscopy

Spectroscopy is used in analytical processes to definitively identify individual atoms or molecules. It is a very powerful technique. The word spectrum comes from the Latin *specere* meaning 'to look at'. The first prism spectroscope was, in fact, designed by Gustav Kirchhoff and Robert Bunsen in 1859 while working on chemical analysis.

Ch29-Walding 4th 25/8/04 1:40 PM Page 663

When light from a hydrogen discharge tube is examined through a spectroscope, distinct lines appear on a black background, each one corresponding to an electron transition in the hydrogen atom. (See Figure 29.9 below.)

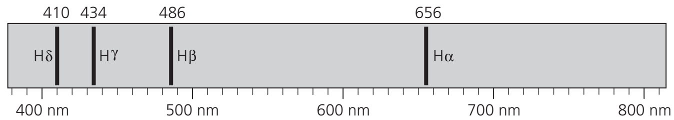


Figure 29.9
The spectrum of hydrogen shows the four lines of the Balmer series, all of which are in the visible region.

As mentioned previously, four of the lines (the Balmer series) are in the visible region. They are labelled alpha, beta, gamma and delta and correspond to the following transitions:

Table 29.1

LABEL	TRANSITION	ENERGY OF PHOTON (J)	FREQUENCY (Hz)	WAVELENGTH (nm)	COLOUR
H_α	$3 \rightarrow 2$	3.02×10^{-19}	4.57×10^{14}	656	red
H_β	$4 \rightarrow 2$	4.09×10^{-19}	6.17×10^{14}	486	green
H_γ	$5 \rightarrow 2$	4.58×10^{-19}	6.91×10^{14}	434	blue
H_δ	$6 \rightarrow 2$	4.85×10^{-19}	7.31×10^{14}	410	violet

When physicists turned their spectrometers towards the heavens and examined the spectra of starlight they found spectral lines characteristic of elements they had examined on Earth. The spectra of hydrogen and helium were particularly noticeable, but other elements such as calcium gave strong lines in their spectrometers. Astronomers were thus able to infer the composition of stars from their spectral signatures. But what was most astonishing was that many of the characteristic patterns were shifted towards the red end (low wavelength) of the spectrum. The term 'red shift' was coined to describe this phenomenon.

Recalling that the frequency of a sound changes as a source moves in relation to an observer, physicists used this very same Doppler effect (see Chapter 16, Section 13) to propose that the red shift was due to the motion of stars speeding away (receding) from us. This is called **radial** or **recessional velocity** (RV).

In a star that is at rest with respect to us (the Sun), or in a hydrogen discharge tube in the laboratory, the hydrogen line wavelengths are 410, 434, 486 and 656 nm. By measuring the amount of shift towards the red, we can determine how fast the star or galaxy is moving away. For example, Figure 29.10 shows the line spectrum of standard hydrogen (Spectrum I) and for two objects that have red-shifted spectra.

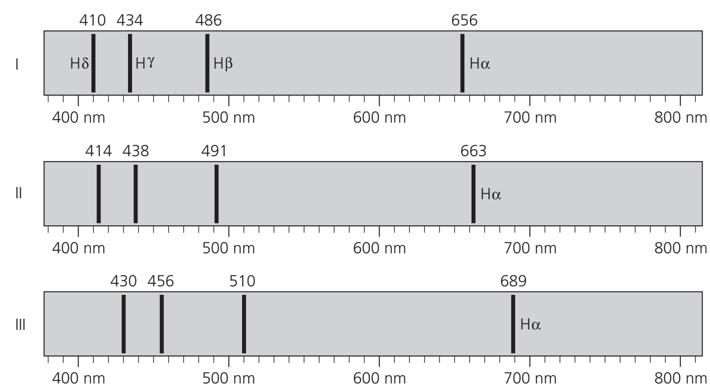


Figure 29.10
Comparison of hydrogen spectra from a laboratory source (I) and from stars speeding away from us (II and III). Note the red shift.

Spectrum II is for the galaxy Centaurus. Note that the Hα line is red-shifted by 7 nm from the standard 656 nm to 663 nm. A shift of 7 nm from 656 nm is a ratio of 7/656 or 0.01, which means the galaxy is travelling away from us at 0.01 times the speed of light (0.01*c*). This corresponds to a speed of 3200 km s^{-1}. The other hydrogen lines are shifted by the same ratio.

The red shift z is defined such that:

$$z = \frac{\lambda_0 - \lambda}{\lambda_0} = \frac{\Delta\lambda}{\lambda_0} = \frac{v}{c}\text{, so } v = c\frac{\Delta\lambda}{\lambda_0} = cz$$

Example

Spectrum III is for the galaxy Ursa Major I. Calculate its radial velocity.

Solution

The Hα line has been shifted from 656 nm (λ_o) to 689 nm (λ), hence $\Delta\lambda$ = 23 nm. The red shift ratio (z) = 23/656 = 0.05, so it is moving at 0.05 times the speed of light (0.05c) away from us. This equals 10 500 km s^{-1}.

Note that this equation only works for galaxies moving a few tenths the speed of light or slower. Those with large z values need a relativistic version of the above equation:

$$1 + z = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

If you want a challenge, show that the speed of the galactic cluster named 3C295 is actually 0.46c (using the relativistic formula), not 1.64c as predicted using the non-relativistic formula. The Hα line has an observed wavelength of 1076 nm.

When z is larger than 1, cz is faster than the speed of light, and while recessional velocities faster than light are allowed, this approximation using cz as the radial velocity of an object is no longer valid. Thus for the largest known red shift of z = 6.3, the recessional velocity is not 6.3c = 1 890 000 km s^{-1}. It is also not the 285 254 km s^{-1} given by the special relativistic Doppler formula. The actual recessional velocity for this object depends on the cosmological parameter omega (Ω) which is a measure of the expansion of the universe (see Chapter 6).

Activity 29.1 IN A GALAXY FAR, FAR AWAY...

1 The spectrum of light from the galaxy Hydra is shown in Figure 29.11. Compare the radial velocity of Hydra as calculated by using both the non-relativistic and relativistic formulas.

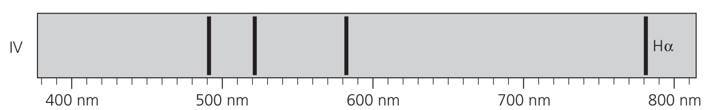


Figure 29.11
Spectrum of the galaxy Hydra.

2 The calcium-K line and the calcium-H line for the galaxy Leo are 419 nm and 398 nm respectively. Locate the standard Ca-K and Ca-H wavelengths and calculate the velocity of Leo.

3 The galaxy Persus is known to have a radial velocity of 5430 km s^{-1}. Draw a simple line spectrum to show the relative spacing of the four hydrogen lines in Persus's spectrum.

Questions

5 Show that the emitted photon from a mercury atom dropping from its first excitation energy level to the ground state is, in fact, an ultraviolet photon.

6 Is it fair to say that Compton scattering between photons and electrons is like billiard balls colliding? Explain.

NOVEL CHALLENGE

In 1801, German scientist J. W. Ritter put a piece of photographic paper in the spectrum of sunlight and found that the greatest blackening was beside violet where there was no colour. Explain that if you can.

7 Define these terms as applied to quantum atomic theory: **(a)** quantum orbital; **(b)** excitation energy; **(c)** ionisation energy; **(d)** principal quantum number; **(e)** Nielsbohrium.

8 Calculate the wavelengths of the first three lines of the Lyman series in the spectrum of hydrogen. To what part of the electromagnetic spectrum do they belong? (See Figure 29.6.)

9 Explain why the spectrum of hydrogen contains several very bright lines while the atom itself contains only one electron and one proton.

10 In a Franck–Hertz experiment carried out with potassium vapour, it is found that current falls off rapidly at an applied voltage of 1.62 V. Calculate the wavelength of the expected spectral line in the emission spectrum of potassium when this voltage is reached.

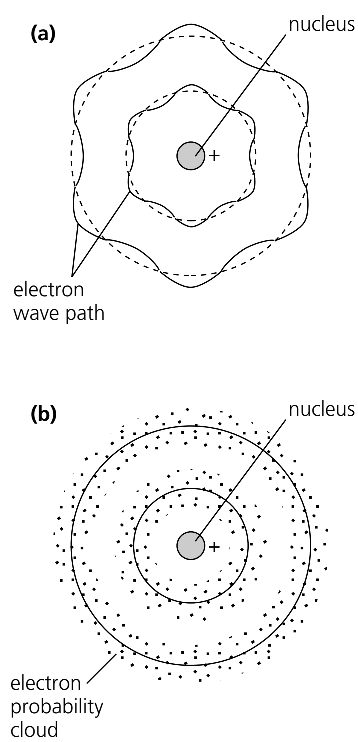
Activity 29.2 SPECTACULAR COLOURS

1 Set up some spectrum discharge tubes using a high voltage induction coil. Obtain your teacher's assistance to do this and use a simple direct vision spectroscope or diffraction grating to observe the spectra. Note both the colour to the eye and the primary lines of the emission spectra. You should try gases such as H_2, He, Ne, CO_2.

2 Placing small amounts of crystalline ionic salts (preferably chlorides) into the flame of a bunsen burner using a clean platinum loop provides a display of characteristic metallic atom spectral colours. Use this to explain the brilliant colours of fireworks.

QUANTUM MECHANICS 29.4

Figure 29.12
Electron standing waves: **(a)** orbital wavelengths; **(b)** electron clouds.



The problem of how an electron could exist in quantum orbitals without losing energy was solved in 1924 when the French physicist Louis Victor de Broglie suggested that matter could also exhibit wave-like characteristics. He called these **matter waves**. Louis de Broglie postulated that an electron particle could have a wavelength $\lambda = \frac{h}{mv}$, just as the photon has a wavelength $\lambda = \frac{h}{p}$ as a result of its momentum.

This idea allowed Bohr's quantum orbitals to be considered as electron wave orbits whose circumference contained an integral number of wavelengths (Figure 29.12**(a)**). The **standing waves** of the electrons in orbit would not require any loss of energy and the angular momentum of the electrons in their orbits is quantised. This de Broglie prediction was experimentally verified by the American team of Clinton Davisson and Lester Germer, as well as the British physicist George Thomson. They showed that a beam of electrons scattered by crystals does, in fact, produce a characteristic wave diffraction pattern.

The de Broglie wavelengths of anything, except subatomic particles, are very short. It makes little sense, for instance, to think of the de Broglie wavelength of a Ford Fairlane driving down the road at 80 km h^{-1} even though such a quantity exists. Assuming a mass of 1642 kg for the car, show that the de Broglie wavelength is 1.8×10^{-38} m. In practice, this value is immeasurably small and can be neglected.

The wave particle concept has led to very complex mathematical models of the nature of atomic structure, called wave mechanics. **Wave equations**, developed by the Austrian physicist Erwin Schrödinger, describe the wave properties of electrons in both hydrogen and helium atoms. The solutions of Schrödinger's wave equations also indicate that no two electrons can possess the same set of characteristics defined by quantum numbers. This verified the exclusion principle established by Wolfgang Pauli in 1925. Further mathematical refinements by German theorists Max Born, Ernst Jordan and Werner Heisenberg led to 'matrix mechanics' theory, which is very successful in making predictions about atomic behaviour.

Although quantum mechanics describes an atom in purely mathematical terms, a verbal description and a visual model can be constructed for our modern view of the atom. Surrounding the dense nucleus of any atom is a series of standing wave electron orbitals with wave crests at certain points. The square of the wave amplitude at any point is a measure of the **probability** that an electron can be found at that point at any given time. This gives us a picture of an **electron cloud** around the nucleus. (See Figure 29.12**(b)**.) This probability is as good as we can get to defining the position of any electron and is a result of the **uncertainty principle**, developed by Werner Heisenberg in 1927. His work pointed out that any measurement made on a physical system will, in fact, change the system itself and introduce a fundamental uncertainty into measurements of all other properties of that system. Heisenberg was awarded the 1932 Nobel prize for physics for his contribution to quantum mechanics. There is a hotel in England with an inscription above the door that reads 'Heisenberg may have slept here!'

The principle states: 'It is impossible to measure the position and the corresponding momentum of a particle simultaneously with complete accuracy. The product of the uncertainty in the position and momentum is greater than, or at best equal to, $h/4\pi$.'

Mathematically: $$\Delta x \, \Delta p = \frac{h}{4\pi}$$

Again, it might be obvious that this effect is really only important in the subatomic domain. For example, if an electron is measured with a velocity of 4.4×10^6 m s^{-1}, with an uncertainty of 0.1%, then the value $\Delta x = 1.3 \times 10^{-8}$ m represents the positional uncertainty of the electron. Check it out. This uncertainty is, in fact, about 100 times the diameter of the hydrogen atom, so this principle will not even allow us to determine if the electron is within the atom. The uncertainty principle places large limits on measurement of atomic properties.

Quantum mechanics has solved a lot of the great scientific problems that have troubled physicists. It is interesting to note, however, that even Albert Einstein had difficulties with the ideas of quantum mechanics and had many famous arguments with Niels Bohr on the subject. It was Einstein, though, who proposed Heisenberg for the Nobel prize with the endorsement: 'I am convinced that this theory undoubtedly contains part of the ultimate truth'. Perhaps one of the most striking features of quantum physics that has only recently been discovered is that it is not possible in general to say when things 'actually happen'. Time itself is very peculiar indeed in quantum physics!

Quantum mechanics has given us a picture of atomic structure, and explained spectral emissions and chemical bonding processes. Most importantly, it has led to an almost complete picture of the fundamental forces and particles of nature. Let's take a look at these now.

NOVEL CHALLENGE

In 2002, Brisbane student John Prior put a piece of photographic paper in a microwave oven and turned it on high. He knew that microwaves have a longer wavelength than infrared but nothing happened. Are you surprised to hear of his findings? Explain.

29.5 THE STANDARD MODEL

Fundamental forces

We have so far in this book recognised two fundamental forces of nature. We've examined mathematically both the gravitational and electromagnetic forces and understood that they both act over large distances and obey inverse square laws with distance. (See Table 29.2.)

The gravitational force keeps planets in orbit, controls the expansion of the universe and stops us from falling off the Earth into space. The electromagnetic force holds electrons in their atoms and binds matter together as molecules. Four forces in total are required to completely describe the universe around us, and in this section we look more closely at the interactions of the other two fundamental forces called the **strong interaction** or strong nuclear force and the **weak interaction** or weak nuclear force.

Table 29.2 FOUR FUNDAMENTAL FORCES

FORCE	EFFECTS	RELATIVE STRENGTH	RANGE
Gravitational	all interactions	1×10^{-38}	large distances inverse square
Electromagnetic	charged particle interactions	1×10^{-2}	large distances inverse square
Weak nuclear	weak interactions e.g. beta decay	1×10^{-13}	to 1×10^{-18} m
Strong nuclear	strong interactions e.g. nucleon bonds	1.0 (reference)	to 1×10^{-15} m

Hadrons and leptons

The strong interaction occurs within a class of particles called **hadrons**, of which the proton and the neutron are the best examples. The strong nuclear force is responsible for keeping protons and neutrons together in stable nuclei, despite the very obvious electrostatic repulsion that also occurs and the extremely high nuclear density. In a typical atomic nucleus the density of matter is about a billion tonnes per cubic centimetre. The only other place in the universe that such high matter density occurs is within pulsars and neutron stars. The strong nuclear force does not depend on electric charge and, within the confines of the nucleus, has the peculiar property of increasing in strength as the particle separation increases.

The weak interaction occurs between members of a class of particles called **leptons**, of which the electron is the best example. Weak interactions may also occur between lepton and hadron particles and are also independent of electric charge. The weak nuclear force is primarily responsible for slow nuclear processes such as radioactive decay of atoms and seems to control the energy-producing fusion reactions going on in stars. Physicists have also surmised that this force played a vital role in the building up of heavy elements from light nuclei in the early stages of formation of the universe. A typical strong-force interaction takes a trillion trillionth of a second whereas a typical weak-force interaction, such as the decay of a neutron, takes about fifteen minutes.

All knowledge of these nuclear forces has come from high energy physics using very powerful particle accelerator machines. These accelerators have also given us knowledge of the basic particles from which all matter is composed. Today, high energy particle theorists refer to a **standard model**, which summarises the known constituents of matter as well as the interactions between them. Table 29.3, as well as the diagram of Figure 29.13, portrays the links between these force interactions and fundamental particles. Before looking at the nature of these fundamental particles, let us complete the story of the force interactions.

The standard model consists of two parts, a part that is used to explain the weak nuclear interactions, called the **electroweak theory**, and a second part used to explain the strong nuclear interactions, called **quantum chromodynamics**, or QCD. Both parts are historically based on an earlier theory called **quantum electrodynamics**, or QED, which was formulated by Richard Feynman, Julian Schwinger and Sin-itiro Tomonaga in the late 1940s. QED theory explains the hydrogen atom as being stable because the proton and electron are continuously exchanging a photon particle between themselves. It's like two tennis players being considered as connected together while they are hitting the ball backward and forward. The photon is called a gauge **boson** particle and acts as the force carrier providing attraction. QED is called a relativistic quantum theory and was one of the first attempts at combining Planck–Bohr quantum theory with Einsteinian relativity.

Ch29-Walding 4th 25/8/04 1:40 PM Page 668

Table 29.3 THE STANDARD MODEL

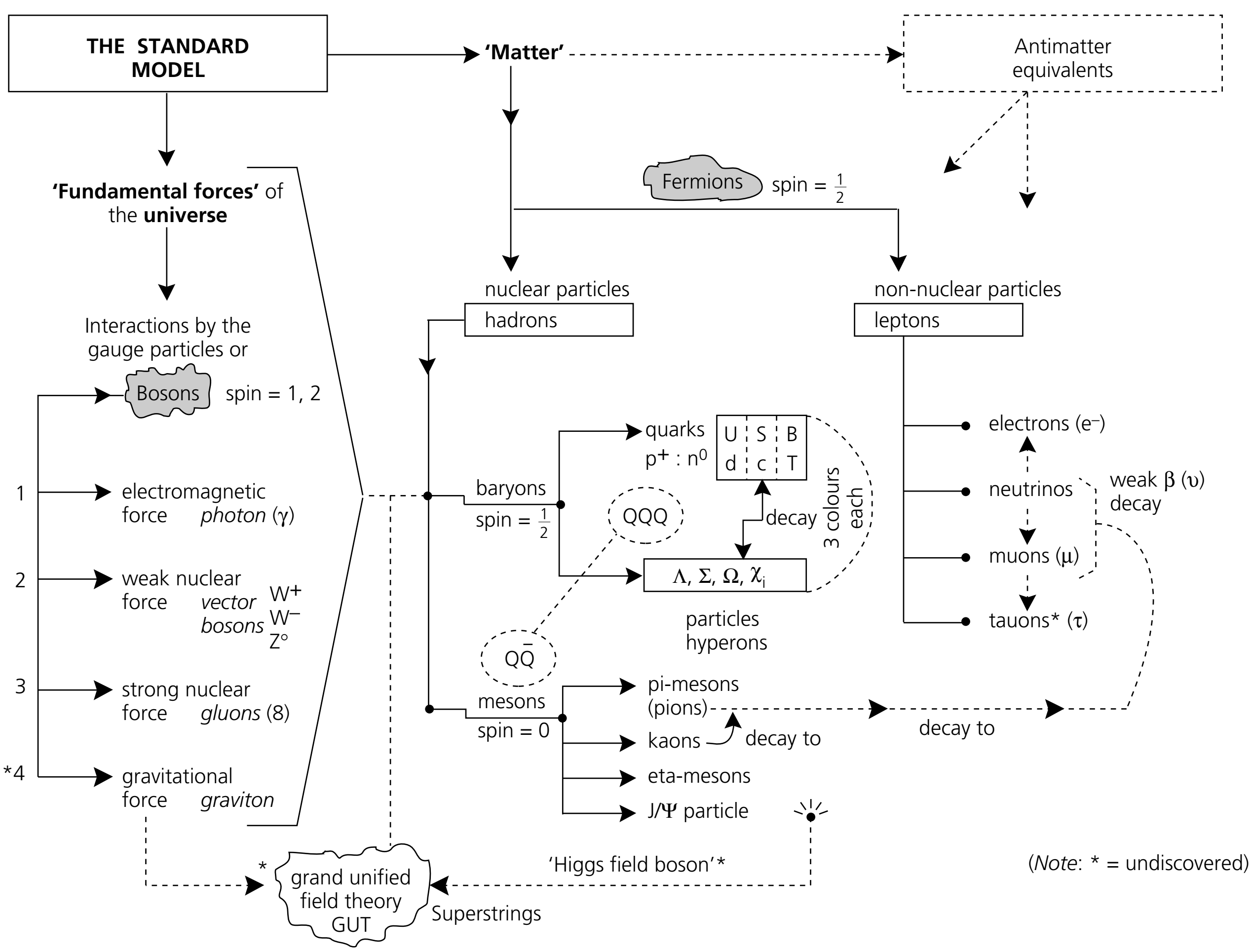
Matter: $m_e = 9.1 \times 10^{-31}$ kg

	First generation Normal matter Symbol	Charge	Mass	Second generation Symbol	Charge	Mass	Third generation Symbol	Charge	Mass
6 QUARKS	Up (u)	+2/3	610 m_e	Charmed (c)	+2/3	2900 m_e	Bottom (b)	+2/3	9800 m_e
	Down (d)	−1/3	610 m_e	Strange (s)	−1/3	300 m_e	Top (t)	−1/3	44 000 m_e
6 LEPTONS	Electron neutrino (ν_e)	0	$\cong 0$	Muon neutrino (ν_μ)	0	< 0.5 m_e	Tauon neutrino (ν_τ)	0	< 68 m_e
	Electron (e)	−1	m_e	Muon (μ)	−1	207 m_e	Tauon (τ)	−1	3491 m_e

GAUGE BOSONS ←	Act on ←	Force interaction
8 Gluons	Nucleons and quarks	Strong nuclear
Intermediate vector Bosons W^+ W^- Z^0	Leptons	Weak nuclear
Photons (γ)	Nucleons and leptons	Electromagnetic
Gravitons *	All particles	Gravitational

* = Not yet discovered

Figure 29.13 Overview of the 'standard model'.



In 1979, the Nobel physics prize was awarded to Steven Weinberg, Abdus Salam and Sheldon Lee Glashow for their work in applying QED to the electroweak theory. In the weak nuclear interaction of radioactive decay, a neutron effectively decays into a proton, an electron and an almost massless neutral particle called an **antineutrino**. The force that leads to the decay of a neutron is very weak. The electroweak theory explains this interaction or breakdown in terms of exchange particles called **intermediate vector bosons** designated as the W^+, W^- and Z^0 particles. These particles are very heavy and were not discovered until 1983 at CERN in Geneva.

The antineutrino involved in weak nuclear decay is characteristic of a complete range of antimatter particles that are now known to exist. In fact, every particle in physics has its own antiparticle, with the same mass but opposite electrical charge. **Antimatter** is very scarce in the universe generally because it has all been annihilated by normal matter during the early formation stages of the universe. (See Figure 29.14.)

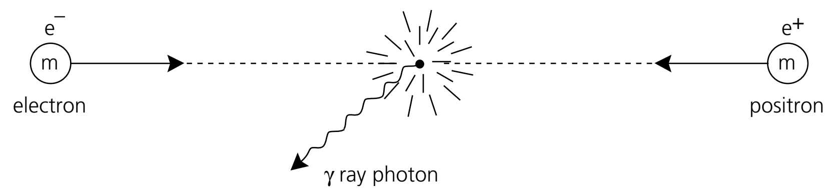


Figure 29.14
Particle–antiparticle annihilation.

The second part of the standard model, called quantum chromodynamics (QCD), attempts to account for the behaviour of theoretical particles, called **quarks** and **gluons**, in forming elementary hadron particles such as protons and neutrons. Again, the theory suggests that the strong nuclear force holding neutrons and protons together is due to the exchange of a force-carrying boson, the gluon, between constituent quarks. The standard model allows for eight gluons and six quarks, although each quark has an associated mathematical property called **colour** charge. The word 'chromo' means colour. An analogy to this is the way that spectral colours can combine to produce light without colour, that is, white. Refer again to Table 29.3 and Figure 29.13.

Fundamental particles

All matter can be considered as divisible into the three major classifications of the standard model, namely **leptons**, **hadrons** and **bosons**. The bosons are unique in that they are their own antiparticles. They are force carriers between other particles. The best known is the electromagnetic photon, while the proposed graviton is yet to be discovered.

The lepton particles are involved in weak interactions as well as electromagnetic and gravitational interactions. The group includes electrons, **muons** and **neutrinos** and are particles that mathematically have spin of 1/2. Quantum theory prescribes that spin angular momentum can only occur in certain discrete values. These discrete values are described in terms of integer or half-integer multiples of the fundamental angular momentum unit, $h/2\pi$, where h is Planck's constant. In general usage, stating that a particle has spin 1/2 means that its spin angular momentum is $1/2\ (h/2\pi)$. Electrons and muons are electrically negative, while the neutrinos are neutral. The six lepton particles occur in 'Flavour' pairs as the:

- electron and the electron neutrino
- muon and the muon neutrino
- tauon and the tauon neutrino.

The word lepton comes from the Greek *leptos* meaning 'small and fine', although the tauon neutrino is nearly 68 times the mass of an electron.

Neutrinos are the most mysterious of the known elementary particles. They were postulated to exist by Wolfgang Pauli in the 1930s, when they were necessary to conserve conservation laws in radioactive β decay. Pauli called them neutrinos, meaning 'little neutral ones'. In 1956 the neutrino was verified in experiments at the Savannah River reactor in the USA by physicists Fred Reines and Clyde Cowan Jr. Because the neutrinos have no charge and

MASSIVE NEUTRINOS

On 5 June 1998, at the Neutrino-98 physics conference at Takayama, Japan, it was announced that the Japanese and American Super-Kamiokande experimental group had detected evidence for the non-zero mass of neutrinos. By studying neutrino interactions in a 50 000 tonne underground tank of purified water, the group had concluded that neutrinos were oscillating between types as they interacted with the water and produced faint light pulses. This would only be possible if they actually had mass. The team concluded that the missing universe dark matter may now be associated with neutrinos.

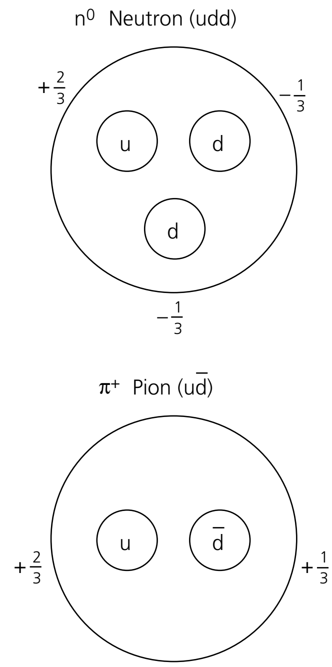
NOVEL CHALLENGE

A columnist in *The Times* (London) newspaper on 11 October 1996 asked: 'What use are quarks; can you eat them?' The distinguished Cambridge metallurgist Sir Alan Cottrell replied, 'I estimate that he eats 500 000 000 000 000 000 000 000 001 each day.' Was Sir Alan correct? Make some rough estimates about food intake and derive your own amount. *Hint:* the mass of a proton or a neutron is 1.67×10^{-27} kg.

negligible mass, they can only be observed by measuring their momentum recoil effects on other particles. They interact very weakly with matter. It has been estimated that solar neutrinos will pass through about 100 light years of water before losing energy. They have no trouble passing through the Earth, for instance. The Sun is a powerful source of natural neutrinos providing an electron neutrino flux density at the Earth's surface of about 6.6×10^{10} cm^{-2} s^{-1}. Let's face it, you are always being literally blasted with neutrinos. It's just as well they appear to do no harm.

The hadron particles may undergo strong nuclear interactions and are subdivided into two classes called **mesons** and **baryons**. Sometimes they may be referred to as nucleons. The most interesting feature of hadrons is that they can be broken down into even more fundamental particles called quarks. Separate quarks do not exist but mesons are quark–antiquark pairs, while baryons are made of three-quark combinations. (See Figure 29.15.)

Figure 29.15
Quark structure of a neutron and a pi-meson.



The mesons have mathematical spin of 0 and are electrically charged, either positive, negative or neutral. Most have masses somewhere between the proton and the electron. Meson comes from the Greek *mesos* meaning 'middle'. They are very short-lived particles. For example, the neutral pion may only last for about 1×10^{-16} s and will decay into leptons.

The baryons have mathematical spin of 1/2. The commonest are the proton and the neutron, with many being more massive than either of these. Baryon comes from the Greek *barus* meaning 'heavy'. The proton is stable and lasts indefinitely, while a free neutron decays in about 15 minutes into a proton, an electron and an antineutrino. Most larger baryons, called **hyperons**, decay into protons and neutrons.

In the early 1960s, American physicists Murray Gell-Mann and George Zweig suggested that hadrons were composed of more fundamental particles called quarks. Gell-Mann coined the word from a phrase in James Joyce's *Finnigan's Wake*, 'three quarks for muster Mark'. These quarks, along with the leptons, constitute the true elementary particles of nature. This concept brought a simpler order to the multitude of particles that had been discovered. The symmetry of six quarks, six leptons and eight bosons was quite simple and allowed all other particles to be classified into the standard model that we have today.

The six quarks are the up, down, strange, charmed, bottom and top; whimsical names introduced to describe various mathematical properties within the theory of their behaviour. An interesting story surrounds the naming of the 'strange' quark. When kaon particles were first discovered in 1947 as a result of cloud chamber studies of cosmic rays, they were noted to do something peculiar with time. Kaons can be created extremely quickly in about one trillion trillionth of a second by colliding protons and neutrons, but once formed they take a considerably longer time of about one nanosecond to decay into pions. This seems to violate the law of time symmetry and reversibility of fundamental physical processes, which generally require formation and decay processes to be opposites. This strange behaviour was subsequently called 'strangeness' when also noted for several other particles. Today, it is known that strangeness is due to the presence of a strange–antistrange quark pair and the decay processes of kaons involve the weak nuclear force in such a way that the time reversibility laws are not violated.

The quark proportional electric charges and masses are tabulated in Table 29.3. For example, a neutron particle contains an up quark, u, and two down quarks, d:

$$\text{neutron (n) udd} = +\frac{2}{3} - \frac{1}{3} - \frac{1}{3} = 0 \text{ or neutral charge}$$

while a positive pion particle contains an up quark, u, and an antidown quark, $\bar{d}$:

$$\text{pion } (\pi^+) \text{ u}\bar{\text{d}} = +\frac{2}{3} + \frac{1}{3} = +1 \text{ or positive charge}$$

The last quark to be discovered, the top quark, was reported only in 1995 from an international team working at FERMILAB, near Chicago, USA.

PHYSICS UPDATE

An international team of physicists has made a batch of 'strange' particles in experiments that could further our understanding of the universe and help with the understanding of collapsed stars called neutron stars, which are thought to contain them.

In 2001, the team created atomic nuclei containing two strange quarks at the Brookhaven National Laboratory in the United States. Since the 1960s only a handful of such particles have been detected and then only in small quantities; but the Brookhaven team of specialists said, 'This is the first experiment to produce large numbers of these doubly strange nuclei'.

The experiment took place within a particle accelerator, where atoms were smashed into their constituent particles, the building blocks of matter producing significant numbers of nuclei containing two strange quarks. (Out of 100 million collisions, 30–40 examples of the doubly strange objects were found).

Normal everyday matter around us consists entirely of only four particles — the electron, the electron neutrino, the up quark and the down quark. These particles are what you consumed for lunch, for instance, so next time someone asks you what you had to eat, tell them! The second generation of particles, the muon, the muon neutrino and the charmed and the strange quarks, are found only in accelerator experiments and in cosmic rays. The third generation particles, discovered since the mid-1970s in very high energy particle accelerators, are thought to describe the state of matter in the early formation stages of the universe.

The dark matter problem

Physicists who study the evolution of the universe (called cosmologists) have calculated that there is a critical density that will determine whether the universe will collapse under its own gravitational attraction (a 'closed' universe) or continue to expand as it currently does (an 'open' universe). When they began looking at the total amount of mass existing in the universe they defined a quantity called omega (Ω). This is the ratio of the measured density of the universe divided by the critical density required for collapse. If omega is greater than 1.0 then the total amount of matter in the universe will eventually cause it to collapse. If omega is less than 1.0 then the universe will continue to expand as it currently does forever. Cosmologists have evidence to suggest that the universe is currently in a state of inflationary balance which suggests that omega equals 1.0.

In actual observations and measurements, modern cosmologists have come across a big problem — the amount of visible matter in the universe produces an omega equal to 0.05 or thereabouts. This means that there is about 95% of the total necessary mass of the universe that is missing! The missing mass has come to be known as '**dark matter**'. A bigger problem is that there also doesn't seem to be enough mass in the observable universe to account for the predictions made by a very successful theory called 'big bang nucleosynthesis' or BBN. This is the current theory which correctly predicts the amounts of hydrogen and helium in the universe and gives the methods for forming larger elements from baryons. If BBN is correct, then the original amount of baryonic matter formed in the Big Bang was about omega = 0.1. So the big question in modern cosmology is: 'What constitutes the universe's missing dark matter?'

Several candidates are being investigated apart from the types of matter that are not easily observable, such as other planets, dim stars, brown dwarfs and exotic particles. The most important of these are the neutrinos which, even if they have a very small mass of about 90 eV, are so numerous as to almost completely account for the missing dark matter. If the universe dark matter is mostly neutrinos, or similar particles, then the dark matter will be termed 'hot dark matter — **HDM**' as the particles are very light, move very fast and will help form large scale structures such as walls, filaments and strings.

The universe dark matter may also be made up of **WIMPS** or 'weakly interacting massive particles' — so called because they are assumed to be extremely massive particles (about 10–100 times the mass of a proton) that do not interact with normal matter very strongly. These particles feel only gravity and the weak nuclear force; and they are impervious to the strong nuclear force and the electromagnetic force (hence we haven't discovered them yet). Such heavy particles would be slow-moving and are known as 'cold dark matter — **CDM**'. This form of dark matter will have assisted in the formation of smaller structures such as galaxies.

Dark matter may in fact it be a mixture of CDM and HDM called 'mixed dark matter — **MDM**'. Cosmologists believe that as the universe grows older, dark matter will become the dominant energy-generation mechanism for the entire universe. Whatever is discovered in the future, dark matter obviously has a lot to do with the way the universe works, and is responsible for the way that the universe is structured.

29.6 THE FUTURE

Beam me up, Scotty!

Australian engineers and physicists are at the forefront of quantum technology. Collaboration between research staff members at the Australian Research Centre (ARC) Quantum Computer Technology facility at the University of Queensland and the Australian National University are at the cutting edge in developing the next generation of computer technology called **quantum computers**, which use particles of light (photons) and fibre optics rather than silicon chip conduction.

Current computer technology is thought to be heading for a 'brick-wall' barrier due to physical semiconductor size constraints and IC chip production costs around the year 2010. Quantum computers will provide a means of overcoming this barrier, by building computers at the level of single electrons and atoms, using principles of quantum physics rather than semiconductor physics and electronics. This technology will dramatically increase the speed and quantity of digital information that can be transmitted over fibre-optic cables. At the heart of the process is the technique of '**teleportation**' which means breaking down an object at one location and reconstructing it at a completely different location. This brings to mind the line 'Beam me up, Scotty' in the famous *Star Trek* series. The researchers have demonstrated models for the teleportation of photons, which is the first step. Teleportation is not only necessary for quantum computing but will also have applications in general communications and security encryption techniques for data (cryptography).

You can see that quantum theory and particle physics have been very intense areas of research since the 1920s. It is remarkable that for physicists and **cosmologists** to understand the universe it is necessary to understand the smallest elementary particles. This is because the elementary particles were formed in the first fragments of time following the Big Bang. In fact, time itself only has meaning following the Big Bang! Understanding how these elementary particles form and interact gives insights into how the universe has evolved and where it is going. Eminent cosmologist Stephen Hawking, who became the Lucasian Professor of Mathematics at the University of Cambridge in 1979, has spent most of his life theorising about the universe, elementary particles and black holes. Hawking's theory combines general relativity and quantum theory into quantum gravity, in which he regards the universe as an expanding entity in which space and time form a closed surface without boundaries. Hawking's cosmology ideas relating to the history of time and the universe have been some of the most important since the original Einstein field equations of general relativity and Edwin Hubble's discovery of the expanding universe. You may have seen Professor Hawking on television speaking through a voice computer and confined to an electric wheelchair.

Today, with data from the cosmic background explorer satellite (COBE) and the Hubble space telescope, physicists are beginning to gain knowledge and understanding about the probable age and fate of the universe. The importance of particle physics to cosmology was in evidence during the famous 1987 A supernova explosion. This was the first exploding star visible to the naked eye for 384 years. Three hours before it was observed with telescopes, two separate underground particle detectors in Painesville, Ohio, USA and Kamioka, Japan detected an influx of neutrino particles that were later analysed.

In physics, a so-called **theory of everything**, or TOE, would provide a complete description of all the forces and particles of nature. In other words, it would contain all parts of the standard model described earlier. It might also explain why the laws of physics are the way they are! It is this question of 'why?' that is so important and not just a description of 'what?'! **Grand unified theories** or GUTs will unify the gravitational, electroweak and strong

PHYSICS UPDATE

Two international teams of astronomers, using NASA's Hubble Space Telescope and ground-based telescopes in Australia and Chile, have discovered the first examples of isolated stellar-mass black holes adrift among the stars in our galaxy.

All previously known stellar black holes have been found in orbit around normal stars, with their presence determined by their effect on the companion star. The two isolated black holes were detected indirectly by the way their extreme gravity bends the light of a more distant star behind them. 'These results suggest that black holes are common and that many massive but normal stars may end their lives as black holes instead of neutron stars,' said David Bennett of the University of Notre Dame.

It has been confirmed recently that a supermassive black hole exists at the centre of our Milky Way galaxy. The black hole has a mass of 3.7 ± 1.5 million solar masses, and the discovery team now believes that a supermassive black hole exists at the centre of every galaxy.

interactions of nature. All this cosmology research is mathematically complex, but there are numerous general interest books available on the subject that give the ideas without the maths. Try to find some!

High energy particle physics is looking forward to 2007 when the exciting large hadron collider (LHC) machine is due to come online at CERN. The LHC will produce head-on collisions between pairs of protons with energies of about 8.0 TeV. Australian research groups are already part of the many projects planned. It is expected that the LHC machine will have enough energy to be able to confirm the existence of the Higgs field boson particle, important in the standard model of matter because it produces spontaneous symmetry breaking and 'allows' normal particles to have mass. Even further into the future, high energy particle physicists are expecting to produce electron-positron colliders to complement the LHC. These machines will really begin to answer 'What next?' type questions.

Questions

11 Calculate the de Broglie wavelength of an electron travelling at 7.5×10^5 m s^{-1} in a cathode ray tube.

12 How does the notion of electron position probability allow us to view the modern atom?

13 Under what conditions is the Heisenberg uncertainty principle important?

14 Explain the differences between leptons, hadrons and bosons.

15 Make notes on the important contribution made to quantum mechanical theory by the following physicists: Wolfgang Pauli, Erwin Schrödinger, Richard Feynman, Murray Gell-Mann, Stephen Hawking.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***16** Give two reasons why Compton scattering provides evidence that supports a particle model for photons.

***17** What is the energy in both joules and electronvolts of a photon whose wavelength is 5.5×10^{-7} m?

***18** The threshold frequency for a particular metal is 2.5×10^{14} Hz. If light of frequency 6.0×10^{14} Hz falls onto the surface, calculate **(a)** the colour of the incident light; **(b)** the incident photon energy; **(c)** the metal's work function; **(d)** the maximum kinetic energy of the photoelectrons; **(e)** the maximum velocity of the photoelectrons.

***19** Calculate the de Broglie wavelength of an electron travelling at 80% the speed of light. Compare this with the diameter of a hydrogen atom.

****20** Use the hydrogen energy level diagram in Figure 29.6 to answer the following:
- **(a)** How much energy must be supplied to raise the atom from quantum state $n = 1$ to $n = 4$?
- **(b)** How much energy is needed to ionise the atom?
- **(c)** What is the frequency of the photon emitted in an electron transition from $n = 5$ to $n = 1$?

***21** Use Einstein's famous equation $E = mc^2$ to determine the energy released when an electron annihilates a positron, each of mass 9.11×10^{-31} kg.

***22** Name the general types of particles that are influenced by the interactions of **(a)** the weak nuclear force; **(b)** the strong nuclear force; **(c)** the gravitational force.

***23** In the quark theory, a normal proton is described as a uud particle. Describe what this means and prove that its electric charge is +1.

***24** Why is the standard model currently regarded as a very good description of the fundamental interactions of nature?

****25** Table 29.4 contains data obtained from a photoelectric experiment. By graphing $E_{K(max)}$ versus frequency, use these data to obtain values for:
(a) Planck's constant in electronvolts; **(b)** the threshold frequency for the metal; **(c)** the work function of the metal.

Table 29.4

$E_{K(max)}$ (eV)	0.5	0.8	1.2	1.75	2.3	2.5
$f(\times 10^{14})$ (Hz)	3.75	4.5	5.5	7.0	8.0	8.9

****26** Produce a quantum mechanical argument as to why it is easier to predict the path of a more massive object, such as a bicycle, rather than a very small object, such as an alpha particle.

****27** The energies of possible quantum states for a gas are listed below. Reorganise these data and represent them on an appropriate energy level diagram. Assume that the ground state energy has been included. Use the diagram to answer the questions that follow:

-8.64×10^{-19} J -5.76×10^{-19} J -16.6×10^{-19} J
-11.5×10^{-19} J -6.72×10^{-19} J

(a) What is the shortest and longest wavelength expected in the emission spectra of this gas under excitation?
(b) How much energy is required to cause the gas atoms to change from energy level 3 to energy level 4?

Extension — complex, challenging and novel

*****28** It is found that a neutron travelling at 1.98×10^4 m s^{-1} has the same energy as a light photon of frequency 5×10^{14} Hz. What is the mass of the neutron?

*****29** In an experiment similar to that of Franck and Hertz, electrons of energy 12 eV are fired into a gas. Electrons penetrating the gas are collected and their energies measured at 12 eV, 1.4 eV and *x* eV. If the spectrum of the light emitted from the gas is also analysed and found to contain photon energies of 11.4 eV, 10.6 eV and *y* eV, deduce the values of *x* and *y*.

*****30** Calculate the de Broglie wavelength of an electron in a cathode ray tube that uses a gun accelerating potential of 750 V.

*****31** The energy levels of a particular type of atom are as follows:

0.00×10^{-19} J 3.36×10^{-19} J 4.96×10^{-19} J
5.76×10^{-19} J 5.92×10^{-19} J 8.16×10^{-19} J.

(a) If atoms in the ground state are bombarded with electrons, what is the minimum energy required to detach an electron from the atom?
(b) If the atoms are bombarded with electrons of energy 5.28×10^{-19} J, what will be the photon energies emitted?
(c) What will be the maximum energy of the scattered electron if a 1.60×10^{-19} J photon is produced in the collision of a 5.58×10^{-19} J bombarding electron with this atom?

*****32** Propose a reason for a head-on collision of two beams of 1 GeV protons being more useful to physicists studying high levels of energy than the collision of a single beam of 2 GeV protons with a fixed target.

*****33** In 1993, the United States Congress voted to stop funding one of the largest particle accelerator projects ever conceived. It was the 20 TeV superconducting supercollider (SSC) to be completed in an 87 km circumference tunnel near Waxahachie, Texas. The reason for the project's abandonment was its projected cost of over $US10 billion. Try to list arguments for why this type of project has both advantages to science and disadvantages to society. How would you vote if you had the chance?

****34** **Max Planck and the 'Deutsche Physik'** The following extract has been taken from the book *Heisenberg probably slept here* by Richard Brennan and published by John Wiley & Sons, NY, 1997. It is the story of Max Planck, one of the most famous German scientists who ever lived. He has been mentioned in this chapter for his discoveries, especially the quantum nature of heat radiation.

Unlike Einstein, Max Planck was caught up in the patriotic frenzy in Germany before the First World War and fully supported Germany's position in what he believed to be a defensive and inevitable war against evil opponents. Planck was the father of two boys of military age and the rector of a university soon to be depopulated by the calling up for military service of both students and the younger instructors. Soon Planck's children were all involved in the war. His girls, Greta and Emma, had trained with the Red Cross and were awaiting assignment to military hospitals. Planck's oldest son, Karl, was at artillery school, and his youngest son, Erwin, was already at the front. 'What a glorious time we are living in', Planck wrote to his sister. 'It is a great feeling to be able to call oneself a German.' How the Plancks ever tolerated their friend Einstein passing out antiwar propaganda on street corners is a mystery. Possibly they considered him a hopeless eccentric. By 1915, the horrors of the First World War became personal for Planck. His nephew, a physicist, his brother's only son, was killed. His own son Erwin had been taken prisoner, and Karl was injured and died of his wounds.

In late 1917 defeat was in the air and the German government was near collapse. But even given all of the tragedy visited on his family and the imminent defeat, Planck refused to sign a proclamation calling for the resignation of the Kaiser, as Einstein had. He was loyal to the end. Despite political differences, Planck's relationship with Einstein remained cordial.

Continuing family tragedies caused Planck great grief. In 1917, his daughter Greta died suddenly a week after giving birth. Her twin sister, Emma, came to Heidelberg to care for the infant, and in 1919 married the widower. By that year's end she too was to die shortly after giving birth. This double tragedy almost destroyed Max Planck. 'There are times now', he wrote, 'when I doubt the value of life itself.'

Planck found solace from public and domestic tragedy both in his work and in helping to raise his grandchildren. His quantum principles were becoming more and more acceptable in the world of science and had expanded into virtually every area of physics. Planck's theorised constant h *came to be regarded as a fundamental constant of nature, the equal of Einstein's* c*, the velocity of light.*

The Nazis and 'Deutsche Physik'

The next period of special note in Planck's life started at the dawn of the Nazi era. In 1930, Planck became president of the Kaiser Wilhelm Society of Berlin, which was then renamed the Max Planck Society. In his seventies at the time, Planck's renown in the world of science was second only to that of Einstein.

The days of Nazi ascendancy in Germany were difficult both for science and for Max Planck personally. The issues were Einstein, because he was a Jew, and the theories of relativity and quantum physics. Anti-Semites (anti-Jews) identified relativity and quantum theories as the decadent work of Jews. In contrast, the right wing extolled the virtues of applied physics, called 'Deutsche Physik', as opposed to contaminated theoretical or Jewish physics. Many German scientists lined up on the Nazi side, and Planck found himself drawn into this ugly fight. The position he took was ambivalent. On the one hand, the major prestigious scientific societies of which he was a leading member remained silent and did not come to Einstein's defense. Privately Planck condemned the Nazi attacks on Einstein as 'scarcely believable filth'. Publicly he tried to stay out of what he called 'political issues'. On the other hand, Planck vigorously defended the theories of relativity. As president of the Society of German Scientists and Physicians, Planck proposed that Einstein be invited to address the annual meeting. Planck hoped that the irrefutable logic of Einstein's science could win the day. Einstein at first accepted the challenge but was forced to withdraw after threats were made on his life. Planck was fighting a losing battle to separate ivory tower science from street politics.

In January of 1933, Adolph Hitler became Reich chancellor and the Nazis were in full power. Max Planck was secretary of the Academy of Science and president of the Kaiser Wilhelm Society, key positions in the scientific establishment in two organisations that depended on the government for financial support. Planck was faced with the choice of either resigning and leaving the country or staying and attempting to moderate Nazi policies. He chose the latter. His hope was to cause compromises for the sake of science, but compromise was not to be had.

Einstein by this time had decided to emigrate to the United States. Letters between the two physicists revealed their separate states of mind with regard to the advisability of compromise with the Nazis, and they would eventually split on this issue. Planck fought long and hard to protect his Jewish students and colleagues, but in the end his efforts could do no more than delay their persecution. Although he never lent his voice and prestige to the Nazi regime in any way, he never stood up firmly or publicly against it. When the Nazis barred all Jewish faculty and students from the universities and Planck remained silent, Einstein broke off their long relationship and never spoke to him again.

Despite the fact that Planck never publicly opposed the Nazi regime, the regime had mixed feelings about Planck. On the one hand, he was a world-renowned scientist, and he and his fame were used in Nazi propaganda efforts. On the other hand, he continued to espouse relativity (even though he ceased using Einstein's name in connection with the theories). This was a typical Planck compromise, for which his reputation suffered abroad. On Planck's eightieth birthday Hitler sent him good wishes, while at the same time Nazi minister for propaganda Joseph Goebbels was trying to prove that he was one-sixteenth Jewish and therefore not fit to lead German science.

Late in 1944, Max Planck's last living child, his beloved son Erwin, was arrested in connection with the plot to kill Hitler. A Nazi court quickly found him guilty, and he was condemned to death. Planck used every political means at his command to save his son. According to one account of what followed, a high Nazi official contacted Planck with a proposed bargain:

Planck would at last join the Nazi party, adding his still-considerable international prestige to their cause. In appreciation, Planck was told, they would seek to commute Erwin's sentence to a prison term. The old man refused. On 23 February 1945, Erwin was executed. Planck was devastated by this loss. To a niece and nephew he wrote, 'He was a precious part of my being. He was my sunshine, my pride, my hope. No words can described what I have lost with him.'

SR

Question: Jewish physics and 'Deutsche Physik' were labels applied by the Nazis to two scientific viewpoints current in Germany in the 1940s. Which one did Planck support and what factors contributed to his position? Was he correct? Support your argument by comparing and contrasting the two types of German science referred to in the article above.

****35** Read the following, about strange new particles, and answer the question that follows.

The tremendous energies available in cosmic rays and particle accelerators led to the discovery of large numbers of additional particles such as kaons and mesons, but also some exotic ones which were referred to as 'strange' particles. Whereas once physicists had only three nuclear particles to deal with, they now had a bewildering array of new particles with a great variety of charges, spins, masses and quantum numbers. There was not the simplicity they had come to expect, so a new theory was needed.

Quarks

A ray of hope came along in the early 1960s with Murray Gell-Mann and George Zweig, who proposed independently that the many hadrons (i.e. baryons and mesons) consisted of smaller particles, which became known as quarks (from James Joyce's novel *Finnegan's Wake*). The quarks, along with the photon and the leptons, would be the true elementary particles.

Three different types of quark were suggested, called 'up' (u), 'down' (d) and 'strange' (s). Later it was found that three more were required; these were known as 'charm' (c), 'top' (t) and 'bottom' (b). All six have antiparticles. Four of the quarks, with some of their properties, are listed in the table below. For charge to work out correctly quarks must all have fractional charges of either

$\pm \frac{2}{3}$ or $\pm \frac{1}{3}$.

A baryon is made up of three quarks and a meson is made up of one quark and one antiquark.

TYPE OF QUARK	SYMBOL	CHARGE	STRANGENESS
up	u	$+\frac{2}{3}e$	0
down	d	$-\frac{1}{3}e$	0
strange	s	$-\frac{1}{3}e$	−1
antistrange	$\bar{s}$	$+\frac{1}{3}e$	+1

Adapted from *Advanced Physics*, J. Murray, 4th Edition by Tom Duncan (1994)

Question: Use the information above to determine the quark structure of each of the fundamental particles in Table 29.5, given that each is composed of two or three quarks. Show all reasoning.

Table 29.5

PARTICLE	SYMBOL	CHARGE	STRANGENESS
pion +	π^+	+1	0
sigma −	Σ^-	−1	−1

Ch30-Walding 4th 25/8/04 1:44 PM Page 679

UNIT 10

Extension Topics

CHAPTER 30

Special and General Relativity

30.1 WHY 'SPECIAL', WHY 'GENERAL'?

Einstein's name is always attached to the **theory of relativity**, yet the work of many famous scientists before him underpins his theory. He questioned the accepted theories of time and motion of earlier nineteenth-century physics and came up with a special theory of his own. People today still ask some of the questions that bothered Einstein:

- Can you travel faster than light?
- Can you travel into the past or into the future?
- If I ran at the speed of light with a mirror in my hand, could I see my own reflection?
- When two rockets are moving relative to each other, can you tell which one is really moving?
- If a torch was moving, wouldn't its light travel faster than if the torch was at rest?
- In *Star Trek*, 'warp speeds' faster than light are equal to $2^n c$, where c is the speed of light and n is the 'warp number'. Can this be true?

In this chapter, we will investigate Einstein's Special Theory of Relativity and later the General Theory.

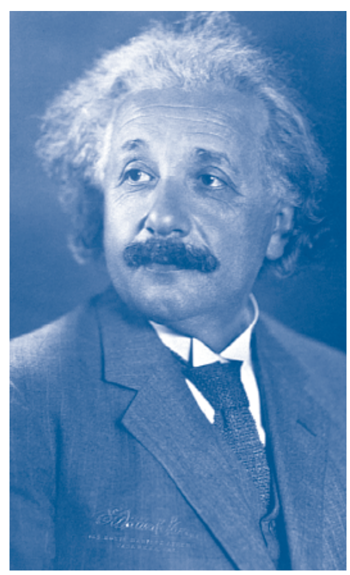
In the beginning

For more than two centuries after its inception, the Newtonian view of the world ruled supreme, to the point that scientists developed an almost blind faith in this theory. And for good reason: there were very few problems that could not be accounted for using this approach. Nevertheless, by the end of the nineteenth century new experimental data began to accumulate that were difficult to explain using Newtonian theory. New theories soon replaced the old ones. In 1884 Lord Kelvin said that there were 'nineteenth-century clouds' hanging over the physics of the time, referring to certain problems that had resisted explanation using the Newtonian approach. Among the problems of the time were the following:

- Light appeared to be a wave, but the medium for its propagation (the 'ether') was undetectable.
- The equations describing electricity and magnetism were inconsistent with Newton's description of space and time.
- The orbit of Mercury didn't quite match the Newtonian calculations.
- Materials at very low temperatures did not behave according to the predictions of Newtonian physics.
- Newtonian physics predicted that a hot object (a black-body radiator) at a stable constant temperature would emit an infinite range of energies — not so!

During the first quarter of the twentieth century, Albert Einstein created revolutionary theories that explained these phenomena. They also completely changed the way we understand nature. To deal with the first two problems he developed the **special theory of relativity** (in 1905). The third item required the introduction of his **general theory of relativity** (1915).

Photo 30.1
Albert Einstein.



The last two items can be understood only through the introduction of a completely new mechanics: quantum mechanics. This chapter deals with special and general relativity. The previous chapter introduced quantum mechanics.

Special Relativity is a deceptively simple theory and has only two assumptions or 'postulates'. They are presented here so you know what is coming, but without any explanation:

- The laws of physics are the same in all uniformly moving reference frames. No preferred frame exists.
- The speed of light in free space has the same value, c, in all uniformly moving reference frames.

Hmm! That doesn't seem too complicated. In fact, most physicists agree that the second postulate is redundant as it is a logical consequence of the first.

General Relativity does away with the restriction of 'uniform motion' which tends to make it more complicated — philosophically, physically and mathematically. In fact, it took Einstein 10 years, with many false starts and wrong turns, from introducing special relativity to the complication of his general theory in its final form. Along the way, the general theory became a whole new way of understanding gravity.

FRAMES OF REFERENCE 30.2

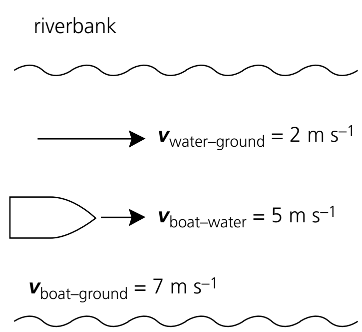


Figure 30.1
The speed of the boat is affected by the speed of the current.

In your earlier work on mechanics, you generally used the ground or Earth as your frame of reference. For example, when a car is going at 60 km h^{-1} along a road, this is with reference to the ground. But when a boat travels down a river, we can state its motion relative to the ground or relative to the water (Figure 30.1). The choice is arbitrary. This notion of reference frames had been discussed at length by Galileo and Newton and we need to begin there.

Inertial frames of reference

This chapter deals with **inertial reference frames** — that is, frames in which Newton's first law (the law of inertia) is valid. If an object experiences no net force due to other bodies, the object either remains at rest or remains in motion with constant velocity (in a straight line). Accelerating frames of reference, rotating or otherwise, are non-inertial frames, and we will not be concerned with them here. The Earth is not quite an inertial frame because it rotates. But it is close enough that for most purposes we can consider it an inertial frame. We could also carry out inertia experiments aboard a ship that is travelling at constant speed. It, too, is an inertial frame.

For Newton, there was a 'master' or absolute inertial frame: a frame stationary relative to absolute space. And any reference frame that is moving at a uniform velocity in a straight line relative to this master inertial frame, he said, will also be an inertial frame. Any reference frame that is accelerating with respect to absolute space, such as the car's frame when the light turns green and the driver accelerates, will not be inertial.

Now imagine that you are riding in the car at, say, 100 km/h down a straight highway and fluffy dice are hanging motionless from the rear view mirror. The principle of inertia is true for you. A second observer is standing beside the highway, watching the car go by. For her the dice are moving in uniform motion in a straight line. So the second observer is also in an inertial frame.

In this case, a good question is: 'Who is moving?' The answer is that you are moving relative to the observer beside the highway, but the observer beside the highway is moving relative to you. So you are both moving relative to each other. Both your inertial frame and her inertial frame are equally 'valid'. This realisation is often called 'Galilean relativity'.

An old favourite to illustrate this further is a cannonball dropped from the mast of a boat sailing along past an observer on the shore (Figure 30.2). For a sailor on the ship the cannonball appears to fall straight down (Figure 30.2**(a)**). From the point of view of an observer on shore, the ball falls with a uniform acceleration downwards while moving with constant speed in the horizontal direction — that is, it follows a parabolic path relative to the shore just like a rock thrown horizontally off a cliff (Figure 30.2**(b)**). However, for both observers the cannonball lands at the base of the mast, and the laws of inertia are the same in both reference frames although the paths are different. We can say:

> A reference frame that moves with constant velocity with respect to an inertial reference frame is itself also an inertial reference frame.

However, in frames moving relative to each other, the velocity of an object will appear different.

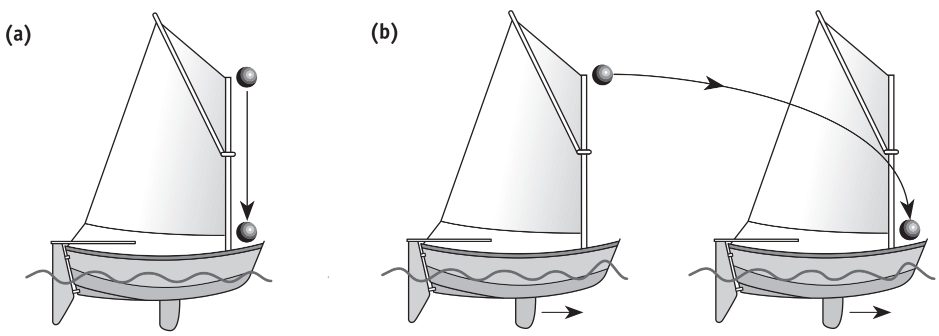


Figure 30.2
A falling cannonball travels different paths depending on your frame of reference: **(a)** from aboard the boat; **(b)** from the shore as the boat travels past you.

SR Activity 30.1 A 'GEDANKEN' (THOUGHT) EXPERIMENT

Before you read any further, you should sort out these questions (well, except for **(f)**):

(a) What would the path in Figure 30.2**(b)** look like if gravity was **(i)** less than that on Earth, **(ii)** more than that on Earth, **(iii)** zero?

(b) How would the path in Figure 30.2**(b)** differ if the cannonball was half the original mass?

(c) If the mast was 20 m high, and the boat sailed at 20 m s^{-1} relative to the shoreline, how many seconds would the cannonball take to hit the deck in Figure 30.2**(a)** and in Figure 30.2**(b)**?

(d) How far would the boat have travelled to the right in this time?

(e) Relative to the shore, what is the displacement and average velocity of the cannonball in its journey shown in Figure 30.2**(b)**?

(f) A very difficult one! How far would the cannonball have travelled in Figure 30.2**(b)** relative to the shore line? You will need to work out the 'arc length' of the parabola. How's your calculus?

Not all things change when viewed in different reference frames. For example, the number of atoms in an object doesn't change. If you time your pulse rate on Earth as 72 per minute, then you'll time it as 72 per minute aboard a moving bus. But if you are sitting down on the bus as it travels along a road at 5 m s^{-1}, you could say your speed is zero relative to the bus ($\boldsymbol{v}_{\text{person–bus}}$ = 0 m s^{-1}), and the speed of the bus relative to the Earth ($\boldsymbol{v}_{\text{bus–Earth}}$) = 5 m s^{-1}. Your speed relative to the Earth ($v_{\text{person–Earth}}$) would then also be 5 m s^{-1}. However, if you walk inside a bus towards the front with a speed of 1 m s^{-1}, and the bus moves at 5 m s^{-1} with respect to the Earth, then your speed is 6 m s^{-1} with respect to the Earth. (See Figure 30.3.)

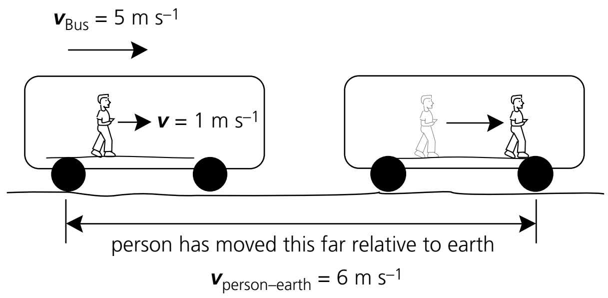


Figure 30.3
At ordinary speeds, the addition of relative velocities is quite straightforward.

Mathematically we can set this out using the Newtonian relativity equation as:

$$\begin{aligned} v_{\text{person–Earth}} &= v_{\text{person–bus}} + v_{\text{bus–Earth}} \\ 6 \text{ m s}^{-1} &= 1 \text{ m s}^{-1} + 5 \text{ m s}^{-1} \end{aligned}$$

You may be more familiar with the equation in the following form where the Earth is the assumed frame of reference and left out of the subscripts. The answer is the same.

$$\begin{aligned} v_{ab} &= v_a - v_b \\ v_{\text{person–bus}} &= v_{\text{person}} - v_{\text{bus}} \\ v_{\text{person}} &= v_{\text{person–bus}} + v_{\text{bus}} \\ &= 1 \text{ m s}^{-1} + 5 \text{ m s}^{-1} \\ &= 6 \text{ m s}^{-1} \end{aligned}$$

SR Activity 30.2 WHO'S REALLY MOVING?

Next time you are on a bus or car that is moving, try this 'thought' experiment. Imagine that you are stationary and it's the Earth that is moving. If so, why do your wheels turn? And why do the wheels of cars beside you turn? It might seem dumb but this is the very same question Einstein pondered 100 years ago. He called it a 'Gedanken experiment' (*Gedanken* is German for 'thought').

THE NATURE OF LIGHT 30.3

NOVEL CHALLENGE

Can a shadow travel faster than light? In the late afternoon, shadows are long and when you stand up your shadow shoots out along the ground much faster than you rose up — well over twice as fast. If you could fire a projectile upwards at 0.8*c* late in the afternoon then its shadow would scoot across the ground at 1.6*c*. *What is the problem here?*

As you may have seen in earlier chapters, changing magnetic fields produce electricity; conversely, changing electric fields produce magnetism. In the mid-nineteenth century, the great Scottish physicist James Clerk Maxwell deduced that as each field could create the other a 'wondrous new phenomena' would result. You get the idea: once a changing field of one type appears, self-perpetuating systems of electric and magnetic fields take on an independent existence, no longer associated with what started them, and would propagate through space as an electromagnetic wave. Maxwell explored the properties of these waves theoretically and calculated their speed as 3.00×10^8 m s^{-1}, and equal to the speed of light. The symbol '*c*' was chosen to represent the speed of light. It was the initial letter in the Latin word *celeritas* meaning 'swiftness' (as in ac<u>celer</u>ate). That speed had been solidly established to an accuracy of a few per cent. So there was no doubt that the speed of Maxwell's electromagnetic waves was the speed of light and his brilliant conclusion was inescapable: light is an electromagnetic wave.

However, a question arose: in which frame of reference did light have this speed? It was thought that light would have a different speed in different frames of reference. For example, if a rocket ship was leaving Earth at half the speed of light (0.5*c*) and shone a beam of light forward at 1.0*c* relative to the spaceship, then the speed of light relative to observers on Earth would be 1.5*c*:

$$v_{\text{light–Earth}} = v_{\text{light–rocket}} + v_{\text{rocket–Earth}}$$
$$1.5c = 1.0c + 0.5c$$

Likewise, when you drive along a road with your headlights on, the light should travel more quickly than if your car was stationary. Something should start to sound a bit odd by now. Light just doesn't do these things!

Maxwell's equations made no provision for a frame of reference. They just said that the speed of light was c (3.00×10^8 m s^{-1}). So physicists thought there must be some special frame of reference where light had this value, and this frame would be the 'absolute' frame by which all other things could be measured. This was a problem for physicists at the beginning of the twentieth century.

Measuring the speed of light

Early attempts to measure the speed of light failed because light moves so quickly. Galileo attempted to measure the speed of light in the early part of the seventeenth century by measuring the time lag between one observer turning on a lamp and another observer noting this and turning on a second lamp on a distant hill several kilometres away. The method failed since the reaction time of the observers exceeded the time (10^{-5} s) that it took the light to travel the distance.

The Danish astronomer Olaus Römer made the first real measurement of the speed of light by using astronomical rather than terrestrial distances (Figure 30.4). By studying eclipses of the moons of Jupiter he was able to measure the speed of light as 2.26×10^8 m s^{-1} — about 75% of the value accepted today. Modern technology has enabled scientists to use Römer's method and obtain a value of 3.0×10^8 m s^{-1} for the speed of light. Albert Michelson, a Prussian-born American physicist, used rotating mirrors in 1926, and by reflecting light between two mountains 35.4 km apart, he was able to measure the speed of light as $2.997\,96 \times 10^8$ m s^{-1}.

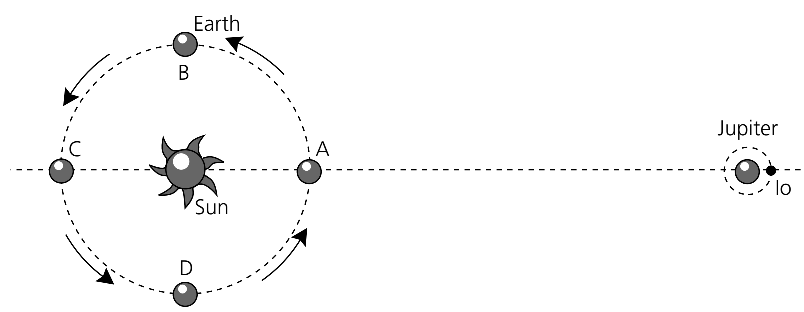


Figure 30.4
Römer's method of measuring the speed of light gave the first reliable value. The eclipse of Jupiter's moon Io occurs 16.6 minutes later than expected when seen from position C. This is the time it takes light to travel the diameter of the Earth's orbit A to C.

The accepted value for the speed of light is now $2.997\,925 \times 10^8$ m s^{-1} but for the purposes of simple calculations in this book we shall take the value to be 3.0×10^8 m s^{-1}. This is a distance of 300 000 km every second, about seven and a half times around the equator of the Earth. It takes about 5 microseconds for light to travel from Brisbane to Cairns.

NEI Activity 30.3 IS LIGHT SLOWING DOWN?

Use your library, CD-ROM or whatever means you can, to find out about and report on one of the following:

1. Some people believe that the speed of light is slowing down. Is there any evidence for this from published historical data or is it just that accuracy has been improving over the years? Use data to support your case.
2. Other people besides Römer and Michelson have tried to measure the speed of light. Report on one of these attempts, using drawings to illustrate how the process worked.

ABSOLUTE FRAMES OF REFERENCE 30.4

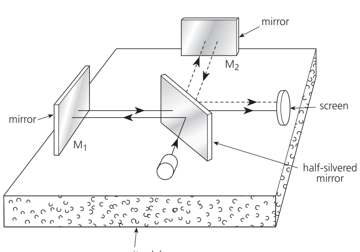
Nineteenth-century physicists were familiar with the properties of water waves, sound waves and waves on springs. These waves all needed some medium for their propagation, be it water, air or steel. There was no reason to think that Maxwell's electromagnetic waves should be any different. They called this transparent medium the *ether* and assumed it permeated all space. But don't think of this ether as the organic liquid used in chemistry. It came from the Greek *aithein* meaning 'to burn', referring to the invisible vapour given off by fires. By 'ether' the physicists referred to some mysterious fluid that filled the universe. The medium for light waves could not be air, since light travels from the Sun to Earth through nearly empty space.

It was therefore presumed that the velocity of light given by Maxwell's equations must be with respect to this 'ether'. Scientists soon set out to determine the speed of the Earth relative to this absolute frame, whatever it might be.

Michelson–Morley experiment

Photo 30.2
The American physicist A. A. Michelson refuted the 'ether wind' theory.

Figure 30.5
Light path in the Michelson–Morley experiment.



In 1887, two American scientists, A. A. Michelson (1852–1931) and E. W. Morley (1838–1923), were concerned that there was no experimental proof of the ether, which was supposed to be the absolute frame of reference for light. They argued that as the Earth moves around the Sun, it should be moving through an ether 'wind'. If there was an ether, and a beam of light was shone in the same direction as the Earth's movement through it, the velocity of light would be measured as greater compared to light travelling at right angles to the direction of motion of the Earth through the ether. Similarly, if light was shone in the opposite direction to the Earth's movement through the ether, the velocity should be measured as smaller. That seems logical!

However, light travels so rapidly that direct measurement of its speed was not possible at the time. But they were able to use an **interferometer** to measure the difference in the speed of light travelling with, against or across the ether. The path diagram for the interferometer is shown in Figure 30.5. It was a huge instrument consisting of a source of coherent (single-wavelength) light and some mirrors on a platform screwed to a massive stone block floating in a pool of mercury so that it could be rotated. This idea comes from lighthouses which also had their rotating lights floating in mercury. In fact, we have some mercury at school which came from the Cape Moreton lighthouse when the old light was removed.

Light was shone from the position marked 'source' where it struck a half-silvered mirror. (This is just a mirror with a half-thickness layer of silver over the whole of the glass so that half the light gets reflected and half passes through.) The reflected beam travelled to mirror M1, bounced off and travelled back to the half-silvered mirror where a lot of it passed through and onto a screen, shown at the right-hand side of Figure 30.5. The other half of the beam from the source initially passed straight through the first mirror where it struck mirror M_2. It also reflected back to the half-silvered mirror where a lot of it was reflected onto the screen.

The positions of mirrors M1 and M2 could be adjusted somewhat so that the path lengths for both beams from the source to the screen were equal (about 11 m). This meant constructive interference would occur and you would get a nice strong image on the screen. If the speed of light was different as it travelled the two paths, there would be a slight time delay in one of the beams and the constructive interference would be upset; the waves would be slightly out of phase and you would get a dimmer image (called a 'fringe shift'). They tried everything. They rotated it left and right; they did it morning and afternoon, in summer and in winter, yet there was no fringe shift.

Trying to explain this 'null' result was going to be difficult for physicists. Some suggested that the ether was really there but it was at rest with respect to the Earth because it was dragged along with the Earth; in this case the fringe shift would be zero. But this was quickly dismissed as fanciful nonsense after some experiments with high-flying balloons were undertaken.

In the 1890s, physicists G. F. Fitzgerald and H. A Lorentz argued that any length — including the stone slab on which the Michelson–Morley interferometer sat — would contract by a factor $\sqrt{1 - \boldsymbol{v}^2/c^2}$ in the direction of motion through the ether. This sounds like they just made it up to explain the null result, but Lorentz argued that the contraction — known as the Fitzgerald Contraction — could happen because the atoms moved closer together as a result of the ether upsetting the electrical interatomic forces. It was a good start, but more justification for his hypothesis was required.

In 1893 Michelson became head of the physics department of the University of Chicago. Later, in 1920, he did research work at the California Institute of Technology and the Mt Wilson Observatory and in 1907 he was awarded the Nobel prize for physics for his work on optical instruments and spectroscopic and meteorological investigations. He was the first American citizen to win this prize.

All of the theories attempting to explain the Michelson–Morley 'null' result were eventually replaced by the far more comprehensive theory proposed by Albert Einstein in 1905 — the special theory of relativity.

30.5 THE SPECIAL THEORY OF RELATIVITY

Einstein was eight years old when Michelson and Morley carried out their famous 1887 experiment. By 1905 he was a young father 26 years old, devoted to his family, to his work at the Swiss Patent Office, and to his physics. In this year he produced six scientific papers, all of which stood out as seminal works in the history of physics. The first one was on quantum mechanics, for which he received the 1921 Nobel prize. The second and third papers were on the size of molecules, for which he received his PhD at the Zurich Polytechnic. The fourth paper introduced the world to the famous formula $E = mc^2$, and the fifth paper dealt with relativity. It was stunning in its simplicity and ingenious, penetrating insight. It resolved completely the contradictions posed by Michelson and Morley.

What motivated Einstein for this paper were what he called '*Gedanken*' (thinking) experiments like 'If I rode on a light beam, what would I see? Would I see light with a speed of zero?' He concluded that absolute space doesn't exist. Einstein's resolution was radical but profoundly simple. It can be stated in one brief sentence, called the Principle of Relativity:

> The laws of physics are the same in all uniformly moving reference frames.

That's it! One sentence implying all of Einstein's special theory of relativity. Historically, Einstein presented two postulates. The second one asserted that the speed of light is the same in uniformly moving reference frames. A more modern approach shows that the second postulate follows from the first and, in fact, by 1910 physicists had shown rigorously that the second postulate is superfluous.

Ch30-Walding 4th 25/8/04 1:44 PM Page 687

In the following section we will examine some interesting consequences of Einstein's theory, particularly the 'invariance' (no variation) in the speed of light, even for observers who are moving relative to each other — and that's so troubling that it will lead to a radical revision of your commonsense notions of space and time. Students get particularly unsettled when commonsense appears to be thrown out the window. But remember, science is not about commonsense; most of the major developments in science appear to be intuitively wrong at first (for example, the Earth revolves around the Sun). As Nobel laureate Richard Feynman said, 'We never really understand physics, we just get used to it.' So, get used to the following!

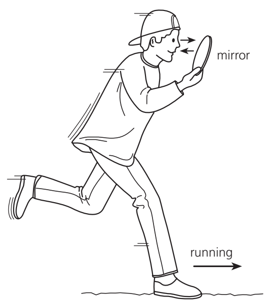
Figure 30.6
The motion of light when you are at rest.

SR Activity 30.4 ANOTHER 'GEDANKEN' EXPERIMENT

When Einstein was a boy he wondered about the following question: a runner holds a mirror at arm's length in front of his face. Can he see himself in the mirror if he runs at the speed of light?

When you look at yourself in a mirror, light travels from your face to the mirror and then is reflected back to your eyes (Figure 30.6). Einstein wondered how light could ever get from your face to the mirror if the mirror is travelling away from the light beam at the speed of light. (Figure 30.7.) The light would never catch up to the mirror! He soon realised the flaw in the logic. Can you? Propose your reasons.

Figure 30.7
Face and mirror moving at speed *c*. Can you see yourself now?



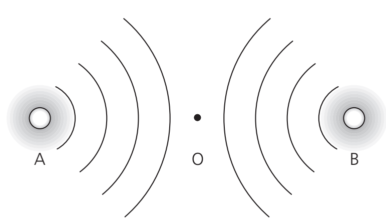
THE MEANING OF 'SIMULTANEOUS' 30.6

Imagine that at your school there are two bells, one at each end of the school. You hear both bells sound at the same time. But could there be a situation where an observer hears one bell before the other? In other words, can an event (sounding of the bells) be simultaneous to one observer but not to the other?

What does simultaneous mean? **Two events are simultaneous if they occur at the same time.** But how can you tell if two bells rang at the same time? If the bells were side by side there would be no problem; but when events are separated in space it gets difficult. If you were midway between the two bells and you heard them ring at the same time, you could say they were simultaneous. But what if you were closer to one bell than the other? If you still heard them at the same time, the more distant one must have sounded first because the sound had to travel further to your ears.

Does this apply to light as well? Say you were looking out your window at dusk and two street lights came on at the same time. You would say the events were simultaneous if you were midway between them (Figure 30.8). If you were not midway you would have to calculate the time it took to get from each event to your position so that you could work out when the events actually occurred. If both lights appeared to turn on at the same time but one was closer to you than the other, the closer one must have occurred after the more distant one. They were not simultaneous. Simultaneity can be defined thus: **Two events are simultaneous if light signals from the events reach an observer who is midway between them at the same time.** So the logic is the same for light as it was for sound.

Figure 30.8
A moment after street lights turn on at A and B, light waves travel outwards. If they arrive at observer O at the same time, she can say they are simultaneous because she is midway between the two lights.



The relativity of simultaneity

To show that two events that are simultaneous in a frame S are not simultaneous in another frame S′ moving relative to S, we will use an example introduced by Einstein.

Imagine a train moving past an observer sitting on an embankment at the side of the track. The train is the moving frame of reference and the embankment is the stationary frame.

Imagine that in the centre of the train carriage there is a person holding a device that can send out a beam of light in the forward direction and at the same time a beam of light in the backward direction (Figure 30.9). Also imagine that the front door and back door are opened automatically by these light beams.

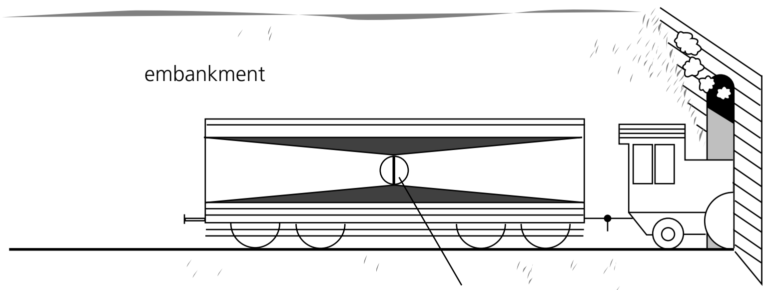


Figure 30.9
The train with the light pulse device.

To the person holding the device, the doors of the carriage will open simultaneously (Figure 30.10**(a)**). But to a person on the embankment, the back door will open before the front door (Figure 30.10**(b)**). This is because the stationary observer sees the back door move forward to meet the light pulse while the front door moves away from the light pulse. So the light gets to the back door before the other light can get to the front door.

Hence, the opening of the doors may be simultaneous to one observer (on the train) but not simultaneous to another (on the embankment). Students often say, 'Who is right? Do the doors really open together or not?' The answer is, 'Both are right.' It depends on your frame of reference. Remember, there is no best frame of reference; some are just more useful than others. You, as an observer, will decide which is the most useful frame.

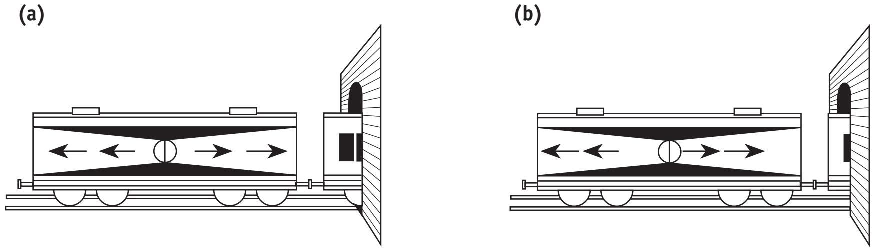


Figure 30.10
(a) Motion of the light pulse as seen by an observer inside the train;
(b) motion of the light pulse as seen by the observer on the embankment — it gets to the rear door first.

30.7 THE VARIABILITY OF TIME

We usually think of time marching on, oblivious to anything we may be doing. Although you may think time drags when you are doing something boring and goes fast when you're having fun, this is not time in a technical sense, only psychological time.

We're now going to convince you that the time interval between two events cannot be the same for two observers in motion with respect to each other. Imagine a bus that has a light source on the floor and a mirror directly above on the ceiling. A brief flash leaves the source and travels upwards to hit the mirror, reflect and return to the source (Figure 30.11**(a)**).

Consider how this looks to an observer sitting at a bus stop at the side of the road. The flash occurs when the bus is located to the left, strikes the ceiling mirror when it is in the centremost position, and returns to the source when the bus is towards the right (Figure 30.11**(b)**).

The labelling of the diagram is as follows. The distance from the source to the ceiling mirror is D. To the roadside observer the bus is travelling to the right at velocity $\boldsymbol{v}$, and moves a distance L in the time between the flash and the light striking the ceiling. The bus moves another distance L by the time the light goes from the mirror back to the source. This makes the total distance moved by the bus equal to $2L$. Light, of course, travels at a speed c for both observers (Einstein's second postulate).

Ch30-Walding 4th 25/8/04 1:44 PM Page 689

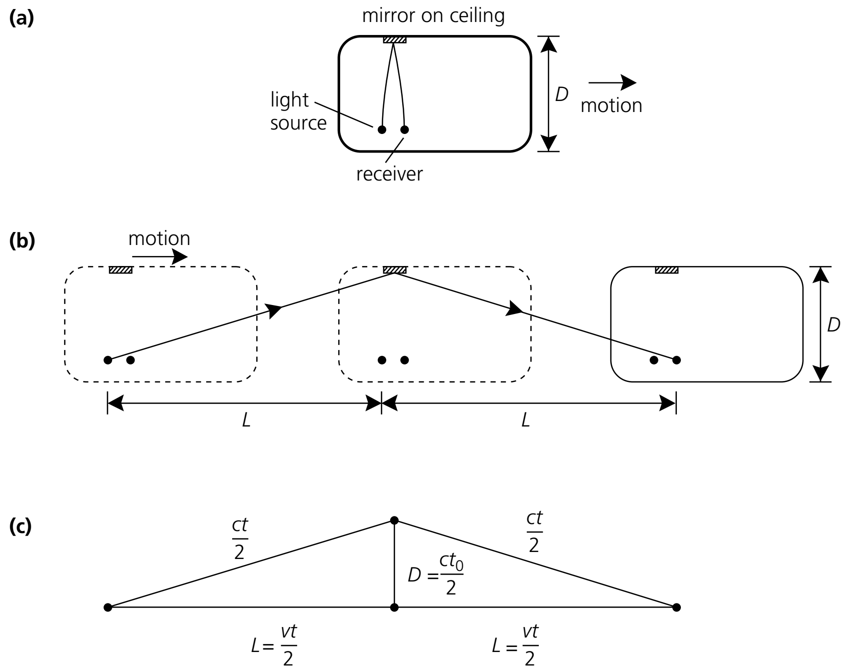


Figure 30.11
(a) Path of light as viewed by the observer aboard the bus; **(b)** path of light as viewed by the observer at the bus-stop — the path looks much longer; **(c)** derivation of Einstein's formula.

To the observer aboard the bus, the light travels a distance of $2D$ (up plus down) in a time t_0. This time is called **proper time** because the start and finish **occur in the same place in space.** Using our velocity formula $\boldsymbol{v} = s/t$, $t = s/\boldsymbol{v}$, hence $t_0 = 2D/c$. When rearranged, the distance $D = ct_0/2$.

To the observer at the bus stop, the light has travelled a triangular path in time t (t with no subscript, as distinct from proper time, t_0). As the speed of light is the same for both observers, the light actually travelled a longer path from the observer at the bus stop's viewpoint, so it must have taken a longer time. This is called 'relativistic' time or 'dilated' time (Latin *dilatare* = 'to spread out'), that is, to bring them away from each other or make them bigger. Dilated time is the time between two events that occur in two different places in space; in this case the two events (the flash leaving the source and the flash arriving back at the source) are separated by a distance of $2L$.

Looking at Figure 30.11**(c)**, the total distance travelled by the light is calculated by using the formula $\boldsymbol{v} = s/t$, or $s = \boldsymbol{v}t$. In this case, the distance travelled by the light will be ct. In the diagram the length of the hypotenuse will equal $ct/2$. The time taken (t) for the bus to go from start to finish will be given by $t = 2L/\boldsymbol{v}$, hence $L = \boldsymbol{v}t/2$ (the base of each triangle).

Using Pythagoras's theorem on one of the right-angled triangles:

$$\left(\frac{ct}{2}\right)^2 = \left(\frac{ct_0}{2}\right)^2 + \left(\frac{vt}{2}\right)^2$$

$$(ct)^2 = (ct_0)^2 + (vt)^2$$

$$t^2 = t_0^{\,2} + \frac{v^2}{c^2}t^2$$

$$t_0^{\,2} = t^2 - \frac{v^2}{c^2}t^2$$

$$= t^2\left(1 - \frac{v^2}{c^2}\right)$$

$$t^2 = \frac{t_0^{\,2}}{1 - \dfrac{v^2}{c^2}}$$

Take square roots of both sides:

$$t = \frac{t_0}{\sqrt{1 - v^2/c^2}}$$

Because v is always less than c, the value of $\sqrt{1 - v^2/c^2}$ must always be less than 1, so we see that $t > t_0$. That is, the time between the two events (the sending of the light, and its arrival at the receiver) is greater for the Earth observer than for the travelling observer. This is a general result of the theory of relativity, and is known as time dilation. Sometimes this is stated simply as **moving clocks are measured to run slow**, but most physicists hate this catchphrase. They say it is utterly meaningless, or worse, since it applies an absoluteness to motion. They prefer the better, but wordier, description of time dilation:

The time between two events is shortest when measured in a reference frame where the two events occur in the same place.

If you want to learn this as 'moving clocks run slow', just remember that you are referring to clocks moving with respect to you. The clocks will be stationary to someone 'moving' along with the clock — that is, to someone in the same reference frame as the clock. Moving is relative! Time really does appear to pass more slowly in a frame of reference moving with respect to you. For example, you could measure the time between two events in a frame moving relative to you as taking 3 seconds. To someone stationary with respect to the clocks, the time between the events may be 2 seconds. They say 2, you say 3, and that seems longer to you. This sounds amazing and is the inevitable outcome of the two postulates of the theory of relativity. Students often ask, 'But is it really true? or does it just appear to be true'. All we can say is that it does not violate the laws of physics and that it has been confirmed by many experiments, so it can be called a 'fact'. However, like all 'facts' in science, they can be replaced if better theories and experiments come along. For now, special relativity works beautifully well, but one day, when you're older ...

We know that the concept of time dilation is hard to accept, for it violates commonsense. We can see from the equation that the time dilation effect is negligible unless v is reasonably close to c. Table 30.1 shows the ratio of v/c (called the speed factor β) for different speeds, and the ratio of t/t_0 (called the **Lorentz factor** γ, after the Dutch physicist H. A. Lorentz who developed the formula before Einstein but didn't realise its significance).

Table 30.1 RELATIVISTIC EFFECTS

SPEED	$\beta = v/c$	$\gamma = t/t_0$
Car at 60 km h^{-1}	0.0000	1.000 000
Jet at 8000 km h^{-1}	0.000 006 8	1.000 000
1 GeV electron	0.999 999 88	2000
20 GeV electron	0.999 999 999 67	40 000
Light pulse	1.000	infinite

Table 30.1 shows that at ordinary speeds (e.g. the car), relativistic effects are negligible, but at speeds approaching the speed of light, the effects are dramatic. An experimenter working with 1 MeV electrons travelling at 0.94 times the speed of light (0.94c) would have to realise that 1 second of time in the electron's frame of reference is the same as 2.9 seconds of time in the laboratory frame of reference. Imagine that the electron spun once on its own axis every second when viewed from a frame of reference stationary to the electron (that is, if you rode along with the electron). To an observer in the laboratory, the electron would take 2.9 seconds to spin once. The electron's 'clock' appears to run slow.

Thousands of experiments have confirmed the theory of relativity. For example, in 1971 extremely accurate clocks were flown around the world on commercial aircraft, and when they were compared to the clocks left back in the laboratory a time dilation effect was confirmed. One of the most famous 'natural' confirmations of relativity is in the 1960s measurement of the lifetime of the unstable elementary particle known as a muon. Muons' 'rest' lifetime (that is, their lifetime as measured by someone at rest to them) is 2.2 microseconds. When they are created in the upper atmosphere by cosmic rays from the Sun (at an altitude of about 5 km) they travel towards Earth at close to the speed of light. Calculations using classical mechanics show that they would be expected to travel $2.2 \times 10^{-6} \times 3 \times 10^{8} = 660$ m before decaying. However, we observe them on the Earth's surface, so their lifetime is longer than expected and in full agreement with the relativity formulas.

Which clock is moving?

The thing that confuses students most of all in this work is defining which is the moving clock and which is the stationary clock. Einstein said that motion is relative so you could say either is stationary. So which is 't' and which is 't_0' in the formula?

It all depends on what event you are timing. Imagine that you are timing a rocket flight to the moon. There is a clock aboard the rocket for the space travellers and there are synchronised clocks on the Earth and on the Moon. The two events — take-off and landing — are measured by the space travellers with a single clock aboard the spacecraft but the observers on Earth need two clocks for their timing — one on Earth to register the time of take-off and one on the Moon to register the time of landing. We say that the space travellers have measured **proper time**, t_0, because the two events were measured in the one place by one clock at rest with respect to both events. The people on Earth measured **dilated time**, t, or **relativistic time**. Dilated time is longer than proper time.

Example

What will be the mean lifetime of a pion (an elementary particle) as measured in the laboratory if it is travelling $0.669c$ with respect to the laboratory? Its mean lifetime at rest is 3.5×10^{-8} s.

Solution

If an observer were to move along with the pion (the pion would be at rest to this observer), the pion would have a mean life of 3.5×10^{-8} s. This is proper time, t_0, and for elementary particles is sometimes called the **rest life**. To an observer in the laboratory, the pion lives longer because of time dilation:

$$t = \frac{t_0}{\sqrt{1 - v^2/c^2}}$$

$$= \frac{3.5 \times 10^{-8}\ \text{s}}{\sqrt{1 - (0.669)^2}} = \frac{3.5 \times 10^{-8}}{0.743} = 4.71 \times 10^{-8}\ \text{s}$$

Space and time units

A **light-year** (ly) is the distance travelled by light in one year. Numerically it is equal to $3 \times 10^{8} \times 60 \times 60 \times 24 \times 365$ or 9.5×10^{15} m. The distance to the red star Betelgeuse in the constellation Orion is 650 ly while our Sun is only 500 light-seconds away from us.

Space travel

If time dilation means that time slows down, it could be possible to live long enough to travel to distant parts of the universe. For example, say you wanted to travel to the star Rigel which is 800 light-years away. If you could travel at close to the speed of light as measured by someone on Earth, it would take about 800 years as measured by Earth observers for you to return. The original observers would all be dead by then. Let's say your speed was 0.999c. Then the time (t_0) for such a trip, as measured by the astronauts, could be calculated:

$$t = \frac{t_0}{\sqrt{1 - v^2/c^2}}$$

$$t_0 = t\sqrt{1 - v^2/c^2} = 800 \times \sqrt{1 - 0.999^2} = 35.8 \text{ years}$$

Thus you could make the trip and get back to Earth within your lifetime. Students often ask: 'Is it just the clocks aboard the spacecraft that slow down?' The answer is that everything slows down — all life processes slow down — and the astronauts would experience 35.8 years of normal sleeping, eating, working and so on.

NOVEL CHALLENGE

You want to get to a star 100 light-years away and have calculated you could do it in 4.5 years if you travelled at 0.999c. The problem is that the human body can't stand accelerations greater than 5 'g'. How much time, as measured by an Earth observer, would it take you to reach 0.999c from rest at an acceleration of 5 'g'? This next question may be too hard: estimate how much time it would take in the spacecraft's frame of reference.

Questions

1. How fast must a pion be travelling if its rest life is 2.6×10^{-8} s but to a laboratory observer it appears to live for 3.1×10^{-8} s?
2. A beam of a certain type of elementary particle travels at a speed of 2.85×10^8 m s^{-1}. At this speed, the average lifetime is measured to be 2.50×10^{-8} s. What is the particle's lifetime at rest?

30.8 CONTRACTION OF LENGTH

There was a young man named Fisk,
whose fencing was exceedingly brisk.
So fast was his action, the Fitzgerald Contraction,
reduced his rapier to a disk.

In a previous example, you saw that a spaceship can travel a distance of 800 light-years in 35.8 years. You may wonder how it can do this in such a short time if nothing can go faster than light, and even it takes 800 years to get there. The answer is that, although time gets stretched in different reference frames, length gets squeezed. This is called c**ontraction of length**. An example involving a rocket departing Earth for the star Rigel may help.

The Earth–Rigel frame of reference

Imagine that an observer on Earth watches a rocket take off for the star Rigel at a speed of $\boldsymbol{v}$. Both the astronauts and the Earth observer will agree on the speed of the rocket. The Earth and Rigel are at rest to one another so they form a single frame of reference. The distance between Earth and Rigel is 800 light-years and has the symbol L_0. This is called the **rest length** or **proper length** (hence the subscript 0) because it is the length measured by an observer at rest to both the Earth and Rigel.

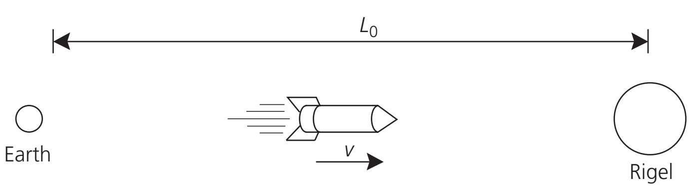


Figure 30.12
As viewed from Earth, the rocket is the moving frame.

The time taken for the journey according to Earth or Rigel observers is **dilated time** (t), because the departure and arrival events are separated in space and require two separate clocks for its measurement.

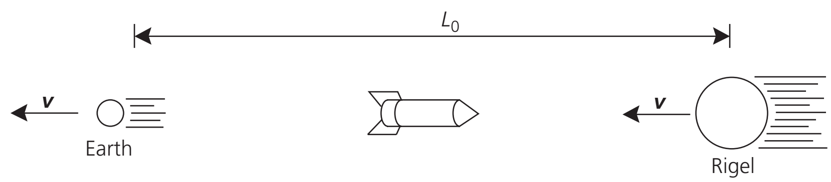
The spaceship frame of reference

Figure 30.13 shows the journey from Earth to Rigel from the astronauts' perspective. They can picture themselves as being stationary and assume the Earth's is rushing away from them and Rigel approaching them. This frame is moving with respect to the astronauts, so they measure the distance between the Earth and Rigel as the **contracted** or **relativistic length** L.

The time taken for the journey as measured by the astronauts is t_0 (proper time), because they are measuring the departure and arrival events in the same place in space (inside their rocket ship) and can use one clock to do this. The space travellers measure t_0, the **proper time**, because the take-off and landing occur at the same place in space for them. Only one clock is needed.

The space travellers and the Earth observers do, however, agree on the relative velocity ($\boldsymbol{v}$) between the two frames of reference.

Figure 30.13
As viewed from the rocket, Earth–Rigel is the moving frame. Either way, the distance is L_0.



Relationships between the frames

The time for the journey is t for the Earth observers and t_0 for the astronauts. The distance is L_0 for the Earth observers and L for the astronauts. They both agree on the velocity of the spaceship as $\boldsymbol{v}$. As we agree that the relationship between t and t_0 is given by: $t_0 = t\sqrt{1 - \boldsymbol{v}^2/c^2}$, the time measured by the astronauts (t_0) is less than that measured by Earth observers (t), hence $t_0 < t$. But as they agree on the velocity of the spaceship ($\boldsymbol{v}$), the distance travelled by the astronauts must also be less than that measured by the Earth observers. In other words, $L < L_0$. We now have two relationships: $\boldsymbol{v} = L_0/t = L/t_0$. If we rearrange the second part we get $L = \boldsymbol{v}t_0$ and if we replace the t_0 with the earlier equation we get this:

$$L = \boldsymbol{v} \times t_0 = \boldsymbol{v} \times t\ \sqrt{1 - \boldsymbol{v}^2/c^2} = L_0\ \sqrt{1 - \boldsymbol{v}^2/c^2}$$

That is:

$$L = L_0\ \sqrt{1 - \boldsymbol{v}^2/c^2}$$

Table 30.2 RELATIONSHIP BETWEEN THE FRAMES

FRAME OF REFERENCE	EARTH–RIGEL	SPACECRAFT
Time for journey	t	t_0
Distance travelled	L_0	L
Velocity	$\boldsymbol{v}$	$\boldsymbol{v}$

Summary of relationships for space travel (Table 30.2): $\boldsymbol{v} = \dfrac{L}{t_0} = \dfrac{L_0}{t}$.

This length contraction applies not only to distances between heavenly bodies but also between atoms — so objects shrink as they speed up. But this contraction occurs only along the direction of motion. For example, if a car travelled forwards at high speed, it would appear to shrink in length (from say 4 m to 2 m) but its height would remain the same at 1.5 m and its width the same at 2 m. If you could run as fast, your height would remain the same but you'd get thinner — but stay just as broad.

Example 1

A spaceship passes you at a speed of 0.80*c*. You measure its length to be 90 m. What length would it be to observers aboard the spaceship?

Solution

- $v = 0.80c$.
- relativistic length $L = 90$ m.
- proper length, $L_0 = ?$

$$L = L_0 \sqrt{(1 - v^2/c^2)}$$

$$L_0 = \frac{L}{\sqrt{(1 - v^2/c^2)}} = \frac{90}{\sqrt{1 - 0.8^2)}} = \frac{90}{0.6} = 150 \text{ m}$$

Example 2

A certain star is 36 light-years away. How many years would it take a spacecraft travelling at 0.98*c* to reach that star from Earth as measured by observers **(a)** on Earth; **(b)** on the spacecraft? **(c)** What is the distance travelled according to observers on the spacecraft? **(d)** What will the spacecraft occupants compute their speed to be from the results of **(b)** and **(c)**?

Solution

- $v = 0.98c$; distance between stars and planets is proper length, $L_0 = 36$ ly.

(a) Observers on Earth measure dilated time, t:

$$v = \frac{L_0}{t} \text{ or } t = \frac{L_0}{v} = \frac{36 \text{ ly}}{0.98c} = 36.73 \text{ y}$$

Note: when a *distance* in light-years (ly) is divided by a *speed* in units of '*c*', the answer is *time* in years (y).

(b) Space travellers measure proper time, t_0:

$$t_0 = t\sqrt{1 - v^2/c^2} = 36.73 \times \sqrt{1 - 0.98^2} = 7.31 \text{ y}$$

(c) Space travellers measure relativistic length, $L = v \times t_0 = 0.98c \times 7.31 \text{ y} = 7.16$ ly.
Alternatively: $L = L_0 \sqrt{(1 - v^2/c^2)} = 36 \text{ ly} \sqrt{(1 - 0.98^2)} = 7.16$ ly.

(d) $v = \frac{L}{t_0} = \frac{7.16 \text{ ly}}{7.31 \text{ y}} = 0.98c$ (same as for Earth observer).

Questions

3 Convert **(a)** 1.8×10^7 m s^{-1} to units of *c*; **(b)** 0.95*c* to m s^{-1}; **(c)** 30 ly to km; **(d)** 3×10^{15} km to ly.

4 An aeroplane whose rest length is 40.0 m is moving at a uniform velocity with respect to Earth at a speed of 630 m s^{-1}. Calculate the length of the aircraft as measured from the Earth.

5 Suppose you decide to travel to a star 85 light-years away. How fast would you have to travel so the distance would only be 20 light-years to you?

6 After the Sun, the nearest star visible to the naked eye is Rigel Centaurus, which is 4.35 light-years away. If a spacecraft was sent there from Earth at a speed of 0.80*c*, how many years would it take to reach that star from Earth as measured by observers **(a)** on Earth; **(b)** on the spacecraft? **(c)** What is the distance travelled according to observers on the spacecraft?

NOVEL CHALLENGE

Imagine you have slid into a parallel universe in which the year is 1840. A US Marshall is travelling by in a train at 0.75*c* and approaches two gunslingers about to have a duel. The gunslingers are both the same distance from the railway line. Gunslinger A is closer to the Marshall.

If the Marshall sees both men draw their guns at the same time, who actually drew first in ***(a)*** *the Marshall's frame of reference;* ***(b)*** *the gunslingers' frame of reference?*

NOVEL CHALLENGE

A physicist driving a very fast sports car is booked by police for travelling through a red traffic light. The physicist argues that because he was travelling fast wih respect to the light, the colour of the light had its wavelength altered and appeared green to him. The judge said that he would let him off the charge of running a red light but would fine him 1 cent for every m/s he was travelling over 100 km/h.

How much was he fined?

Note: the frequency of red light is 4.5×10^{14} Hz and green light, 6.0×10^{14} Hz. The transverse frequency shift formula

$$f_0 = f\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}},$$ is where f_0 = the frequency of the light with respect to an observer in the frame of reference of the source (i.e. the police).

THE TWIN PARADOX 30.9

Not long after Einstein proposed the special theory of relativity, an apparent paradox was pointed out. According to this *twin paradox*, suppose one of a pair of 25-year-old twins takes off in a spaceship travelling at very high speed to a distant star and back again, while the other twin remains on Earth. When the travelling twin returns they have aged differently according to the concept of time dilation (Figure 30.14). The question is: who has aged more? In the examples below we'll assume the travelling twin had an average speed of $0.80c$ for the journey and had aged 30 years aboard the spacecraft before returning.

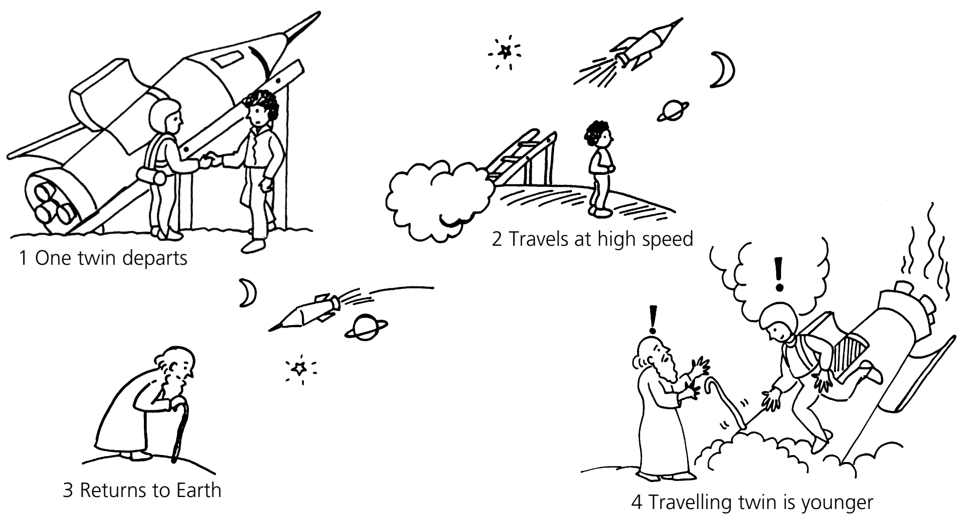


Figure 30.14 The twin paradox — how we can age at different rates.

Scenario 1: Earth's reference frame

According to the Earth twin, the travelling twin will age less.

- Proper time (aboard spacecraft): t_0 = 30 years.
- Velocity: $\boldsymbol{v} = 0.8c$.
- Dilated time:

$$t = \frac{t_0}{\sqrt{1 - \boldsymbol{v}^2/c^2}} = \frac{30}{\sqrt{1 - 0.8^2}} = 50 \text{ years}$$

Hence, the time elapsed on Earth is 50 years; this is how much the Earth twin will have aged. The travelling twin will be 55 years old (25 + 30) whereas the Earth twin will be 75 years old (25 + 50).

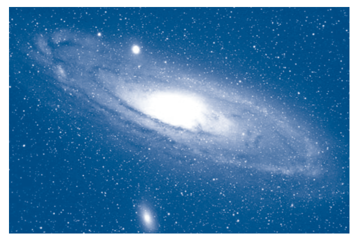
Scenario 2: spaceship's reference frame

Since 'everything is relative', all inertial reference frames are equally good. The travelling twin could make all the claims the Earth twin does, only in reverse. The travelling twin could claim that since the Earth is moving away at high speed, time passes more slowly on Earth and the twin on Earth will age less. In this case, proper time will be that as measured aboard the 'moving' Earth (30 years) and dilated time will be measured by the 'stationary' spacecraft (50 years). This is the opposite of what the Earth twin says. They both can't be right, can they? When the spacecraft returns to Earth they can stand beside one another and compare ages and clocks. Only one of the above scenarios will be correct — but which one?

Resolution

The problem can be resolved by deciding who is travelling with uniform motion. The travelling twin must change velocity at the beginning and end of the trip and also when turning around out in space, so he must be really moving, even if these acceleration periods occupy only a tiny portion of the total time. So the Earthbound twin measures proper length and the travelling twin measures the contracted (shortened) length. But as both twins agree on the relative velocity, the travelling twin must measure a shorter time (to cover the shorter length) and thus returns to Earth having aged less than the Earthbound twin. Even when the acceleration periods are considered, Einstein's general theory of relativity, which deals with accelerating reference frames, confirms this result. The ultimate judge, of course, is experiment, and the 1971 experiment of precise clocks sent around the world in jet planes confirms that less time passes for the traveller.

Photo 30.3
The Andromeda galaxy is 100 light-years away but astronauts would age only 20 years (in their time) if they travelled at 0.98c to get there. Andromeda is the nearest major galaxy to our own Milky Way galaxy, and our galaxy is thought to look much like Andromeda. Together these two galaxies dominate the local group of galaxies. The diffuse light from Andromeda is caused by the hundreds of billions of stars that compose it. The several distinct stars that surround Andromeda's image are actually stars in our galaxy, and are well in front of the background ob



© Jason Ware

Time travel

There was a young man named White,
who wasn't exceedingly bright.
He went out one day in a relative way,
and came back the previous night.

The prospect of travelling into the past or into the future has always excited people. However, it is always discounted as a non-scientific idea. But it's not! It all depends on how you look at events.

In the twin paradox, the space traveller arrived back on Earth as a 50-year-old to meet his Earth twin who was 75 years old. The space traveller had travelled into his twin's future. You can travel into the future — but once you are there, you can't go back.

Likewise, the Earth twin went into his travelling twin's past to see the travelling twin at age 50. You can travel into the past — but only someone else's past.

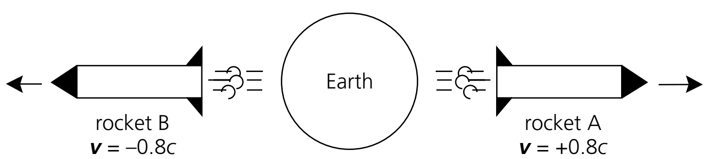
30.10 RELATIVISTIC ADDITION OF VELOCITIES

In the classical view of motion, velocities were vector quantities that could be added or subtracted according to certain simple rules. In modern physics, this notion no longer holds. Consider the following two cases.

Case 1: Anti-parallel tracks

Consider two rocket ships, both leaving Earth in opposite (anti-parallel) directions and both travelling at 0.8*c* relative to the Earth (Figure 30.15).

Figure 30.15
Anti-parallel tracks means moving in opposite directions.



Anti-parallel tracks

The velocity of A relative to B can be calculated in the classical Newtonian method by:

$$\boldsymbol{v}_{AB} = \boldsymbol{v}_{AE} + \boldsymbol{v}_{EB}$$

Or, if rearranged, the relationship becomes:

$$\boldsymbol{v}_{AB} = \boldsymbol{v}_{AE} - \boldsymbol{v}_{BE}$$

It can be simplified by assuming the velocities are with respect to the same stationary frame of reference, in this case the Earth, and omitting the subscript E:

$$\boldsymbol{v}_{AB} = \boldsymbol{v}_A - \boldsymbol{v}_B$$

(this is the basic formula for relative motion from Chapter 2.)
If we substitute the data from the current scenario:

$$\boldsymbol{v}_{AB} = 0.8c - -0.8c = 1.6c$$

This is clearly wrong because no object can travel faster than the speed of light in any reference frame.

Einstein's modification

Einstein showed that since length and time are different in different reference frames, the old addition of velocities formula is no longer valid. Instead, the correct formula, he said, is:

$$\boldsymbol{v}_{AB} = \frac{\boldsymbol{v}_A - \boldsymbol{v}_B}{1 - \boldsymbol{v}_A\boldsymbol{v}_B/c^2}$$

When applied to the current situation:

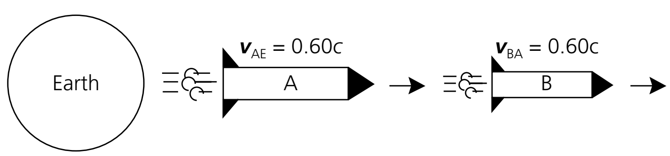
$$\boldsymbol{v}_{AB} = \frac{0.8c - -0.8c}{1 - \dfrac{(0.8c)(-0.8c)}{c^2}} = 0.98c$$

which is less than the speed of light.

Case 2: Parallel tracks

Consider rocket ship A, which travels away from the Earth with velocity $\boldsymbol{v}_{AE} = 0.60c$, and assume that this rocket has fired off a second rocket, B, on a parallel track, which travels at velocity $\boldsymbol{v}_{BA} = 0.60c$ with respect to the first (Figure 30.16).

Figure 30.14
Parallel tracks means moving in the same direction.



We might expect that the velocity $\boldsymbol{v}_{BE}$ (or $\boldsymbol{v}_B$) of rocket B with respect to Earth can be determined by:

$$\boldsymbol{v}_{BA} = +0.60c; \text{ hence } \boldsymbol{v}_{AB} = -0.60c$$

Substituting into $\boldsymbol{v}_{AB} = \boldsymbol{v}_A - \boldsymbol{v}_B$:

$$-0.60c = 0.60c - \boldsymbol{v}_B$$
$$\boldsymbol{v}_B = 1.20c$$

This is again clearly wrong because no object can travel faster than the speed of light in any reference frame.

Using Einstein's relativistic addition of velocities formula:

$$\boldsymbol{v}_{AB} = \frac{\boldsymbol{v}_A - \boldsymbol{v}_B}{1 - \boldsymbol{v}_A\boldsymbol{v}_B/c^2}$$

$$-0.60c = \frac{0.60c - \boldsymbol{v}_B}{1 - \dfrac{(0.60c)(\boldsymbol{v}_B)}{c^2}}$$

$$\boldsymbol{v}_B = +0.88c$$

Questions

7 Spaceships A and B are approaching Earth at velocities of 0.75c and 0.50c respectively. Calculate the velocity of A relative to B if the ships are on **(a)** parallel tracks; **(b)** anti-parallel tracks.

8 The meteorite watch officer on a spaceship reports that two fast micrometeorites are approaching the ship on parallel tracks (i.e. from the same direction), one at a velocity of 0.90c and the other at 0.70c. What is the velocity of either of them with respect to the other?

9 A radioactive lithium nucleus is travelling at 0.70c in an accelerator when it emits a beta particle directly forward at a velocity of 0.80c relative to the nucleus. What is the velocity of the beta particle relative to the laboratory frame of reference?

30.11 MASS AND FORCE

In terms of Newtonian mechanics, if a rocket engine was used to propel a spacecraft through frictionless space, then its speed would continue to increase forever. Newton's second law of motion implies that as long as there is an unbalanced force, acceleration will continue ($\boldsymbol{F} = m\boldsymbol{a}$). For example, if a 10 000 N force was applied to a 100 kg satellite, then its acceleration would be 100 m s^{-2}. To go from rest to the speed of light could be calculated thus:

$$\boldsymbol{a} = \frac{\boldsymbol{v} - u}{t} \text{ or } t = \frac{\boldsymbol{v} - u}{\boldsymbol{a}} = \frac{3 \times 10^8 - 0}{100} = 3 \times 10^6 \text{ s (about 1 month)}$$

If the force was continued for another month then the speed would be twice that of light, and so on. Clearly, this is wrong.

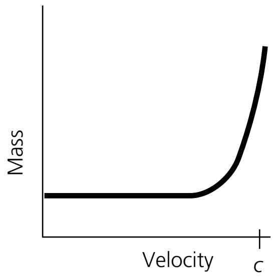
Mass increase

So far, we've seen that time and length are relative, but another relative quantity is mass. In 1909 physicist Hans Bucherer was investigating beta rays (electrons) being emitted by radium. He found that they were being emitted at different velocities but the greater the velocity, the greater the mass. He applied Einstein's relativistic mass formula and found good agreement with the observations. The formula is:

$$m = \frac{m_0}{\sqrt{1 - \boldsymbol{v}^2/c^2}}$$

In applying this formula, Bucherer used the rest mass of the electron for m_0. This is the mass of the electron as measured by an observer at rest to the electron; that is, travelling along with it. The symbol m is used for the relativistic mass; that is, in the frame in which the electron is moving at a velocity $\boldsymbol{v}$. In Bucherer's experiment this was the laboratory.

Figure 30.17
At a speed of c, mass would be infinite.



From the formula it should be obvious that relativistic mass is always greater than rest mass ($m > m_0$). But what may not be as obvious is how mass increases with velocity. A plot of the data produces a graph as shown in Figure 30.17. As velocity increases so does mass, but it increases exponentially; hence, you need more and more force for the same increase in velocity. As velocity approaches the speed of light, mass approaches infinity (as $\mathbf{v} \to c$, $m \to \infty$).

Example

The rest mass of a proton is 1.67×10^{-27} kg. What is the mass of a proton travelling at $\mathbf{v} = 0.865c$?

Solution

$$m = \frac{m_0}{\sqrt{1 - \mathbf{v}^2/c^2}} = \frac{1.67 \times 10^{-27}\ \text{kg}}{\sqrt{1 - 0.865^2}} = \frac{1.67 \times 10^{-27}\ \text{kg}}{0.50} = 3.34 \times 10^{-27}\ \text{kg}$$

(At a velocity of $0.865c$, the mass of an object is double its rest mass.)

Questions

10 What is the mass of a neutral pi-meson $\pi°$ ($m_0 = 2.4 \times 10^{-28}$ kg) travelling at a velocity of $0.87c$?

11 Escape velocity from the Earth is 40 000 km h^{-1}. What would be the increase in mass of a 3.8×10^5 kg spacecraft travelling at that velocity?

The ultimate speed

Perhaps the most astonishing prediction to come out of the special theory of relativity is that there is a certain velocity beyond which nothing can go. The contraction of length formula suggests that, as the velocity of an object approaches c, its length approaches zero. This means that at a speed equal to c the object would disappear. You cannot go faster than light. You just can't! You can't even equal it. This is a fundamental result of the special theory of relativity. The speed of light is a natural speed limit in the universe. As an object is accelerated to greater and greater speeds, you are doing more work on it and giving it more kinetic energy. This extra energy is converted to mass ($E = mc^2$) so its mass becomes larger and larger and, at a speed of c, mass would be infinite and this is impossible. However, Einstein's equations do not rule out the possibility that objects exist whose speed is always greater than c. If such particles exist (the name 'tachyon' — Greek *tachy* = 'fast' — was proposed), the rest mass m_0 would have to be imaginary; in this way the mass m would be the ratio of two imaginary numbers for $v > c$, which is real. Did you do that in maths? For such hypothetical particles, c would be a lower limit of their speed. In spite of extensive searches for tachyons, none have been found. **It seems that the speed of light is the ultimate speed in the universe.**

MASS AND ENERGY 30.12

As you know from Newton's laws of motion, when a net force is applied to a body it accelerates ($\mathbf{F} = m\mathbf{a}$) and so it gets faster. Since the force is acting over a distance, work is done on the object ($W = \mathbf{Fs}$) and its kinetic energy increases as well. In classical mechanics, the work done on an object leads to an increase in speed and kinetic energy, but in relativistic mechanics, the work done on an object also increases its mass.

Hence, the relativistic kinetic energy formula incorporates the notion of relativistic mass. Einstein proposed the following formula:

$$E_k = mc^2 - m_0c^2$$

where m is the relativistic mass of an object travelling at speed v, and m_0 is the rest mass of the same object. Einstein called the first term (mc^2) **total relativistic energy** (or E_{tot}) and the second term (m_0c^2) **rest energy** (or E_0). Hence, kinetic energy is the difference between total energy and rest energy:

$$E_k = E_{tot} - E_0 \text{ or } E_{tot} = E_k + E_0$$

This is a mathematical statement of the principle of conservation of mass–energy: **The total energy (rest energy plus all other forms of energy) in a closed physical system is a constant** (and is equal to mc^2). You can change rest mass into kinetic energy and vice versa. In nuclear reactions or radioactive decay, kinetic energy is produced as mass is lost. In other words, rest mass is converted to kinetic energy as the particles fly away from each other. Einstein's formula $E = mc^2$ is one of the most famous in science. If you know the change in mass of an object, you can calculate how much energy this is equivalent to.

Example 1

(a) How much energy would be released if a pi-meson ($\pi°$) of rest mass of 2.4×10^{-28} kg decayed and its entire rest mass was transformed completely into electromagnetic radiation? (*Hint:* assume it is not moving and has no kinetic energy to contribute.)
(b) How much energy would be released if it was travelling at 0.8c when it decayed? (*Hint:* consider the kinetic energy as well.)

Solution

(a) $E = mc^2 = 2.4 \times 10^{-28} \times (3 \times 10^8)^2 = 2.2 \times 10^{-11}$ J.
(b) The total energy is equal to its relativistic mass (m) times c^2.

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \frac{2.4 \times 10^{-28}}{\sqrt{1 - 0.80^2}} = 4.0 \times 10^{-28} \text{ kg}$$

$$E = mc^2 = 4.0 \times 10^{-28} \text{ kg} \times (3.0 \times 10^8)^2 = 3.6 \times 10^{-11} \text{ J}$$

Note: you could also calculate the rest energy ($m_0c^2 = 2.2 \times 10^{-11}$ J) and add it to the E_k, which equals $(m - m_0)c^2$ or 1.4×10^{-11} J, giving a total the same as above (3.6×10^{-11} J).

Example 2

An electron (rest mass = 9.109×10^{-31} kg) moves with a speed of 0.8c. **(a)** Calculate its kinetic energy. **(b)** Compare this with the Newtonian kinetic energy.

Solution

(a) The mass of the electron at 0.8c is:

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \frac{9.109 \times 10^{-31} \text{ kg}}{\sqrt{1 - 0.8^2}} = 1.5 \times 10^{-30} \text{ kg}$$

Thus its kinetic energy is:

$$\begin{aligned} E_k &= mc^2 - m_0c^2 \\ &= 1.5 \times 10^{-30} \times (3 \times 10^8)^2 - 9.109 \times 10^{-31} \times (3 \times 10^8)^2 \\ &= 5.30 \times 10^{-14} \text{ J} \end{aligned}$$

(b) The Newtonian calculation would give:

$$E_k = \tfrac{1}{2}mv^2 = \tfrac{1}{2} \times 9.109 \times 10^{-31} \times (0.8 \times 3 \times 10^8)^2$$
$$= 2.62 \times 10^{-14}\ \text{J}$$

but this is not the correct formula.

Questions

12 Calculate the kinetic energy of a proton travelling 9.2×10^5 m s^{-1}. The rest mass of a proton is 1.673×10^{-27} kg.

13 What is the kinetic energy of an electron whose mass is 5.0 times its rest mass?

HOW REAL IS RELATIVITY? 30.13

Photo 30.4
Global positioning system: a hand-held unit worth about \$300. The one shown here is the Magellan GPS 310, which can get a precision fix on a location by tracking up to twelve GPS satellites simultaneously, and to an accuracy of 15 m or better.

Because of the complexity of modern scientific theories, an adage has developed among physicists: 'You never really understand new theories — you just get used to them'. This is true of special relativity. Try explaining relativity to someone who knows nothing about it and you'll soon find out that people are quite sceptical. But relativity is included in all sorts of applications these days. When NASA began the Apollo program in the 1960s, they had to figure-in relativity to get their spacecraft trajectories right. Relativity theory works — it is used. It is essential in modern technology.

NEI

Activity 30.5 NAVIGATION SATELLITES

In modern long-range navigation, the precise location and speed of a moving craft are continuously monitored and updated. With a modern system of navigation satellites called NAVSTAR, the location and speed anywhere on Earth can be determined to within about 15 m and 2 cm s^{-1}. However, if relativity effects were not taken into account, the accuracy would be unacceptable to modern navigation systems.

1. Suggest how you would find out if relativity effects are taken into account with the more common global positioning system (GPS) satellite used by small boats and outback adventurers. What would you ask?
2. If you have time, carry out your suggestion, and report back to your class. *Hint:* write, phone, look up a GPS manual or use one of the Internet newsgroups.

GENERAL RELATIVITY 30.14

Since special relativity requires all objects and particles to be limited by the speed of light, all forces and interactions must also travel at or below the speed of light. Newton's gravitational theory is in contradiction with this principle because it states that the gravitational force acts instantaneously. Einstein spent many years attempting to create a gravitational theory that would not require forces to act faster than light. Unlike special relativity, his theory of general relativity — developed in 1915 — allows for a simple treatment of objects moving with non-uniform (accelerated) motion. But like special relativity it begins with another of his simple postulates: 'Inertial mass and gravitational mass are the same.' This is called the **principle of equivalence**. It may not seem to be very profound, as we have never questioned the difference. Let us investigate. You may like to review pages 78 and 79 first.

Ch30-Walding 4th 25/8/04 1:44 PM Page 702

In Newton's theory, two kinds of mass appear:

(a) **inertial mass** from his second law of motion: $F_{net} = m \times a$ or $a = \dfrac{F_{net}}{m_{inertial}}$

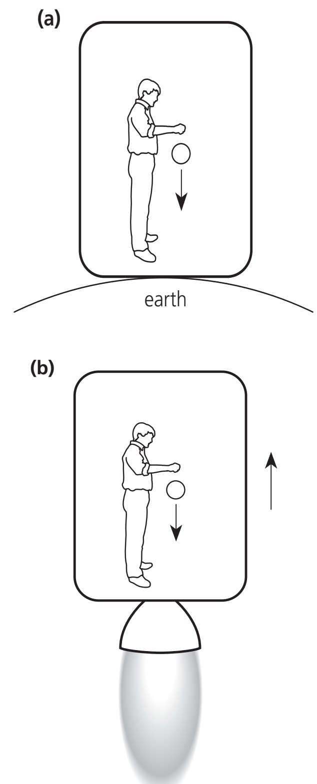
(b) **gravitational mass** from his law of gravity where the force between objects of masses M and m separated by a distance r is $F = GMm/r^2$, where G = the universal gravitational constant. Taking m to be an object near the Earth's surface, with M and r being the Earth's mass and radius, the force on m is:

$$F_{gravitational} = \frac{GM_{Earth}}{r_{Earth}{}^2} \times m = 9.8 \text{ m s}^{-2} \times m_{gravitational}$$

In Newton's theory there is no physical reason why these masses should be related to each other. Why should the pull of gravity on an object be related to the reluctance of an object to accelerate when a net force is applied? Consider what happens if you drop an object from shoulder height. The only force acting on the object is supplied by gravity and hence the net force F_{net} is $F_{gravitational}$. The acceleration is given by $a = F_{net}/m_i$ and when the equation for F_{net} is substituted into it, the acceleration is given by $a = 9.8\ m_g/m_i$. If m_g and m_i were different, the acceleration of an object under the influence of gravity would depend on the inertial mass and thus fall differently. We know this is not true. In fact, the inertial and gravitational masses are equal in value to a very high precision (1 part in 10^{12}), and this is an astonishing mystery in Newton's theory that begs to be explained.

In characteristic fashion, Einstein hypothesised that these two kinds of mass are, in fact, one and the same and he sought to deduce the remarkable consequences of this hypothesis. Einstein was very good at using simple, but profound, physical reasoning to get to the heart of things. His discussion of free-falling elevators provided two good examples.

Figure 30.18
'Einstein's elevator' thought experiment. A stationary observer near Earth experiences similar forces to a person accelerating in a rocket in free space, where gravitational influences are almost zero.



Case 1: On Earth versus an accelerated reference frame in space

Imagine you are in a small closed room on Earth (such as an elevator). You are holding an apple, which you drop, and it accelerates towards the Earth (at 10 m s^{-2}) because of the constant force of gravity acting on it. Now let's attach a rocket motor to the room and travel to a distant region of intergalactic space, so far from any planets or stars that gravity appears to have no effect. The rocket engine propels the room with an acceleration of 10 m s^{-2}. If you hold the apple in your hand it shares your motion and you have the same motion as the room. So the room, you and the apple are all accelerating at the same rate. Now repeat the apple-dropping experiment. As soon as you let go of the apple there is no longer any net force acting on it. So what does it do? It obeys the law of inertia and tries to retain its current velocity. But the room is accelerating and its velocity is increasing relative to the apple so the floor catches up to the apple. To you, it appears that the apple hits the floor. Einstein asked, 'How does this differ from apple-dropping on Earth?' The answer is that it is no different. An unaccelerated frame in Earth's gravity is equivalent to an accelerated frame in the absence of gravity.

Case 2: Free fall on Earth versus an unaccelerated reference frame in space

Imagine you are in an elevator that is at rest relative to the Earth (Figure 30.19**(a)**). The gravitational force on your body, called your **weight** ($\boldsymbol{F}_W$), pulls you down onto the floor of the elevator. However, because you are neither going through the floor nor being thrown into the air it follows that the floor must be pushing up on you with exactly the same force. This is the normal reaction force ($\boldsymbol{F}_N$). You experience this reaction force as your weight. Suddenly disaster strikes; the elevator cables snap and you become weightless as you are in free fall (Figure 30.19**(b)**). The floor of the elevator is accelerating towards Earth as fast as you are (10 m s^{-2}). This has two consequences: the elevator cannot exert any force on your body and you remain at rest relative to the elevator. You take an apple from your pocket and let go of it. The apple appears to remain suspended in thin air. Again, this is because the apple is in free fall and is accelerating towards the Earth at the same rate as the elevator. Einstein asked, 'How different is this from being inside an elevator drifting around in outer space?' (Figure 30.19**(c)**). Einstein postulated that all physical phenomena occur exactly the same way in a frame (i.e. the elevator) accelerating in gravitational free fall as they do in a frame without gravity. Within the confines of the room, no experiment could help you decide which situation you were in. Nor could any other experiment involving the laws of motion.

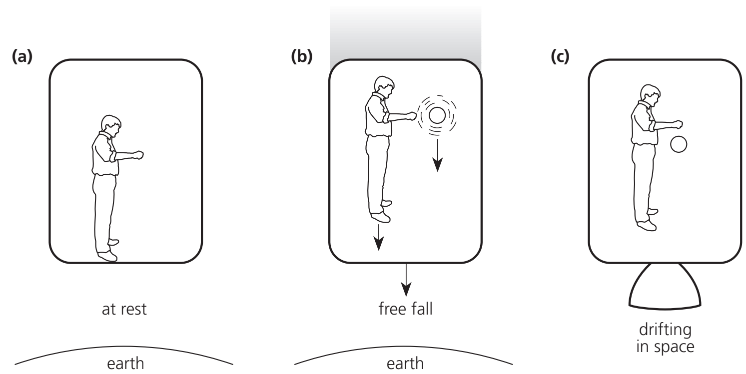
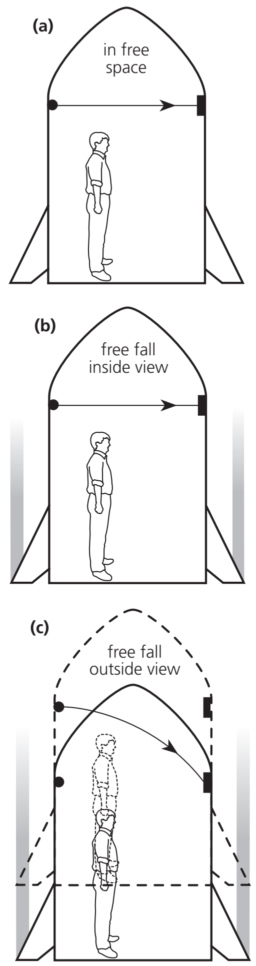


Figure 30.19
'Einstein's elevator' thought experiment. Free fall in a gravitational field is like weightlessness in the absence of a field.

Figure 30.20
A laser beam has different paths to observers inside and outside a falling spaceship. Both are correct.



With his 'Gedanken' experiments, Einstein showed that gravity can be made to vanish, merely by going to a frame of reference that is in free fall. If gravity can be so easily banished, he reasoned that what we call the *force* of gravity may be imaginary. Perhaps gravity is not a force at all, but is somehow related to free motion in '**space-time**'. Einstein concluded that, since gravity has been transformed away within the elevator, all experiments conducted inside it should give the same results as experiments carried out in empty space, where there are no net gravitational influences. In 1915 Einstein translated his thoughts about nature into a rigorous mathematical theory, the **general theory of relativity**. Einstein summarised the results of his reasoning in his **principle of equivalence**, which can be restated as:

> All experiments will give the same results in a local frame of reference in free fall and in a local frame of reference far removed from gravitational influences.

That is, there is no experiment we can perform that will tell us whether we are in a free-falling reference frame (like the elevator above) or in a reference frame far away in space.

Consequences and tests of general relativity

The consequences of Einstein's profound hypothesis are quite remarkable. In the following section you will see some of the logical deductions of the hypothesis and how they have been verified.

Consider Figure 30.20. In a rocket in free space (somewhere far away from stars and planets) a laser beam is emitted from one side of the rocket towards a light detector on the opposite side (diagram **(a)**). The astronaut sees the laser beam travel in a straight line from one side to the other. How would this appear if the same rocket were in free fall near Earth (diagram **(b)**)? From inside, the astronaut will still see the laser beam travel in a straight line across the room. So far, nothing strange! But now consider the same experiment viewed from the reference frame of someone on Earth (diagram **(c)**). The stationary Earth observer also sees the laser beam hit the light detector, but by the time the light beam has crossed the cabin, the rocket and the light detector will have fallen a small distance. So the observer who is stationary with respect to Earth will see the laser beam follow a *curved path*.

Since the stationary observer believes himself to be in a gravitational field (because he feels his weight) he will conclude that gravity bends light. Einstein assumed that light, nonetheless, travels in as straight a line as possible. The fact that light's natural motion is curved could be understood if the space-time through which it travelled were itself curved.

The principle of equivalence is only the *basis* of general relativity. Just as the contraction of length and time dilation equations follow from the postulates of special relativity, a mathematical framework follows from this 'equivalence' postulate. Unfortunately, general relativity theory is too complicated to discuss quantitatively here (it involves the mathematics of **tensors** and **differential geometry**). However, some of its astonishing predictions make interesting reading. Four have become very popular with physicists: (1) the deflection of light by the Sun, (2) gravitational lensing, (3) the precession of the perihelion of Mercury, and (4) gravitational red shift.

Tests of general relativity

Deflection of light

Previously we imagined observing a beam of light in an accelerated elevator and saw that the light path was curved. By the equivalence principle the same must be true for light whenever gravitational forces are present. This was tested by carefully recording the position of stars near the rim of the Sun during an eclipse (see Figure 30.21), and then observing the same stars a year later when the Sun was not in a line between us and the stars.

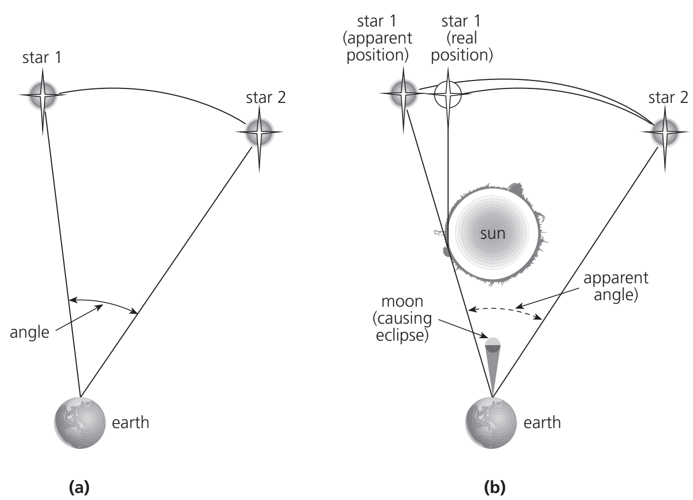


Figure 30.21
Distant stars appear to have a fixed angular distance between them, but if the rays of light from one should pass near the Sun, they are deflected by gravity.

During the eclipse the observed starlight reaches us only after passing through a region where gravitational effects from the Sun are very strong (that is why only stars near the rim are used), but the observations a year later are done at a time where the gravitational effects of the Sun on starlight are negligible.

It is found that the positions of the stars are displaced when photographs of both situations are compared. The deviations are the same as the ones predicted by general relativity. Eddington first observed this effect in 1919 during a solar eclipse.

Gravitational lensing

In the late 1970s, a double quasar was discovered. These quasars are powerful astronomical sources of radiation in the 'radio' band of the spectrum (10 cm to 10 m wavelength). The fact that it was a double quasar was unusual, but both had identical radio 'signatures' so everything about the two quasars seemed to be exactly the same, except that one was fainter than the other. It was suggested that perhaps there was only one quasar and something was producing multiple images. This 'something' was likely to be a massive but optically faint object which had bent its radio waves. A faint galaxy was subsequently discovered between the two quasar images, which confirmed this hypothesis. Other examples have since been discovered (see Figure 30.22).

Figure 30.22
A massive but optically faint galaxy can bend light from a distant quasar, producing multiple images. This is called 'gravitational lensing'.

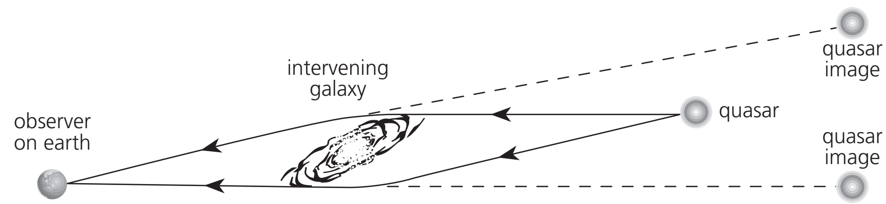
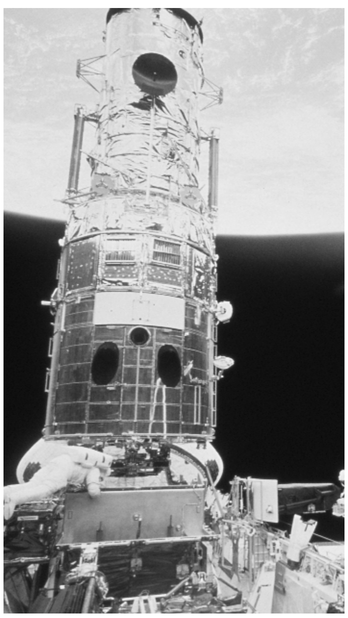


Photo 30.5
Hubble Space Telescope. Now over 10 years old (launched in 1990), the telescope is basically a new machine. Upgrades and maintenance keep Hubble operating in top condition to give us the best scientific data possible.



Gravitational lenses can be considered to be a test of general relativity. This concept has also given general relativity a new role in astronomy. By examining multiple images of a distant galaxy or quasar, their relative intensities and so on, astronomers can gain information about an intervening galaxy whose gravitational field causes the bending of light. The light may take several months or even years longer on one of the light paths than on the other to traverse the ten or more billion light-years to the quasar.

Such an effect has been observed by the Hubble Space Telescope (Photo 30.5). On a photo of a remote galaxy, two mirror images of the structure were observed on opposite sides of the picture (see Photo 30.6: Einstein's Cross). These were believed to be caused by the gravitational lensing of an intervening cluster of galaxies containing much **dark matter** that does not emit electromagnetic radiation and so cannot be detected by regular observations.

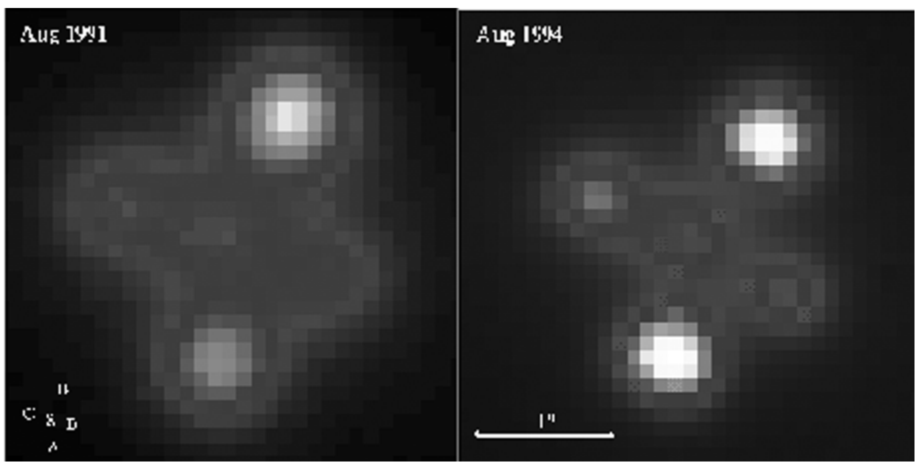
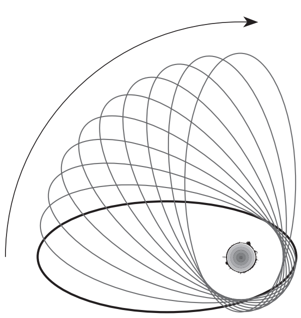


Photo: Geraint Lewis and Michael Irwin, William Hershel Telescope

Photo 30.6
Einstein's Cross is a multiple image of a quasar approximately 8 billion light-years away, formed by gravitational lensing of light as it passes a galaxy twenty times closer (400 million light-years). This is a famous example of an object that is seen four times. Here a very distant quasar happened to be positioned right behind a massive galaxy. The gravitational effect of the galaxy on the distant object was similar to the lens effect of a drinking glass on a distant street light — it created multiple images. But stars in the foreground galaxy have been found to act as gravitational lenses here too! These stars make the images change in brightness relative to each other. The brightness changes are visible on these two photographs of Einstein's Cross, taken about three years apart.
© Robert J. Nemiroff, Jerry T. Bonnell

Figure 30.23
The slightly elliptical orbit of Mercury precesses due to the perturbing effects of the other planets.



Perihelion of Mercury

A long-standing problem in the study of the Solar System was that the orbit of Mercury did not behave as required by Newton's equations. As Mercury orbits the Sun, it follows an ellipse — but only approximately. It is found that the point of closest approach of Mercury to the Sun does not always occur at the same place, but slowly moves around the Sun (see Figure 30.23). This rotation of the orbit is called a *precession*.

The precession of the orbit is not peculiar to Mercury; all the planetary orbits precess. In fact, Newton's theory predicts these effects, as being produced by the pull of the planets on one another. The question is whether Newton's predictions agree with the *amount* an orbit precesses. It is not enough to understand qualitatively what is the origin of an effect; such arguments must be backed by hard numbers to give them credence. The precession of the orbits of all planets except for Mercury's can in fact be understood using Newton's equations. But Mercury seemed to be an exception.

The problem was that Newton's law of gravitation is correct only for a weak gravitational field. A strong field might well cause observation to deviate from the 'classical' expectation. If any planet shows deviation, it should be Mercury, for it orbits where the Sun's field is strongest. Using general relativity, a correction to the classically expected precession rate of Mercury may be calculated. The result of 43 seconds of arc per century is in good agreement with observation.

Gravitational red shift

One prediction of general relativity is that gravitation affects time by causing it to slow down. The greater the gravitational field, the greater is the slowing of time. In a region that has a strong gravitational field, such as the Sun, the slowing of time should be noticeable. The consequence is that the electronic vibrations of the Sun's atoms should also be slower. From your studies of wave motion and visible light you know that the rate of vibration (the frequency) is related to the colour of the light: the lower the frequency, the more the colour is shifted towards the red end of the spectrum. This is different from the Doppler red shift, which is caused by the relative movement of the source (stars) and observer. In 1960, two American physicists, R. V. Pound and G. A. Rebka, Jr, detected the red shift resulting from the Earth's gravitational field in agreement with Einstein's predictions.

The future

The special theory of relativity showed that relative motion can dilate time intervals and contract length. With general relativity we have seen that a gravitational field can also change (warp) time intervals, even when there is no relative motion. So we can ask, 'Can gravity warp space intervals as well?' The answer is 'Yes'. General relativity predicts that a massive heavenly body warps space-time nearby.

Representing warped space-time in three dimensions is difficult. It is easier in two dimensions, in which space is *area*. Figure 30.24 shows a massive heavenly body disturbing the uniformity of a two-dimensional space. All 'cells' in this 2D space are of equal area, but you should be able to see that to 'outside observers' like us, the cells near the heavenly body are really larger — but only from our 'extra-dimensional' viewpoint. You wouldn't notice the warping of real three-dimensional space if you were located within it — it is only apparent to someone outside it. We ourselves live within our space of three dimensions, and are not able to stand back and view our universe on four-dimensional axes. However, 'reduced-dimensional' views such as Figure 30.24 help to provide a qualitative understanding of some features of general relativity. These are the challenges facing physicists today.

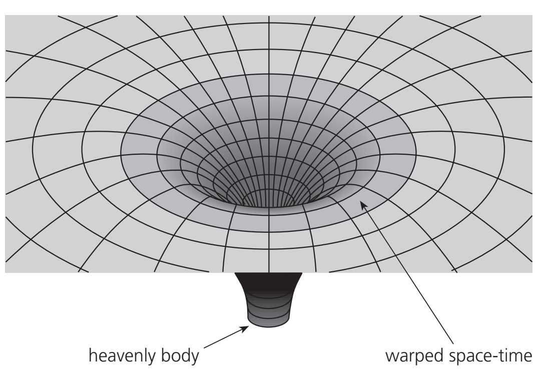


Figure 30.24
A massive heavenly body disturbs the regularity of a two-dimensional space.

Another challenge relates to new evidence for and against relativity. After a century, Einstein's theories have held up remarkably well. But as scientists probe the edges of the current knowledge of physics with new tests, they may find effects that require modifications of the venerable theory.

Several current theories, designed to encompass the behaviour of black holes, the Big Bang and the fabric of the universe itself, could lead to violations of special relativity. So far, recent, updated versions of century-old experiments show no signs that Einstein's vision is reaching its limits. Various tests are ongoing, however, and a new generation of ultra-precise, space-based experiments is set to launch in the next few years, offering some chance — however slim — of observing signs of the laws that will eventually supersede relativity.

In an updated version of the Michelson–Morley test, researchers have already sent laser light into two optical cavities set at right angles to each other. The light forms a standing wave in each cavity, with a frequency that depends on the cavity length and the speed of light in that direction. If light can go faster in one direction of space than another, rotating the apparatus should reveal this effect as a change in the relative frequencies between cavities. The team's preliminary results showed no deviation from special relativity.

In 2003 physicists also found that Einstein's theory passed the most accurate version yet of the Kennedy–Thorndike test – perhaps the most critical test of relativity first done in the 1930s. They compared the resonance of a standing light wave with an atomic clock over a period of 190 days, during which time Earth's orbital speed changes by 60 km s^{-1}. The result was ten times as accurate as previous Kennedy–Thorndike measurements, and no deviation from relativity theory was found.

Others physicists have monitored highly stable atomic clocks. These are collections of atoms that radiate at a certain frequency. Deviations from special relativity would show up as changes in their frequencies, depending on which way Earth is pointing. Earth-bound atomic clocks start to become unstable after just a few hours. Gravity, daily temperature changes and mechanical degradation are all sources of error. So the next generation of atomic clock measurements (the Primary Atomic Reference Clock in Space, PARCS) will compare the frequency of an ultra-cold caesium atomic clock against a hydrogen microwave laser (or maser). These measurements are scheduled to run on satellites or the International Space Station, where microgravity and a shorter rotation time should allow higher accuracy.

Finally, two important tests of relativity will come late this decade: the Superconducting Microwave Oscillator (SUMO) will keep time with a microwave-filled superconductor cavity. It will make measurements on its own and in conjunction with the Rubidium Atomic Clock Experiment (RACE) and the Space Time Mission (STM) will slingshot a satellite containing three atomic clocks around Jupiter for a close view of the Sun. The high speeds involved will offer more sensitive tests of relativity.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***14** If you were standing on top of a moving train and threw a rock straight up (as it appeared to you):
(a) how would the motion of the rock appear to your mother, who is standing on the platform; **(b)** would it land behind the carriage or on top of it;
(c) would you be in trouble when she got hold of you?

***15** If you were on a rocketship travelling at 0.6c away from the Sun, at what speed would the sunlight pass you?

****16** How fast must a pion be travelling if its rest life is 2.6×10^{-8} s but to a laboratory observer it appears to live for 5.2×10^{-8} s?

****17** A certain type of elementary particle travels at a speed of 2.6×10^{8} m s^{-1}. At this speed, the average lifetime is measured to be 2.20×10^{-7} s.
(a) Express its speed in units of 'c'.
(b) What is the particle's lifetime at rest?

***18** Do mass increase, time dilation, and length contraction occur at ordinary speeds, say, 100 km h^{-1}?

****19** A spaceship passes you at a speed of 0.75*c*. You measure its length to be 120 m. How long would it be when at rest?

****20** Suppose you decide to travel to a star 80 light-years away. How fast would you have to travel so the distance would be only 40 light-years?

****21** If you were to travel to a star 24 light-years from Earth at a speed of 2.4×10^8 m s^{-1}, what would you measure this distance to be?

****22** Spaceships Alpha and Beta are approaching Earth at speeds of 0.85*c* and 0.65*c* respectively. Calculate the speed of Alpha relative to Beta if the ships are on **(a)** parallel tracks; **(b)** anti-parallel tracks.

****23** In the medical procedure called positron emission tomography (PET), a radioactive nucleus emits a positron that is captured by an electron. Both particles immediately self-destruct, forming two gamma photons that travel out in opposite directions. Knowing that photons travel at '*c*', calculate the speed of one relative to the other.

****24** At what speed will an object's mass be twice its rest mass?

****25** **(a)** What is the speed of an electron whose mass is 800 times its rest mass?
(b) If the electrons travel in a lab through a linear accelerator 1.5 km long, how long will this tube be in the electrons' reference frame?

****26** **(a)** How much energy can be obtained from conversion of 1.0 μg of mass?
(b) How much mass could this energy raise to a height of 10 m?

***27** If you were travelling away from Earth at a speed of 0.5*c*, would your mass, height, or waistline change? What would observers on Earth using telescopes say about these things?

***28** Consider the piece of paper on which one page of this book is printed. Which of the following properties of the piece of paper are absolute, that is, which are independent of whether the paper is at rest or in motion relative to you?
(a) The thickness of the paper; **(b)** the mass of the paper; **(c)** the volume of the paper; **(d)** the number of atoms in the paper; **(e)** the chemical composition of the paper; **(f)** the speed of the light reflected by the paper; **(g)** the colour of the coloured print on the paper.

Extension — complex, challenging and novel

****29** Because of the rotational motion of the Earth about its axis, a point on the Equator moves with a speed 460 km h^{-1} relative to a point on the North Pole. Does this mean that a clock placed on the Equator runs more slowly than a similar clock placed on the Pole?

****30** A 100 MeV electron, travelling at 0.999 987*c*, moves along the axis of an evacuated tube that has a length of 3.00 m as measured by a laboratory observer S with respect to whom the tube is at rest. An observer S′ moving with the electron, however, would see this tube moving past her. What length would the tube appear to the observer S′?

****31** A friend of yours travels by you in her fast sports car at a speed of 0.760*c*. It appears to be 5.80 m long and 1.45 m high.
(a) What will be its length and height at rest?
(b) How many seconds did you see elapse on your friend's watch when 20.0 s passed in the Earth's frame of reference?
(c) How fast did you appear to be travelling to your friend?

*****32** The star Alpha Centauri is 4.0 light-years away. At what constant velocity must a spacecraft travel from Earth if it is to reach the star in 3.0 years, as measured by travellers on the spacecraft?

*****33** Suppose that a special breed of cat (*Felix schrödingerus*) lives for exactly 7.00 years according to its own body clock. When such a cat is born, it is put aboard a spaceship and sent off at a speed of 0.80*c* toward the star Alpha Centauri.
(a) How far from the Earth (reckoned in the reference frame of the Earth) will the cat be when it dies?
(b) As soon as the cat dies, a radio signal announcing the death of the cat will be sent from the spaceship. How many years after departure will the signal reach Earth (radio signals travel at the speed of light)?

****34** Suppose a spacecraft of rest mass 20 000 kg is accelerated to 0.25*c*.
(a) How much kinetic energy would it have?
(b) If you used the classical formula for kinetic energy, by what percentage would you be in error?

*****35** An object with a rest mass of 1.000 kg is accelerated to high speed. Calculate the dilated mass (to 4 significant figures) of the object for speeds increasing from rest in increments of 0.1*c* up to 0.9*c* and also at 0.95*c*, 0.99*c*, 0.999*c* and 0.9999*c*. Plot a graph with speed on the x-axis. Use a computer spreadsheet if you like.

*****36** You are sitting in your Holden when a very fast sports car passes you at a speed of 0.18*c*. A person in the car says his car is 6.00 m long and yours is 6.15 m long. What do you measure for these two lengths?

*****37** Apart from the Sun, our nearest star is Proxima Centauri which is 4.225 light-years away. How many years would it take a spacecraft travelling 0.80*c* to reach that star from Earth as measured by observers **(a)** on Earth **(b)** on the spacecraft? **(c)** What is the distance travelled according to observers on the spacecraft?

*****38** Prove that the kinetic energy of an electron of rest mass 9.11×10^{-31} kg, which has a relativistic mass of 2.0×10^{-30} kg, is 0.62 MeV (1 J = 6.24×10^{18} eV) and that its speed is 2.7×10^{8} m s^{-1}.

****39** A proton has a total energy three times its rest energy.
(a) Calculate its rest energy in MeV.
(b) Calculate its KE in MeV.
(c) How fast is it travelling?

*****40** In relativity, momentum is conserved just as it is in classical physics. The formula for relativistic momentum is the same except that the relativistic mass (m) must be used instead of rest mass (m_0).
(a) Derive a formula for relativistic momentum ($mv = m_0$ etc.).
(b) Square both sides and rearrange to show that $m^2 = m_0{}^2 + (mv)^2/c^2$.
(c) Multiply both sides by c^4 and show that $m^2c^4 = m_0{}^2c^4 + c^2p^2$.
(d) For a photon (rest mass = zero), show that this equation becomes E_{total} (= mc^2) = cp.
(e) Using the equation in part **(c)**, show that the relativistic momentum of an electron described in Question 38 above is 5.4×10^{-22} kg m s^{-1}.

*****41** Imagine a rocketship takes off for a distant planet and can travel at many times the speed of light. (We know that this is impossible but let's just say you can for this question.) Observers on the planet are viewing the incoming spaceship through a powerful telescope. Describe what they will see from the moment the rocketship leaves Earth until it lands on the observers' planet.

CHAPTER 31

Designing Practical Electronic Circuits

31.1 INTRODUCTION

In Chapters 23 and 24 we looked at the behaviour of basic components in electronics as well as simple systems that could be produced with them. In this chapter we will examine, in more detail, practical applications of electronics, especially using integrated circuit systems. We will look at circuit examples that you might like to try building as the basis of a hobby project. You might become interested enough to purchase and build one of the many hundreds of electronics constructor kits that are commercially available.

Most electronics today is based on combinations of integrated circuits, especially digital ICs, and these are quite easy to use. Always keep in mind the safety aspect of electricity and only deal with kits or projects that involve battery power supplies or use mains plug pack transformers, as described later in this chapter.

31.2 RLC RESONANCE AND TUNING CIRCUITS

Radio waves transmitted from sources such as AM and FM radio stations, television channels and CB or short wave transmitters all involve different frequency electromagnetic signals or voltages. In order to learn how to detect these signals with electronic circuits, we need to examine the AC behaviour of capacitor and inductor components.

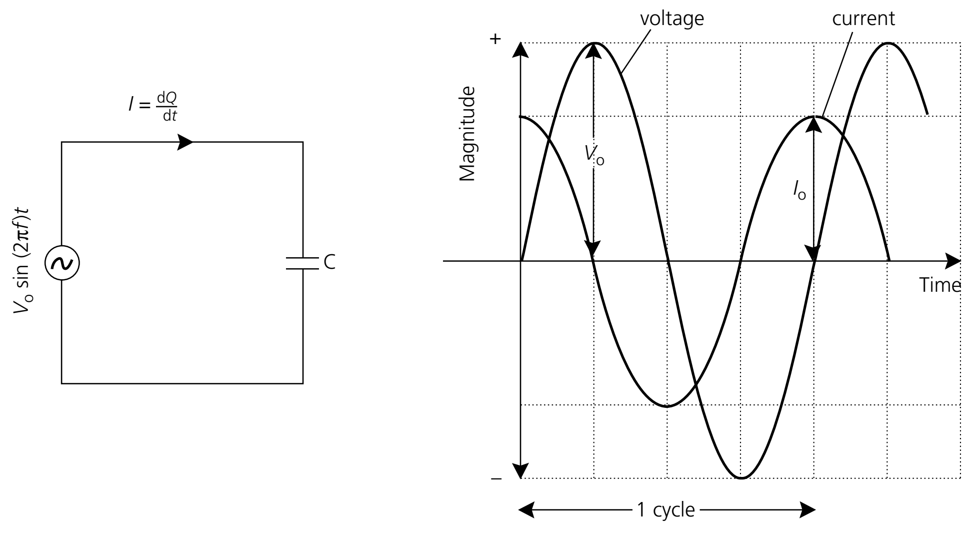


Figure 31.1
Capacitors and AC voltage.

Recall the DC timing constant property of a charging capacitor in an RC circuit. Let's see what occurs when an AC voltage is placed across any capacitor, as in Figure 31.1. The capacitor will charge instantly, with the voltage across it at any time being equal to the supply voltage. A sinusoidally varying AC voltage (sine wave) of frequency f will be given by the equation:

$$V = V_o \sin \omega t = V_o \sin (2\pi f)t$$

(where V_o is the voltage peak amplitude).

and this voltage will at any instant of time be equal to the voltage as defined by the capacitance, namely:

$$V = \frac{Q}{C} = V_o \sin (2\pi f)t$$

or

$$Q = V_o \,.\, C \,.\, \sin (2\pi f)t$$

but the current, I, flowing to the capacitor at any time will be given by:

$$I = \frac{dQ}{dt} \text{ (rate of change of charge with time)}$$

Hence, current flow in any capacitor connected to an AC voltage supply will be:

$$I = \frac{dQ}{dt} = \frac{d(V_o C \sin (2\pi f)t)}{dt}$$

$$I = V_o C\, 2\pi f \cos (2\pi f)t$$

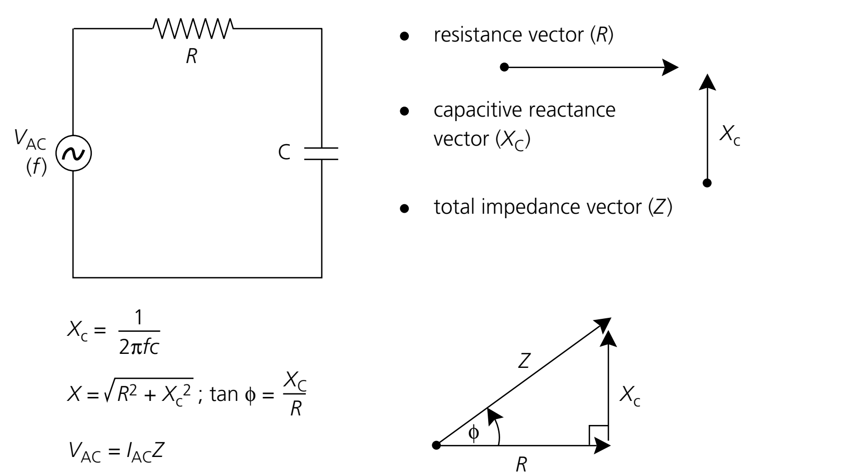
$$I = I_o \cos (2\pi f)t \quad \text{where} \quad I_o = V_o\, 2\pi f C$$

(where I_o is the current peak amplitude).

Note that because the sine curve and a cosine curve are out of phase, the current peak, I_o, leads the voltage peak, V_o, by 90°. This is called a phase shift or **phase angle**. Also notice that the link between I_o and V_o for a capacitor can be written as:

$$V_o = I_o \frac{1}{2\pi f C} = I_o X_C$$

which is similar to Ohm's law and introduces a property of frequency-dependent resistance or **reactance** for a capacitor.



$$X_c = \frac{1}{2\pi fc}$$

$$X = \sqrt{R^2 + X_C^2}\,;\ \tan \phi = \frac{X_C}{R}$$

$$V_{AC} = I_{AC} Z$$

Figure 31.2
RC circuit and impedance.

The **capacitive reactance**, $X_C = \frac{1}{2\pi fC}$, becomes a phasor quantity (similar to a vector quantity) with both magnitude in ohms and a phase angle in degrees. In the RC series circuit of Figure 31.2, the vectors are drawn representing pure resistance, R, in which current and voltage are in phase, and capacitive reactance, X_C, in which the current leads the voltage by 90°. The total AC resistance to the flow of alternating current (AC) from the supply is called the circuit **impedance**, Z, and is calculated by vector addition processes.

In a simple series RC circuit the total impedance:

$$Z = \sqrt{R^2 + X_C{}^2}$$

and the phase angle (ϕ) is given by:

$$\tan \phi = \frac{X_C}{R}$$

The impedance, Z, is measured in ohms. Ohm's law equivalent for this AC series RC circuit becomes:

$$V = I \,.\, Z$$

Example

In the circuit of Figure 31.2, the AC voltage was 12 V RMS and the supply frequency is 50 Hz, the capacitor has a value of 0.33 μF and the resistance is 10 kΩ. Find:

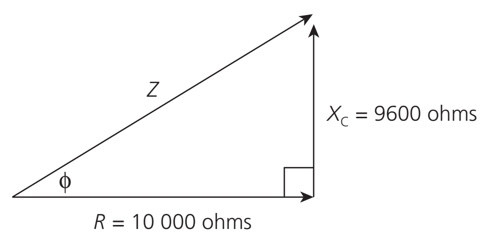
(a) the capacitive reactance X_C;
(b) the total circuit impedance Z;
(c) the AC circuit current (RMS);
(d) the phase angle between voltage and current.

Solution

(a) At f = 50 Hz:

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 50 \times 0.33 \times 10^{-6}}$$
$$X_C = 9600\ \Omega$$

(b) Completing a vector diagram for impedance Z:



$$Z = \sqrt{(9600)^2 + (10\ 000)^2} = 13\ 900\ \Omega$$

(c) Hence, AC current, I_{RMS}, can be calculated using $V = IZ$.

$$12 = I_{RMS} \times 13\ 900$$
$$I_{RMS} = 0.86\ \text{mA}$$

(d) Phase angle is given by:

$$\tan \phi = \frac{X_C}{R} = \frac{9600}{10\ 000} = 0.96 \quad \therefore \text{ angle } \phi = 44°$$

Thus, in this circuit, the current sine wave reaches a maximum 44° before the voltage sine wave reaches its maximum. This would be best shown graphically.

A similar analysis to that above can be applied to a simple inductor coil connected to an AC voltage source. (Refer to Figure 31.3.) The result is that an inductor provides an opposition to the flow of alternating current in a circuit that is frequency-dependent and is called **inductive reactance** (X_L), where $X_L = 2\pi fL$.

Notice that an inductor is a low resistance to low frequency AC and provides high resistance to high frequency AC. This reactance is again measured in ohms. For any inductor coil, the AC current flow lags the AC voltage across the coil by a 90° phase shift and thus the inductive reactance vector can be drawn pointing downward.

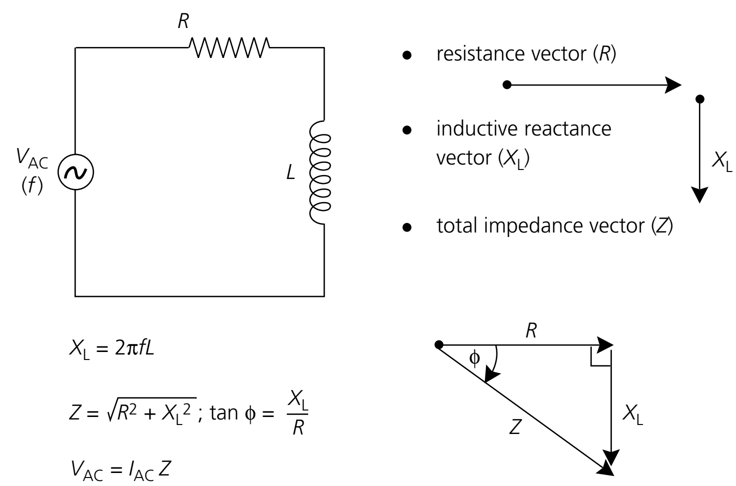


Figure 31.3
RL circuit and impedance.

To calculate total circuit impedance in an RL series circuit, vector processes are again required.

$$Z = \sqrt{R^2 + X_L^2}$$

Series RLC resonance

Consider the circuit of Figure 31.4 showing an RLC series connection to an AC voltage source. In this circuit the inductive reactance will tend to balance the capacitive reactance because of the opposite direction of the vectors. At one particular AC frequency $X_L = X_C$ and a maximum value of current will flow, only restricted by the pure resistance R. At this particular frequency the series RLC circuit is said to **resonate**. An RLC series circuit will resonate at a frequency calculated by equating the reactances, hence:

$$2\pi fL = \frac{1}{2\pi fC} \quad \text{or} \quad f_R = \frac{1}{2\pi\sqrt{LC}}$$

where f_R = resonant frequency.

It is important to realise that in this circuit the total impedance at resonance will be equal to the circuit resistance, R. Figure 31.4 also illustrates graphically the impedance and current flow as a function of frequency.

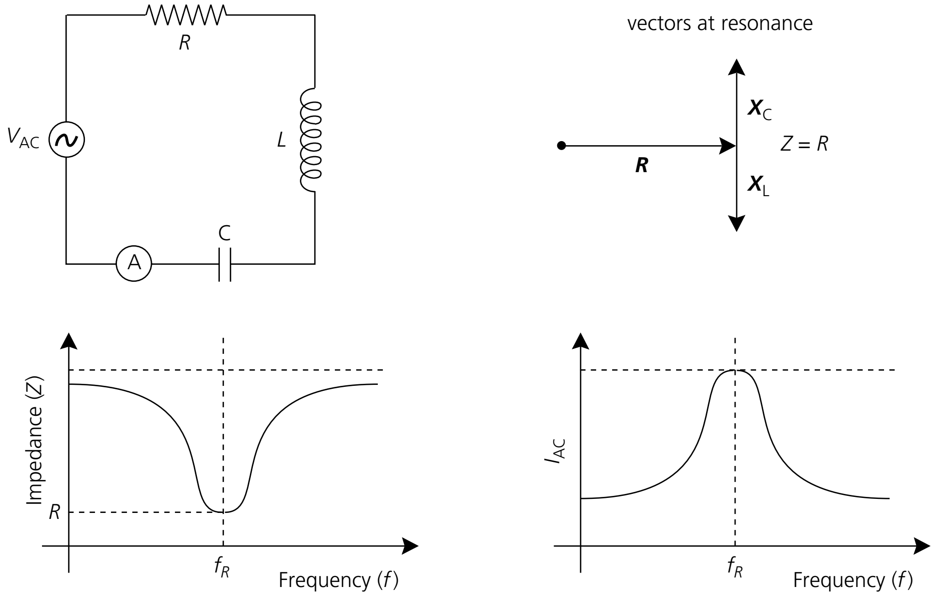


Figure 31.4
RLC series resonance.

In any series circuit at resonance, the voltage across the capacitor and the inductor individually may be many times more than the supply voltage. This is true even if the total voltage across the combination of *L* and *C* together is zero because of the effect of opposite phase angles. This is a danger to anyone who may touch capacitors or inductors in any high voltage AC circuits, as shown in the following example.

Example

An RLC circuit is connected to a variable frequency AC generator whose effective voltage output is 48 V_{RMS}. If the circuit elements have values $L = 100$ mH, $C = 0.02$ µF, $R = 50\ \Omega$, find:

(a) the resonant frequency of the circuit;
(b) the circuit AC current at resonance;
(c) the voltages across each component R, L and C at this resonant frequency.

Solution

(a) The resonant frequency:

$$f_R = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{100 \times 10^{-3} \times 0.02 \times 10^{-6}}}$$

$$f_R = 3.6 \times 10^3 \text{ Hz}$$

(b) At resonance:

$$I_{AC} = \frac{V_{AC}}{R} = \frac{48}{50} = 0.96 \text{ A}$$

(c) Voltage drops across each component:

$$\begin{aligned}
V_C &= IX_C = I \times 1/2\pi fC = 0.96/2\pi \times 3.6 \times 10^3 \times 0.2 \times 10^{-6} \\
V_C &= 2.1 \times 10^2 \text{ V} \\
V_L &= IX_L = I \times 2\pi fL = 0.96 \times 2\pi \times 3.6 \times 10^3 \times 100 \times 10^{-3} \\
V_L &= 2.2 \times 10^3 \text{ V} \\
V_R &= I.R = 0.96 \times 50 = 48 \text{ V}
\end{aligned}$$

Notice the high voltages individually across L and C, which are considerably greater than the AC values of the EMF in the circuit.

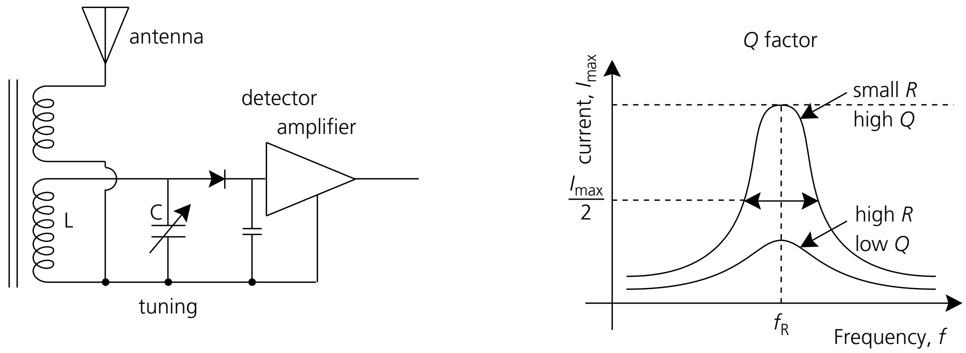


Figure 31.5
Radio tuning circuits.

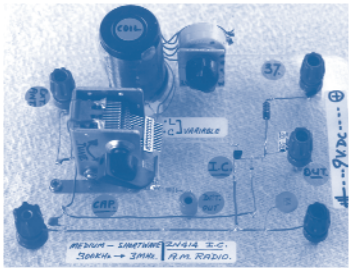


Photo 31.1
Tuning module.

PHYSICS UPDATE

In 2002 the engineers at Motorola's Semiconductor Products Sector in Austin, Texas developed a set of silicon chips that apply sophisticated digital processing to standard analog signals, enabling software code rather than analog circuitry to do the tuning. Called 'Symphony Digital Radio', the system relies on algorithms running at the rate of 1500 million instructions per second on Symphony's 24-bit semiconductor chip set. The device converts any incoming tuned AM or FM signal into an intermediate frequency that can be filtered and conditioned by DSPs (digital signal processors). The result can be almost CD-quality sound from analog radios, given a sufficiently strong signal. The Motorola system represents an early example in a new class of what the electronics industry calls software or software-defined radios, a technology that derives tremendous flexibility by using digital code in place of fixed hardware to accomplish functional tasks.

One of the most useful applications in electronic circuits for the property of resonance is in **tuning circuits**, as used in radio and television receivers. (See Figure 31.5.) Resonance is used to select a desired frequency or radio channel from the multitude of frequencies arriving at the receiver antenna from all available broadcasting stations. A variable tuning capacitor is used in an RLC circuit, known as a **tank oscillator**, to select a resonant frequency equal to the carrier wave or broadcast frequency of the channel selected. For example, the carrier frequency of radio station Brisbane B105 FM is 105.4 MHz. At the resonant frequency, the EMF induced in the antenna by the incident waves causes a large current in the antenna circuit, which can then produce a modulated audio or video signal. Other channels or stations that have carrier frequencies not in resonance with the antenna circuit will produce negligibly small currents. The quality factor or ***Q* factor** of a resonant circuit is defined as the ratio of the voltage across the capacitor to that of the voltage across the resistor at resonance:

$$Q = \frac{V_C}{V_R} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

The larger the Q factor of a circuit, the sharper will be the resonance curve and therefore the more selective will be the tuning ability. This means that the radio receiver will be better able to tune one individual radio station without at the same time receiving another station that happens to be in a close position on the broadcasting frequency band. Have you ever listened at night to your transistor radio and heard more than one station at the same time; it can get rather confusing, can't it? This effect is often due to poor selectivity and it increases at night when reception of weaker strength radio signals is often enhanced.

VOLTAGE REGULATORS AND POWER SUPPLIES 31.3

Recall in Chapter 23, when referring to rectification, it was stated that simple DC power supply circuits suffer from a lack of voltage regulation. The solution offered was to use a zener diode to maintain constant voltage conditions. In more advanced DC power supply designs a better choice is a three terminal linear IC called a **voltage regulator** or a **three terminal regulator**. These devices are one of the most useful IC components in modern electronics simply because they do away with the need for complex voltage regulation circuits when designing the power supply section for consumer items, such as small transistor radios or cassette players. In fact, highly regulated DC voltage supplies are vital for most digital circuits such as computers and audio–video control circuits.

The term regulated power supply means that active devices such as transistors and Op-Amps are used to electronically filter or eliminate variations in output voltage and sometimes current. The stabilised voltage output of a regulated supply allows a circuit to operate more precisely. Three terminal regulator ICs provide very high general regulation characteristics as well as other advantages such as current limiting, thermal overload protection against over-heating, and the ICs are available as either positive, negative or variable output voltage types.

Three-terminal regulators have three-pin connections called input, output and common, as shown in Figure 31.6, which illustrates the common 7800 series of positive regulators in a TO-3 or TO-220 IC case design. For example, a 7805 voltage regulator has an output voltage of exactly +5.0 V. The last two digits of the series generally refer to the regulator's output voltage. Similarly, the 7900 series is a negative voltage regulator family providing similar features to the positive types but with a different pin configuration. As usual, it is always wise to check the manufacturer's pin diagrams before connecting these regulator chips into the circuit. Both of these regulator families provide output currents up to 1.5 A maximum if the chip is adequately mounted onto an aluminium **heat sink** to dissipate the heat produced as the IC chip operates.

Figure 31.6
Three-terminal regulators — 7800 series pin diagrams.

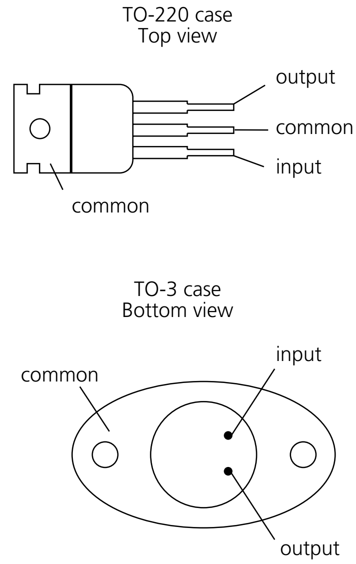


Figure 31.7 illustrates a common circuit used in voltage regulator power supply design. The transformer and diode bridge provide a rectified input to the device and the output is maintained at a very constant positive 5.0 V at 1.0 A maximum current capability. Notice that, as with all regulators, the input voltage has to be at least 2 V higher than the desired output, preferably even higher, in order to maintain regulation and eliminate ripple voltage passing across from the rectifier. As well, capacitors are usually connected between the common terminal and both the input and output terminals, as close as possible to the regulator chip itself.

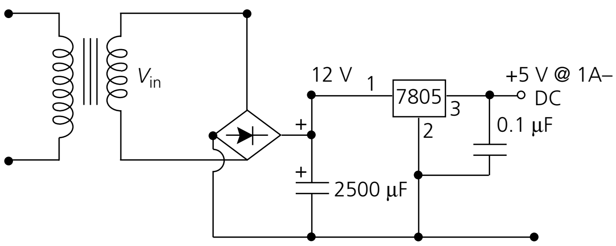


Figure 31.7
Voltage regulator.

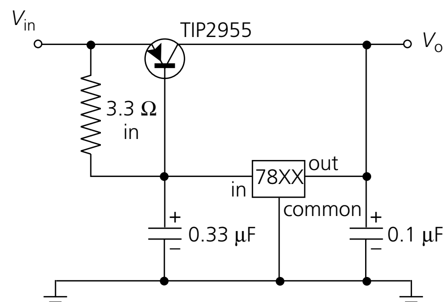


Figure 31.8
Current pass transistor with a voltage regulator.

If a power supply circuit design calls for a high current capability as well as voltage regulation properties, then a **current bypass** power transistor can be used, as shown in Figure 31.8. In this circuit a PNP power transistor, such as a TIP2955, is used to allow for an output voltage of 12 V at a maximum current of 4.0 A. The input to this circuit again would be any rectified and filtered DC voltage greater than about 15 V.

Sometimes the most versatile power supply designs are those that provide variable output voltages all at the same maximum current capability. The LM317 or LM350 series regulators allow this type of function with a circuit as shown in Figure 31.9. In this circuit design the value of resistor R is given by:

$$R = (96V_o) - 120$$

where V_o is output voltage required.

Of course, if R is made a variable resistor or potentiometer, then the circuit is fully variable and not just adjustable. Again, V_{in} should be at least 2.5 V higher than the required V_o and the capacitor voltage ratings must match the required input and output voltages.

[$R = 96V_o - 120$]

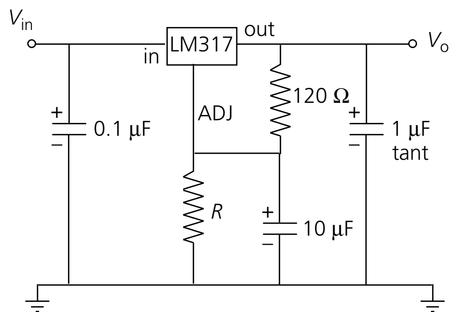


Figure 31.9
LM317 adjustable voltage regulator.

With these circuits it is easy to see that simple power supply design is within reach of most constructors. One final integrated circuit regulator chip that is very versatile is the 723 voltage regulator, which is intended for both positive and negative supply design. Its output voltage is adjustable between 2 and 37 volts. Its own internal transistor can supply currents up to 150 mA, but output current capability can be increased with the appropriate external bypass transistor.

NEI Activity 31.1 DESIGNING A POWER SUPPLY

Obtain copies of the data sections from electronics supplier catalogues, such as can be obtained from Dick Smith Electronics or Tandy electronics stores. Research the full range of three terminal regulators available together with their respective costs and operating characteristics. You might like to see what it would cost to build a complete DC power supply designed to operate from a plug-pack AC transformer. Aim for a variable supply from 3 volts to 9 volts output capable of supplying up to 1.0 amp of current.

Questions

1 In a series RC circuit operating at an AC frequency of 1.0 kHz and voltage 10 V_{RMS}, R = 2700 Ω, C = 0.1 µF. Calculate **(a)** capacitive reactance; **(b)** total circuit impedance; **(c)** the alternating current flow in the circuit; **(d)** the phase angle between voltage and current.

2 An RLC series circuit is connected to an AC generator of voltage 50 V_{RMS}. In the circuit L = 100 mH, C = 2.0 µF, R = 150 Ω.
(a) What is the resonant frequency?
(b) What is the effective current at resonance?
(c) What are the voltage drops across each component at resonance?
(d) What power is delivered to the circuit at resonance?
(e) What is the circuit Q factor?

3 Using the circuit of Figure 31.9, calculate the value of the adjust resistor needed to supply an output voltage of 7.2 V. What would be an appropriate input voltage for this circuit?

TRANSISTORS AS AC VOLTAGE AMPLIFIERS 31.4

Recall that, in Chapter 24, a transistor was described as an active semiconductor three terminal device that could be biased to act either as a direct current amplifier or as an electronic switch. One of the most practical uses of transistors in electronic circuit design is as an amplifier for small AC voltages. Such small voltages are produced, for instance, as the output of a microphone. AC **voltage amplifiers** based on either discrete transistors or integrated circuits (Op-Amps) form the heart of many electronic systems and are one of the building blocks of modern electronic technology. Let's look now into their principles of operation and design.

It has been seen that a small base current (DC) flowing into a transistor will control a much larger collector current. If this base current is made to change in a periodic fashion by a small AC voltage signal connected to the transistor's base, then a corresponding amplified AC voltage will appear across the transistor output or collector resistance. The transistor needs to be biased correctly to avoid various types of output voltage distortion. The voltage across the collector resistor depends on both the size of the collector current and the value of the collector resistance. Remember that for any given base current the size of the collector current depends on the transistor gain, β, hence a transistor AC voltage gain depends on its own current gain and also on the value of the collector resistor used.

The available transistor output AC voltage is usually taken via a capacitor connected directly (coupled) to the collector terminal of the transistor. Let us now look at how to design a practical transistor amplifier circuit that could be used to amplify the very small AC voltage produced by a microphone. The circuit is shown in Figure 31.10 and is correctly called a linear class A common emitter NPN transistor amplifier. Transistors suitable for this type of circuit are often called general purpose small signal amplifier transistors or GPSS types. Common examples are designated as BC108, BC548 or 2N3566, but hundreds of different types are manufactured.

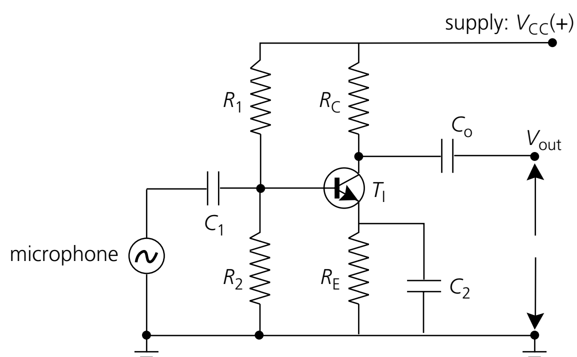


Figure 31.10
Common emitter class A small signal amplifier.

Linear amplifiers should fulfil these basic principles of design:

- The output signal voltage should be an exact replica of the input, but much larger in amplitude.
- The amplifier's input impedance must be as close as possible in value to that of the output of the voltage transducer driving it.
- No unwanted AC signals or DC voltages should enter the amplifier input.
- The amplifier's output impedance must match the load into which it is driving.
- The amplifier should introduce minimum distortion or change of wave shape at the output.

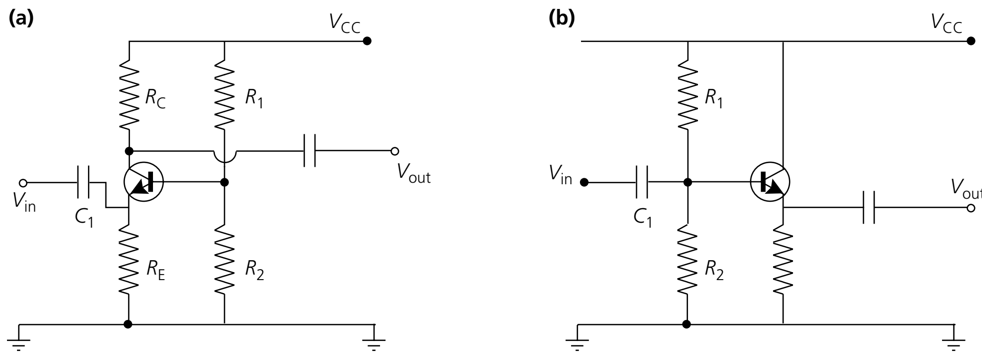
In a **class A amplifier**, such as we are designing, the transistor conducts during the complete input signal cycle, that is, for 360°. In class B and class C amplifiers, transistor conduction is not for the full input cycle and while this is often more efficient for higher voltage and power levels, they are more difficult to design and will not be considered here. The transistor in Figure 31.10 is connected in **common emitter** mode because the transistor emitter terminal forms part of the current loop for both input and output circuits. Other methods of transistor connection are called common base and common collector or emitter follower and are shown in Figure 31.11. These are less often used in general purpose amplifier circuits and are restricted to special applications.

INVESTIGATING

The symbol for the power supply in these circuits is V_{CC}. Find out what the 'CC' means.

NPN bipolar transistors will be used in our amplifier design because these devices have a high β gain, which can produce large voltage gains in the output circuit comprising resistor R_C in parallel with the output terminal load device. Manufacturers provide β gain values on data sheets, both an AC and a DC value, but because our designs assume only small input signal voltages, we will take these as equivalent. **Biasing** is the name given to the fixed DC voltage connected between transistor emitter and base, which causes a steady current to flow in the base–emitter circuit when no input voltage signal is applied through capacitor C_1. The

Figure 31.11
Other amplifier modes: **(a)** common base; **(b)** emitter follower.



(a) Common base (b) Emitter follower

biasing network consists of voltage divider R_1 and R_2 and the emitter resistor R_E. Biasing is necessary to provide the correct operating DC conditions for our amplifier, specifically:

- to provide a conduction voltage of more than 0.7 V for silicon transistors
- to avoid output distortion due to the rectifying properties of the base–emitter junction
- to prevent thermal runaway occurring as a result of amplified leakage current being temperature-dependent and adding to the base current — this could cause thermal breakdown of the semiconductor
- to provide for a variation of characteristics among manufactured transistors. This means the bias network has to cope with the fact that no two transistors, even of the same type, are identical and so if one device has to be replaced by another, the overall circuit does not have to be redesigned. One characteristic that varies widely from one transistor to another is the current gain factor, β. For example, BC108 transistors can vary from a value of 100 to a value of 800 in this factor.

The amplifier of Figure 31.10 uses **voltage divider bias** because the base bias voltage is determined by the resistors acting as a voltage divider to the supply voltage. Let us look at a typical set of steps for designing a transistor amplifier of this type. We will assume a transistor has been selected that has a maximum DC collector current of 5 mA and we are using a power supply V_{CC} of 12 V.

Step 1 Calculate value of R_E

It is appropriate to drop about 10% of the supply voltage across R_E. Hence, 10% of 12 V = 1.2 V. For class A operation, the quiescent or steady state collector current should be about half the maximum allowed, that is, in this circuit, a value of 2.5 mA, which is the same as I_E. Hence:

$$R_E = \frac{V_{RE}}{I_E} = \frac{1.2\ \text{V}}{2.5\ \text{mA}} = 480\ \Omega$$

Step 2 Calculate value of R_C

For a class A amplifier, R_C, the collector resistor, should have a value such that half the supply voltage appears across the transistor's collector–emitter terminals, with the quiescent collector current flowing. In our amplifier, at an I_C value of 2.5 mA:

$$V_{CE} = \frac{V_{CC}}{2} = \frac{12}{2} = 6\ \text{V}$$

Now the remaining 6 V drop will appear across R_C and R_E. From Step 1 and Ohm's law:

$$\begin{aligned} 6\ \text{V} &= 2.5\ \text{mA}\ (R_C + 480) \\ R_C &= 1920\ \Omega\ \text{(value of 2 k}\Omega\text{ is close)} \end{aligned}$$

Step 3 Calculate the voltage at the transistor's base, V_B

Assuming a silicon transistor, V_{BE} = 0.7 V, hence the voltage at the base $V_B = V_{BE} + V_{RE}$, V_B = (0.7 + 1.2) V = 1.9 V.

Step 4 Calculate maximum base current, I_B

A small signal voltage amplifier transistor will have a worst case β value of, say, 200.

Hence $I_B = \frac{I_C}{\beta} = \frac{2.5 \text{ mA}}{200}$ = 0.013 mA, which will represent the maximum value of base current required. This figure allows a calculation of both R_1 and R_2 values.

Step 5 Calculate R_1 and R_2 divider values

The current drawn into the transistor's base is supplied by the voltage divider $R_1 : R_2$. If the voltage at the base is to remain constant irrespective of the current value, I_B, it is necessary to make the current flow through the divider $R_1 : R_2$ much larger than I_B — about 10 times, in fact, as a rule of thumb.

Hence:

$$\text{current through } R_2 = 10 \times 0.013 \text{ mA} = 0.13 \text{ mA}$$

Given that the voltage required at the base, V_B = 1.9 V, the value of:

$$R_2 = \frac{1.9 \text{ V}}{0.13 \text{ mA}} = 14.6 \text{ k}\Omega = 15 \text{ k}\Omega \text{ closest}$$

Also: the voltage drop across resistor R_1 = 12 − 1.9 = 10.1 V, and the value of the current through R_1 = 0.13 mA. Thus, the value of:

$$R_1 = \frac{10.1 \text{ V}}{0.13 \text{ mA}} = 77.7 \text{ k}\Omega = 82 \text{ k}\Omega \text{ closest}$$

Step 6 Calculate values of capacitors

These capacitors are the least critical in practical amplifier design. The input and output **coupling capacitors** are usually chosen to be in the range of 1.0–10 μF. The **emitter bypass capacitor**, C_2, is necessary to obtain maximum voltage amplification from this type of circuit. A portion of the output signal voltage appearing at the collector can find its way into the resistor R_E. This leads to a decrease in overall gain and in order to prevent this happening, resistor R_E has to ignore any voltage at the signal AC frequency. It has to become an AC short circuit path to ground at signal frequencies. If a capacitor is placed in parallel with R_E, so that its X_C reactance at signal frequencies is very low, then this short circuit path is produced and the possible decrease in amplifier gain is prevented. In most practical audio amplifier circuits, such as those used to amplify small microphone voltage signals, a typical emitter bypass capacitor value is from 20 μF to 50 μF.

Our finished practical amplifier of Figure 31.10 has the following components and will act as a reliable small signal voltage amplifier:

T_1 = BC108 NPN	V_{CC} supply = 12 V DC
R_1 = 82 kΩ	$C_1 = C_{out}$ = 1 μF
R_2 = 15 kΩ	C_2 = 47 μF
R_C = 2 kΩ	R_E = 480 Ω

If this amplifier circuit is constructed it is usually found that values close to those calculated will still make the circuit work quite satisfactorily. This is an advantage of the voltage divider bias conditions. If the bias conditions for a transistor amplifier are not correct, then various forms of **distortion** of the output waveform can occur. Distortion, such as saturation and cut-off, can occur with incorrect resistor values, and total signal clipping

Ch31-Walding 4th 25/8/04 1:46 PM Page 721

can occur if the input voltage signal is too large in amplitude. Engineers designing critical amplifiers need to be careful not to introduce these forms of amplifier distortion as they will lead to eventual distortion of the sound output if, for instance, the amplifier is part of an audio system.

Refer to Figure 31.12. Notice also that the output signal in these common emitter class A amplifiers is always phase inverted. This means that if the input signal is a maximum at a particular time then the consequent output signal will be a minimum.

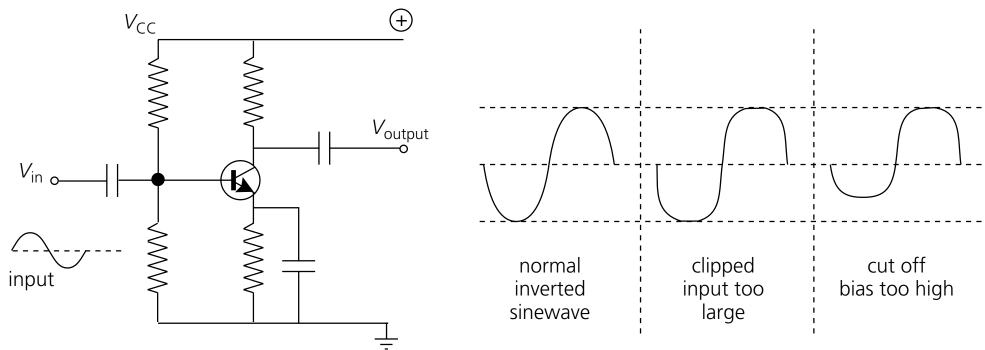


Figure 31.12
Amplifier output distortion.

The final amplifier design will increase small changes in input voltage to much larger changes in output voltage. This is called **voltage amplification** or voltage gain, A_V, where:

$$A_V = \frac{V_{out}}{V_{in}} = \frac{R_C}{R_E}$$

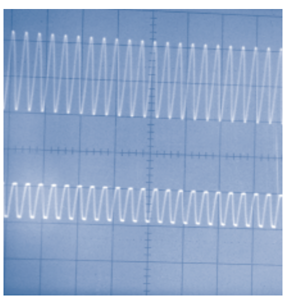


Photo 31.2
Signal amplification.

In circuits with an emitter bypass capacitor present:

$$A_V = \frac{R_C}{r_e}$$

where r_e is the internal emitter resistance. Its value is usually about 10–20 Ω so that voltage gain values, A_V, of 100–200 times are quite common. The voltage gain of an amplifier circuit is often easily measured with an oscilloscope and is usually quoted at a particular frequency of, say, 1000 Hz. The photo illustrates the voltage amplification of a transistor amplifier on a CRO screen.

Questions

4 Explain the operating conditions necessary to make a transistor function as a voltage amplifier. What bias method is commonly used?

5 You are designing a single transistor amplifier circuit using a device with a maximum collector current of 20 mA, β = 250 and voltage supply of 9 V. Using the design steps, calculate all values of the circuit components needed. Sketch your designed circuit and estimate its maximum voltage gain.

IC APPLICATION CIRCUITS 31.5

Linear IC applications

Let's continue our investigation of Op-Amp integrated circuits applications. Recall that, in Chapter 24, the Op-Amp chip was used in its inverting amplifier mode. This method of operation is similar to the discrete transistor AC voltage amplifier discussed in the previous

Ch31-Walding 4th 25/8/04 1:46 PM Page 722

section, but IC chip amplifiers are much more stable and are easier to use in electronics design. You should also realise that manufacturers often make integrated circuits with multiple operational amplifiers in the one chip package. Good examples are the LM324 Quad GP operational amplifier or the TL074 Quad JFET low noise operational amplifiers, where Quad refers to four amplifiers available in each chip package. Figure 31.13 illustrates the pin configuration of the LM324 Op-Amp chip in a 14 pin DIL package.

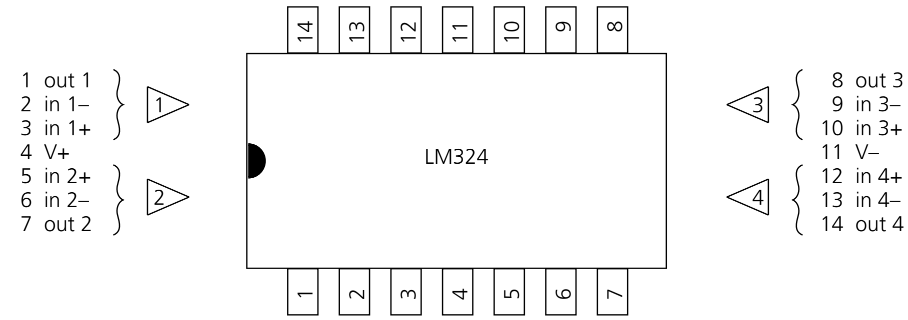


Figure 31.13
Quad Op-Amp LM324 chip.

Further application circuits of the linear operational amplifier IC are found as the wave function generator, adder and comparator as well as the audio amplifier. A brief discussion of each circuit type follows. Notice that quite often the Op-Amp chip requires a dual polarity power supply in these circuits.

Wave function generators

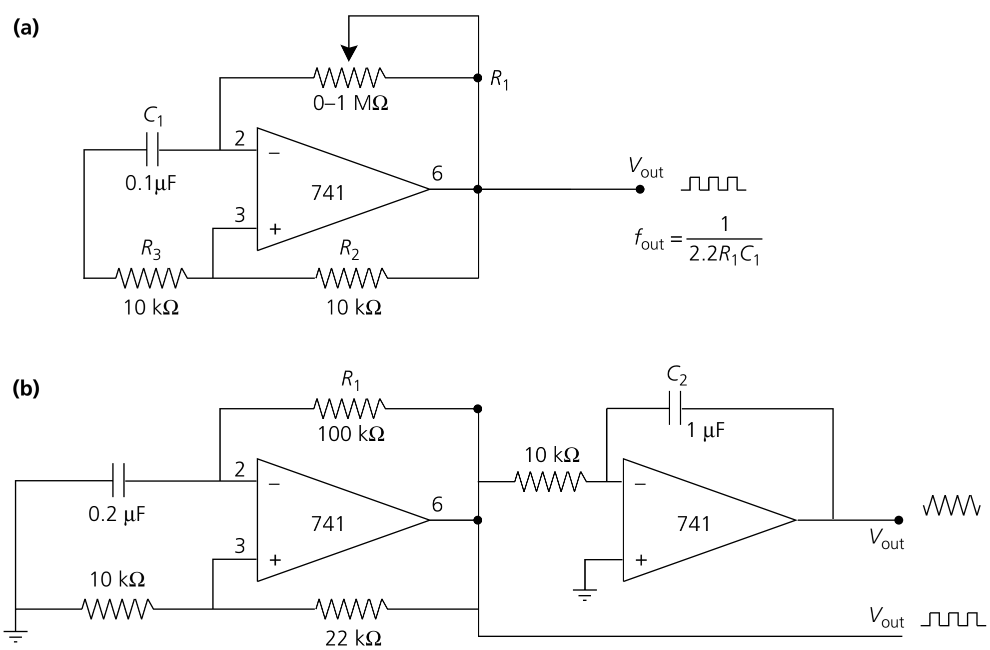
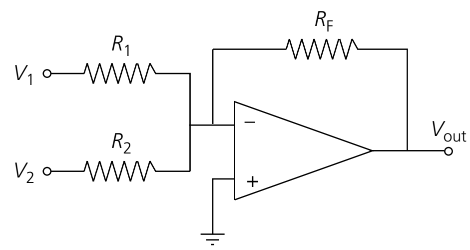


Figure 31.14
Wave function generator circuits.

The circuit of Figure 31.14**(a)** is using the 741 Op-Amp as a free running multivibrator circuit that produces a continuous square wave clock pulse. In circuit **(b)** the output of a similar circuit is connected to the input of a circuit that has a capacitor in place of the feedback resistor and has the function of carrying out an integration operation on the input voltage. This circuit will produce a second triangular wave output at pin 6 of the second 741 Op-Amp. Note the frequency formula for both circuits.

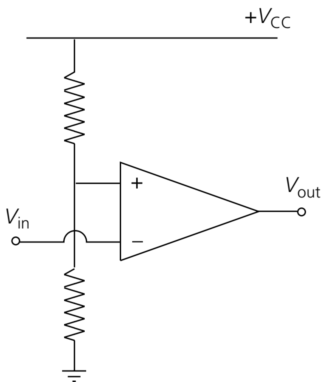
Adders and comparators

Figure 31.15
Op-Amp adder.



$$V_O = \left(\frac{R_F}{R_1}\right) V_1 + \frac{R_F}{R_2}(V_2)$$

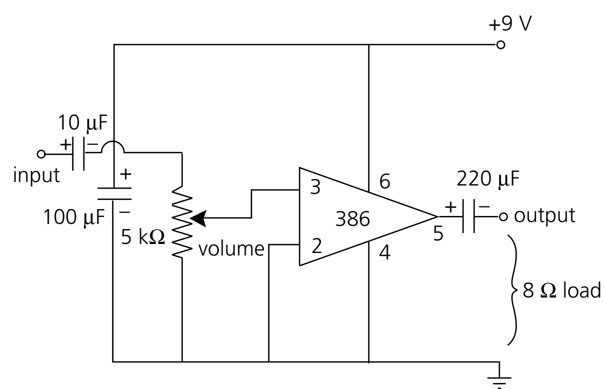
Figure 31.16
Op-Amp comparator.



Refer to Figure 31.15 which illustrates the Op-Amp used as an adder. It is simply an inverting amplifier with more than one input resistor. Resistors R1 and R2 control the amount of each voltage input that will appear added together at the circuit output. The comparator circuit of Figure 31.16 is used as a switch, which changes when a certain threshold voltage is reached, determined by the voltage divider resistors present at the non-inverting input of the Op-Amp. Voltage comparators drive the familiar bar graph LED displays in stereo amplifiers that indicate volume or recording level changes and contain a series of illuminating level indicators.

Audio amplifiers

Figure 31.17
Linear LM386 audio amplifier.



Before leaving linear IC circuit applications let's look at a very practical IC, the LM386, which is a low voltage audio amplifier that could be used as the basis of a construction project, say, as an amplifier to drive a small set of speakers from the headphone output socket of your portable stereo Walkman. Of course, a separate circuit would be needed for each channel of the available stereo output (Figure 31.17). The 5 kΩ potentiometer is used to control the volume because it subdivides the voltage from the input and applies the appropriate proportion to pin 3 of the chip. The internal gain of the LM386 chip is set to a value of 20. The circuit operates quite well from a simple 9 V DC power supply. The LM386 power amplifier is a very versatile integrated circuit that can supply about 400 mW of power into a typical 4 Ω speaker load impedance using the 9 V supply.

In the same category but able to supply higher power ratings, especially if properly connected to thermal sinks to dissipate heat generated, is the LM380 2.5 W amplifier IC and the LM1875 20 W amplifier IC. Usually these devices are driven from a properly designed regulated power supply. Again, pin configuration diagrams for these chips are very easy to obtain and the chips are relatively inexpensive to purchase.

Ch31-Walding 4th 25/8/04 1:46 PM Page 724

Digital devices and application circuits

Integrated circuits are used to perform a wide variety of functions in systems such as telephones, calculators and computers. Two basic classes of circuit types exist in digital electronics. Firstly, there are the **logic circuits**, which act as directional switches, latches and counters, and secondly, the **multivibrators**, which perform memory and timing functions. Both of these circuit classes can be represented by various **digital ICs**. Most of the digital ICs discussed in this section are manufactured in multiple unit packages in the form of 14 pin DIL chips. The series 4000 CMOS logic family are commonly used in circuit applications

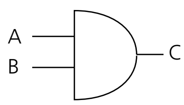
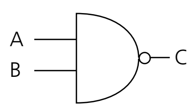
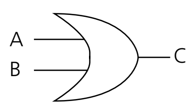
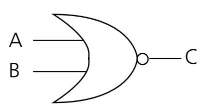
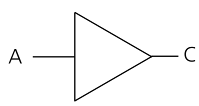
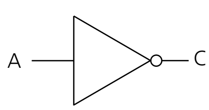
Logic gate	Circuit symbol	Truth table			Typical IC chip CMOS
		Input A	Input B	Output	
AND		0 0 1 1	0 1 0 1	0 0 0 1	4081 quad buffered 2-input
NAND		0 0 1 1	0 1 0 1	1 1 1 0	4011 quad 2-input
OR		0 1 0 1	0 0 1 1	0 1 1 1	4071 quad 2-input
NOR		0 1 0 1	0 0 1 1	1 0 0 0	4001 quad 2-input
Buffer		0 1		0 1	4050 Hex-non inverting
Inverter (NOT)		0 1		1 0	4049 Hex-inverting

Figure 31.18
Digital gates.

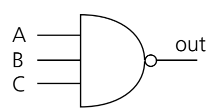
with both quad (4) or hex (6) multiples available in the one IC package. Manufacturers make available IC pin configuration diagrams that describe the necessary input, output and power supply pin connections.

No matter how complicated, all digital ICs are made from simple building blocks called logic gates or just gates, which are the equivalent of electronic switches. The circuits are called logic gates because they make logical decisions with the output state being dependent on the input states. Let's look at the basic set of logic gates, their input–output characteristics or **truth tables**, and typical IC packages that contain them. (See Figure 31.18.) Remember that 1 = digital ON or HIGH and 0 = digital OFF or LOW. In actual voltage terms, a digital high represents the chip supply voltage and a digital low represents ground or zero volts. Refer back to Chapter 24 (Section 24.2) for discussion of ADC.

Logic gates may have more than two inputs, which increases the decision-making power of the gates. Multiple input gates can increase the number of ways that the logic functions can be interconnected to form advanced digital logic circuit blocks. For example, Figure 31.19 shows the truth table for a 3 input NAND gate.

Figure 31.19
Three-input NAND gate.

Symbol

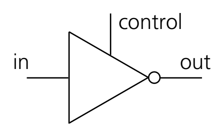


Truth table

A	B	C	Out
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

Figure 31.20
Three-state inverter.

Symbol



Truth table

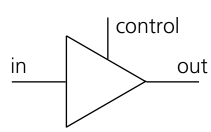
Control	In	Out
0	0	1
0	1	0
1	X	Hi-Z

X means either 1 or 0

Often the single input inverter designed to reverse the output from another gate can have an extra control input, in which case it is called a three-state inverter (Figure 31.20). Also, the buffer gate, which is a circuit that isolates gates from each other electrically or provides outputs that are capable of driving higher resistance loads, can have a control input as well. This gate is called a three-state buffer (Figure 31.21). Three-state buffers and inverters have an output that can be electronically disconnected from the rest of the circuitry. The output then is neither high nor low, but 'floats' and appears as a very high resistance.

Figure 31.21
Three-state buffer.

Symbol



Truth table

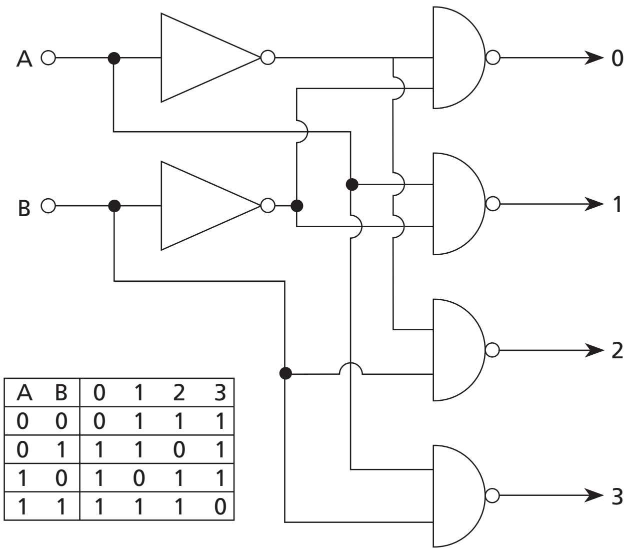
Control	In	Out
0	0	0
0	1	1
1	X	Hi-Z

X means either 1 or 0

Logic gates can be interconnected in a network of gates referred to as logic circuits. The methods of logic circuit connections can either be combinational or sequential.

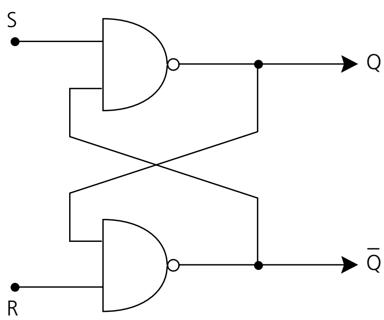
Combinational logic circuits respond to incoming data pulses immediately and their decisions do not depend on a series of previous logic events. Any combinational circuit can be constructed using the basic NAND and NOR logic gates. For example, the combinational logic circuit of Figure 31.22 and its truth table illustrate the process of converting a two-bit binary number to its decimal equivalent. Advanced combinational networks are often available as separate digital ICs; for example, data selectors (multiplexers) and digital encoders–decoders, such as the 74HC154 CMOS 4 to 16 decoder/demultiplexer chip.

Sequential logic circuit outputs are determined by the previous states of the circuit's inputs. That is, data bits move through the circuits step-by-step. Often a separate clock pulse



A	B	0	1	2	3
0	0	0	1	1	1
0	1	1	1	0	1
1	0	1	0	1	1
1	1	1	1	1	0

Figure 31.22
Binary to decimal decoder (BCD).



S	R	Q $\bar{Q}$
0	0	Disallowed
0	1	1 0
1	0	0 1
1	1	No change

Figure 31.23
RS Flip-flop (latch).

(square wave) signal is required in order to make the data bits move. The basic building block of sequential logic circuits is the **Flip-flop**. A basic reset-set Flip-flop circuit (RS Flip-flop), also called a **latch**, is shown in Figure 31.23, together with its truth table. The outputs Q and $\bar{Q}$ are always in opposite digital states. Flip-flop circuits can become quite complex and form the basis of counters and registers as used in computer circuits.

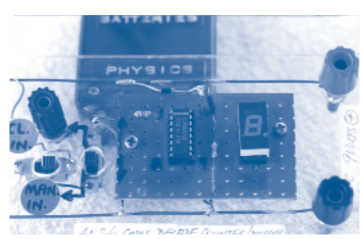
Circuit designers use **Boolean algebra** statements to show the respective outputs of logic gates. This algebra was invented by an English mathematician, George Boole, in the middle of the nineteenth century as a system of analysing logic statements mathematically. (See Figure 31.24.)

- The Boolean operation 'AND' is represented by a dot (●).
- The Boolean operation 'OR' is represented by a plus sign (+).
- The Boolean operation 'NOT' or inversion is represented by a bar over the letter ($\bar{A}$).

Combinations of the Boolean operations are also possible; for example, the NAND gate produces a combination of NOT and AND, whereas the NOR gate produces a combination of NOT and OR.

Figure 31.25 illustrates two circuits that use chips combining both combination and sequential logic circuits. The block diagram shows the simple decimal **counter** system, which is able to count incoming clock pulses from a circuit such as the clock of Figure 31.14. The BCD counter advances one count for each incoming pulse. When the count reaches binary (1001) or decimal (9) the counter recycles to 0000. The decoder activates the appropriate segments of a seven-segment LED display so that the count is obtained. Notice that a single chip can replace the function of both counter and decoder driver. This is the CMOS 4026 chip. Notice that this chip has a carry out (CO) pin that allows counts to other 4026 chips in order to cascade them together so that a total count of hundreds or thousands is possible. (See Photo 31.3.)

Photo 31.3
4026 counter chip.



Ch31-Walding 4th 25/8/04 1:47 PM Page 727

Figure 31.24
Boolean operators.

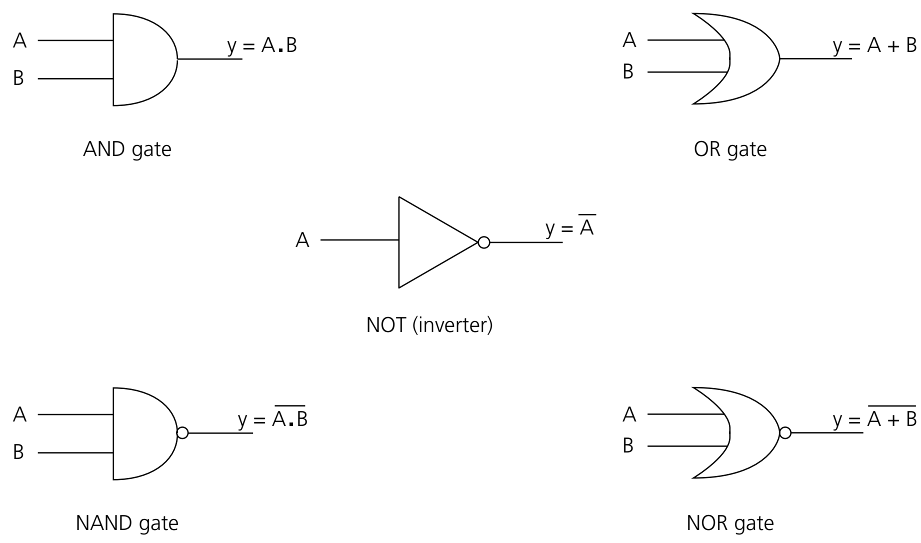
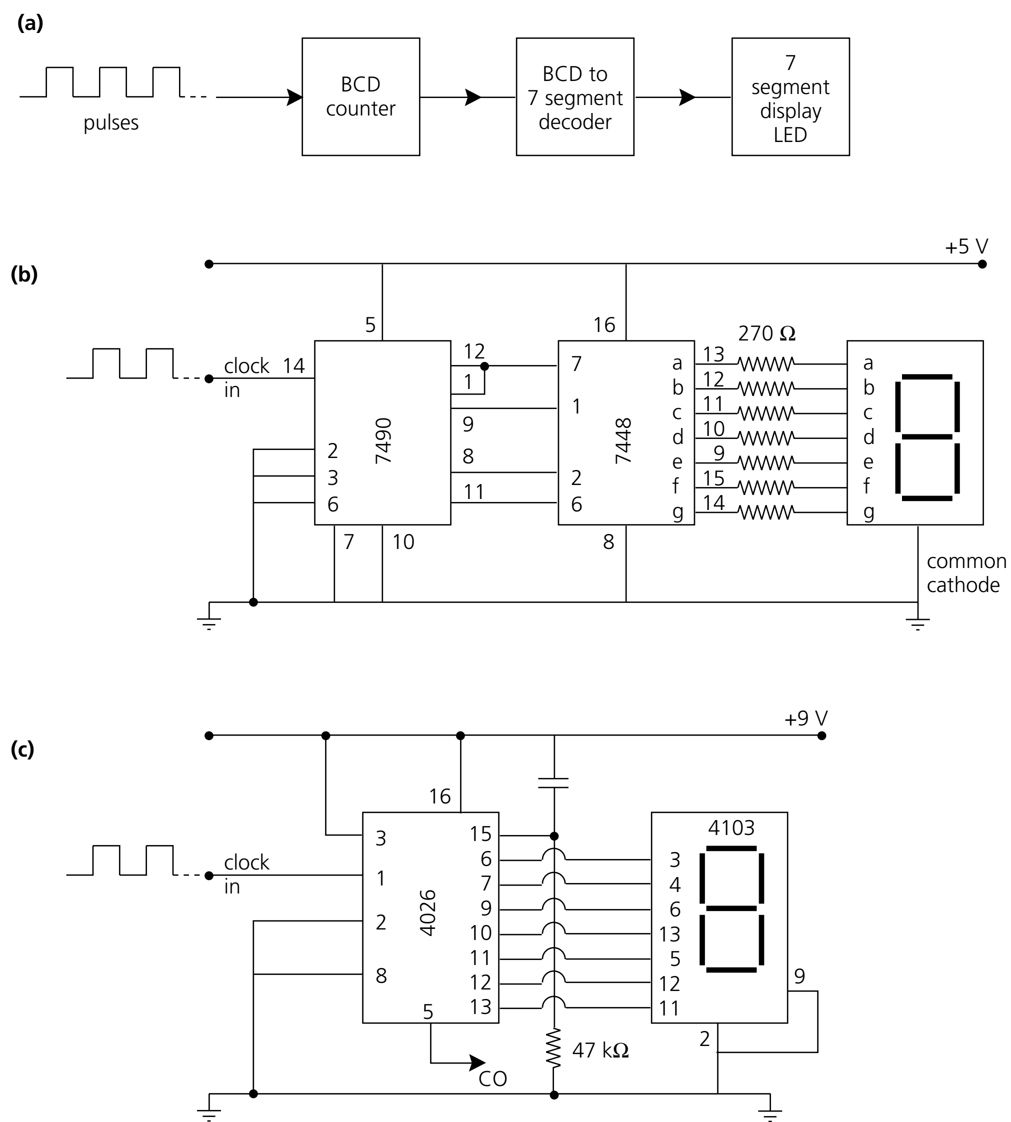


Figure 31.25
Decimal counting system: **(a)** block diagram; **(b)** separate chip circuit, 7490 + 7448; **(c)** single chip circuit, CMOS 4026 decade counter-display driver.



It is great fun to work with these digital counter and display driver chips. Using a protoboard and simple LEDs to indicate output states of gates or logic circuits, much experimentation can be done. Always be careful to observe correct power supply connections to all IC chips and handle them with care, especially if the type you are using is a CMOS design. An important circuit building tip for use with all digital logic circuits is to ensure that the inputs and outputs of unused gates in an IC package are connected to earth. This will avoid spurious noise signals triggering the gates falsely.

Questions

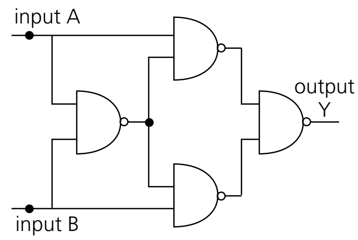
6 Select a simple application for a 741 Op-Amp. Sketch the circuit used including chip power supply and explain the operation of the circuit.

7 Draw the symbols used to represent the following logic gates: **(a)** 2 input AND gate; **(b)** 3 input NAND gate; **(c)** 2 input OR gate; **(d)** 3 state buffer.

8 The output of digital logic gates can be summarised with Boolean algebra statements. Which gates correspond to the following statements in Boolean algebra?
(a) A + B; **(b)** A●B; **(c)** $\overline{A + B}$; **(d)** A●B●C; **(e)** $\overline{A}$.

9 Figure 31.26 illustrates a combination of NAND gates. Deduce the truth table for the combination and decide if it is equivalent to any single logic gate.

Figure 31.26
For question 9.



Activity 31.2 BUILDING WITH ICs

If you have access to circuit building boards especially designed for ICs such as the SK40 protoboards, then your teacher may supply you with some integrated circuits as described in this text so that you can actually build some of the circuits. Remember to always connect power supply voltages to the correct pins of the IC and connect the power supply last of all. If you are using multiple gate digital chips always connect unused gate inputs and outputs low to avoid false triggering.

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***10** Explain the differences in AC circuits between resistance, capacitive and inductive reactance. How are these concepts linked to impedance?

***11** For the circuit of Figure 31.27, calculate **(a)** capacitive reactance, X_C; **(b)** total circuit impedance, Z; **(c)** the RMS current flowing; **(d)** the phase angle between current and voltage.

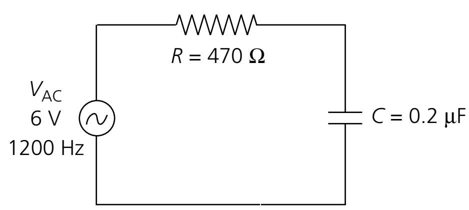
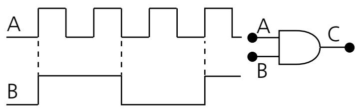


Figure 31.27
For question 11.

****12** Explain the function of each of the resistors used in a voltage divider common emitter class A transistor amplifier. Draw the common circuit diagram for this amplifier, marking all necessary connections including power supply.

****13** Calculate all necessary circuit component values in the design of a single transistor amplifier if it is to operate with the following parameters: V_{CC} = 12 V, β = 280 and I_C (max) = 18 mA. Sketch your designed circuit and estimate its voltage gain.

Figure 31.28
For question 14.



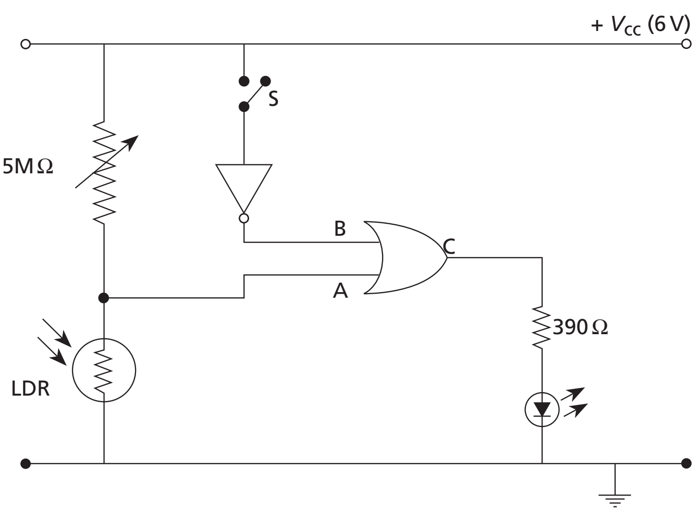
***14** Figure 31.28 illustrates the input waveforms applied to an AND gate. Correctly draw, at the same scale, the output waveform of this gate.

****15** An AC source (f = 100 Hz) is connected in series with a resistor of 1000 Ω.
(a) Draw a circuit diagram.
(b) If the peak voltage is 12 V, plot a curve (V–t) for three cycles of the waveform. What is the peak current value?

****16** Suppose you wish to build an AC powered battery replacement circuit for your portable Walkman player so that it can be used at your study desk. Draw a circuit that would provide the necessary 9 V output. Your circuit must be capable of supplying about 120 mA of smooth output current.

****17** Figure 31.29 illustrates a circuit that could be used as an automatic night light switch. The circuit automatically switches on the LED when darkness falls or when the sensor is covered.
(a) Identify each electronic component used in this circuit and make a listing.
(b) When it is dark how does the resistance of the LDR alter?
(c) Explain how the circuit works!
(d) What is the purpose of the potentiometer at the input A of the circuit?
(e) What is the function of the switch in the input B of the circuit?

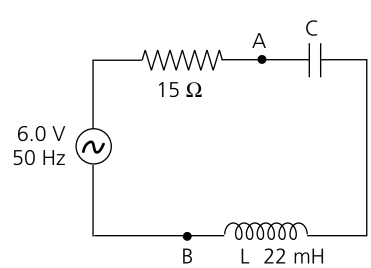
Figure 31.29
For question 17.



Extension — complex, challenging and novel

*****18** The circuit of Figure 31.30 contains an AC source of 50 Hz, 6.0 V. If the capacitor has a value of 65 μF:
(a) what is the total impedance, Z_{AB};
(b) what is the total circuit impedance if the source internal impedance is 5 ohms;
(c) what is the RMS current flowing in the circuit;
(d) what is the value of RMS voltage between points A and B?

Figure 31.30
For question 18.



*****19** Consider Figure 31.31. Recall that a magnet moved inside a coil of wire will generate a small AC voltage. Look at the circuit diagrams supplied and analyse them to answer these questions:

(a) How is the simple rectifier power supply working?

(b) Why is it needed in this circuit application, and how is it connected?

(c) How is the Op-Amp modifying the signal produced when the magnet is moved into the multi-turn coil?

(d) What will occur when the circuit variable resistor is adjusted?

(e) To what range would the multimeter need to be set?

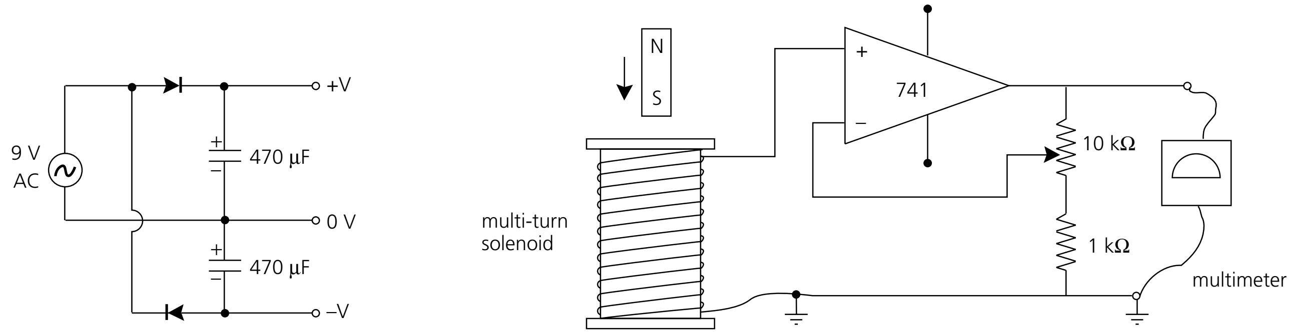


Figure 31.31
For question 19.

****20** In Boolean algebra, an exclusive OR (XOR) function is represented by the plus symbol inside a circle (⊕). An XOR gate produces a digital high (1) output only when one of its inputs is high (1). If both inputs are either high or low then the output is low. Draw a logic gate circuit diagram that would represent the following Boolean algebra statements:

(a) $Y = \bar{A} + \bar{B} + \bar{C}$

(b) $Y = \bar{A} \oplus \overline{B \bullet}$

(c) $Y = \bar{A} \oplus \bar{B}$

*****21** Consider Figure 31.32, showing the circuit of a Quad NAND gate CMOS IC, the 4011 operating from a 9 V DC supply. The circuit is one method of testing the truth table for the NAND logic gate. Analyse the circuit given and answer these questions:

(a) Would the 9 V DC supply need to be regulated?

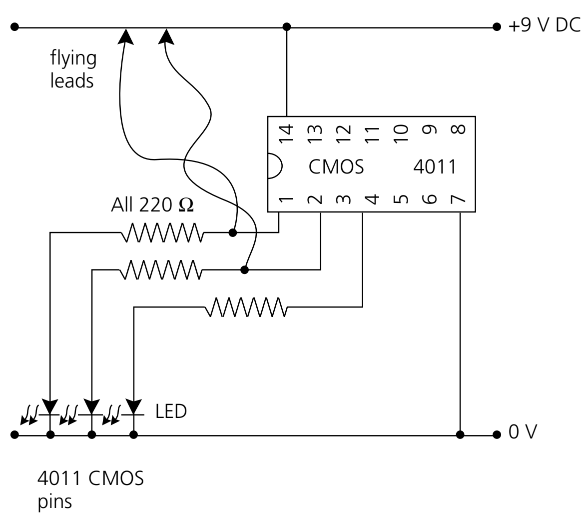
(b) How many different gates could be tested?

(c) What is the purpose of the flying leads?

(d) Why are the resistors needed?

(e) Explain how the three LEDs display the input and output combinations of the gate. Draw a truth table.

(f) Could this circuit be used to test other gate types? Explain.



1	in A1	8	in B3
2	in B1	9	in A3
3	out 1	10	out 3
4	out 2	11	out 4
5	in A2	12	in B4
6	in B2	13	in A4
7	V_{SS} (0 V)	14	V_{DD}

Figure 31.32
For question 21.

CHAPTER 32

Solar Physics

INTRODUCTION 32.1

As you awake one morning, a cheery radio weather presenter announces:

> The weather forecast this morning is for moderate to high temperatures, sunlight early yielding 620 watts per square metre rising to 980 watts per square metre on the coast and 1100 watts per square metre inland. Those in southern regions will require battery reserves.

In the future, this scenario could prove true for communities that have their electrical power generation supplemented by solar energy. The Earth's star is a self-sustaining nuclear fusion reactor whose output is an incredible 4×10^{26} W, of which central Australia receives only about 1.0×10^{3} W m^{-2} on the ground. If the nuclear reactions in the Sun's core were to be switched off now, it would be 10 million years before the outer solar surface started to cool and before the Earth would feel the effects. Such is the power of the Sun!

Even animals are affected by solar processes. On 7 July 1988, 3000 homing pigeons were released from cages in northern France for their annual race back home towards southern England. Two days before, unusual solar flare activity had sent vast clouds of charged protons and subatomic particles into space, some of which disrupted the Earth's magnetic field patterns. In poor weather the pigeons used internal magnetic compasses to guide them. Misled by the solar disturbances caused to the Earth's magnetism, the pigeons flew way off course. Most of the 3000 never returned!

Solar physics is the study of the Sun's energy processes and the ways in which modern technology can use both its heat and its light. One of the best examples of solar technology assisting engineering was the feat of American aeronautical engineer Paul MacCready, who designed the famous *Solar challenger* human-powered aircraft. In July 1981, Steve Ptacek pedalled this aircraft, whose wings were covered with solar cells on their upper surfaces, over 262 km from Cormeilles en Vexin near Paris, across the English Channel to Manson in Kent.

In this chapter we will look at the Sun itself, the major methods of obtaining energy from the Sun, both passively and actively, as well as one of the main dangers to the health of all Australians, namely, ultraviolet radiation.

SOLAR RADIATION 32.2

The Sun as a star

Our Sun, called *Sol*, dominates the planetary system that includes the Earth. The Sun provides the input energy for most of the food webs that make up our natural environment. The Sun radiates energy at the tremendous rate of 4.0×10^{26} W, of which the Earth receives approximately 1.8×10^{17} W at its outer atmosphere. About half of this actually reaches the ground and provides the driving energy for our climate and weather systems as well as the photosynthetic requirements of plants as the food chain producers.

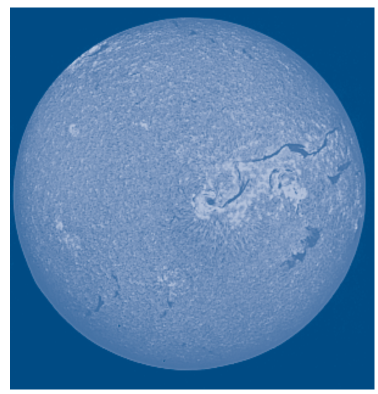
The Sun has been studied scientifically since the time of Galileo (1611), who used the first telescopes to discover sunspots on its surface. Table 32.1 lists the physical data of the Sun. It is a very average star by comparison with those in the rest of the universe and is about half-way through its lifespan — middle-aged you might say, with only about 4.5 billion years left to keep radiating its energy. Our next nearest stellar neighbour is the bright star in the Centaurus (pointers) constellation called Alpha Centauri, at 4.3 light-years distance.

Table 32.1 SOLAR DATA

Diameter	1.39×10^9 m	Spectral type	G
Mass	1.99×10^{30} kg	Mean distance	149 597 000 km (8.3 light-minutes)
Specific gravity	1.409	Rotation period (equatorial)	25.38 days
Axial inclination	7° 15′	Absolute magnitude	+4.71
Effective temperature (black body)	5800 K	Escape velocity	617.5 km s^{-1}

In 1814 the German physicist Joseph von Fraunhofer used a spectroscope to break the Sun's light radiation up into its component wavelengths and examined it carefully. Recall that Isaac Newton had also performed spectral dispersion with a prism as early as 1666. Fraunhofer's spectral analysis enabled an explanation of the solar atmosphere. He found that the continuous emission spectrum of the Sun was crossed by a complex set of dark lines. It was Gustav Kirchhoff in 1859 who showed that these dark **Fraunhofer lines** were actually absorption lines caused by atoms present in the low pressure solar atmosphere lying between the Sun's source of light, called the **photosphere**, and the experimental observer. By comparing these lines with the emission spectra of known elements on Earth, it was shown that the Sun itself contains most of the known elements. The inert gas helium was named because of this. In 1889 the American astronomer George E. Hale invented the spectroheliograph (Greek *helios* = 'Sun'), which enabled the Sun to be studied in the light emitted by one element alone, such as the light of hydrogen or calcium. Today, spectral filters can do much the same job.

Photo 32.1
Sun's surface features.



Like all stars, the Sun is composed mostly of hydrogen, together with helium (27%) and other heavy elements (2%). At the core of the Sun, the temperature is 16 000 000 K and has a density about 150 times that of water. In these conditions hydrogen nuclei fuse to produce helium via reactions called **nuclear fusion**. The solar nuclear fusion process is actually a series of three collisions between atomic particles called the 'proton–proton cycle'. The three collision processes are not of equal probability but end up fusing four hydrogen nuclei (protons) into one helium nucleus. As the final helium nucleus is only 99.3% as massive as the original four protons the missing energy appears as gamma rays and a neutrino according to Einstein's $E = mc^2$ formula. The neutrino is totally unreactive and escapes the Sun very quickly, while the gamma rays may bounce around internally for millions of years. Even though the timescale for the proton–proton cycle is large, the staggering number of particles in the Sun means a massive amount of total energy is continuously released via gamma rays that radiate outward toward the convection layers of the Sun. It is in this outer one-third of the Sun's volume that large-scale convective turbulence not only reduces the temperature but produces most of the Sun's radiation energy. The photosphere is the top surface of these convection cells, which give it a mottled appearance called **solar granulation**. The granulated cells on the photosphere last typically for 5–15 minutes and are about 2000 km in diameter. The temperature of the photosphere is about 6000°C. The German-born American physicist Hans Bethe was awarded the 1967 Nobel prize for physics for his work on the fusion cycle reactions that are the source of the Sun's tremendous energy production.

Sunspots

Sunspots are smaller regions of the photosphere that are, on average, 2000°C cooler than their surroundings and depressed into the surface slightly. They tend to occur in pairs. In 1908 George E. Hale discovered that they contain very strong magnetic fields, with each pair containing magnetic flux that points in opposite directions, either into or out of the Sun's interior. A sunspot cycle occurs, in which the number of sunspots varies from a maximum to a minimum over an 11-year cycle. As a new 11-year cycle begins, the magnetic polarity of the leading sunspot of each pair in each hemisphere of the Sun's surface reverses. This represents a full solar cycle of 22 years. Sunspots forming early in the cycle in each hemisphere tend to start at higher latitudes (45°) than sunspots later in the cycle (10°). These cyclic changes seem to indicate a definite connection between the Sun's magnetic field, the convection zone in the Sun's outer layers, and the Sun's rotation period itself, which is faster at its equator than at its poles.

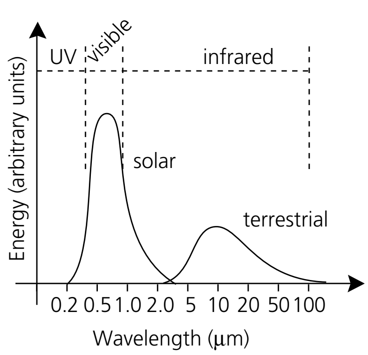
Solar flares and prominences

The Sun's **chromosphere** (Greek *chroma* = 'colour') rises to about 9600 km above the photosphere, with an average density about one thousand times less than the photosphere and a temperature of about 30 000 K. Elements in the chromosphere absorb light and produce the Fraunhofer lines in the Sun's emission spectrum. The upper layer is not uniform but produces **spicules** or high temperature gas plasma eruptions that are continuously penetrating the outer layer or **corona** (Greek *corona* = 'crown'). Because of the continuous agitation, plasma particles are being thrown off into space, causing the **solar wind**, which eventually reaches the Earth. Near sunspots the chromosphere radiation is more active, producing very rapid releases of magnetic energy and plasma particles called **solar flares**. Among the phenomena that accompany solar flares are intense X-rays, radio waves and other energetic particles that may also eventually reach the Earth to cause auroral displays and disrupt radio and telecommunications services.

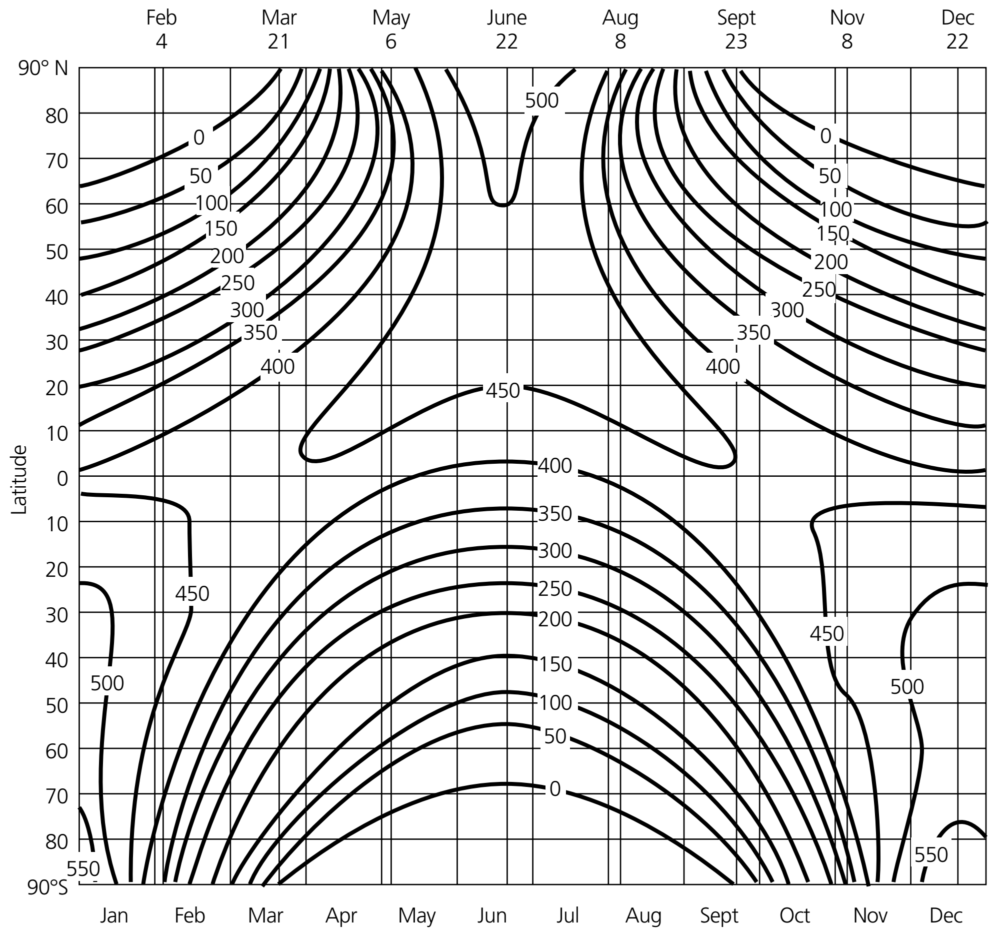
The corona extends for several solar radii from the disc of the Sun itself. All the structural detail within the corona is due to the solar magnetic field. The corona is at a very high temperature of about 1 000 000°C indicating very high particle velocities. Occasionally, the corona traps low temperature plasma emissions on a large scale from the chromosphere. These produce **prominences**, which may extend out from the Sun's surface for hundreds of thousands of kilometres and are best seen during periods of solar eclipses at the edge of the Sun's disc. These prominences can also release tremendous numbers of particles into the solar wind. The largest recorded could have swallowed the Earth many times over.

Solar radiation at the Earth

Figure 32.1
Solar and terrestrial energy emissions.



The Earth receives a constant energy flow from the Sun of about 1.23×10^{17} W (122 500 TW). As the Earth gains thermal energy its temperature will rise but, like any hot body, its rate of energy emission also increases with temperature. If the received and emitted thermal energies were equal in wavelength this would lead to an average equilibrium temperature of about −17°C for the Earth. Fortunately, the Earth reradiates its thermal energy at much longer wavelengths, as shown in Figure 32.1. These longer wavelengths are absorbed by the atmospheric water vapour and carbon dioxide. This absorbed energy is reradiated with about 85% of it returning to further heat the Earth to an average global value of 15°C (288 K). This effect is called the **greenhouse effect** and without it most life forms on Earth would die. The commonly held view that the greenhouse effect is bad stems from a misunderstanding of the basic effect. Modern technology needs to be applied to reduce the emissions of greenhouse gases into the atmosphere. This will prevent an increase in the natural greenhouse effect, which would lead to a rise in average temperatures or **global warming**. The planet Venus has an atmosphere of dense carbon dioxide, which produces a surface temperature of 470°C through its natural greenhouse effect. Let us hope the Earth never gets to this point.



The variation with latitude and time of year of the maximum amount of solar radiation (in kW m^{-2}) received on a horizontal surface at sea level

Figure 32.2
Variation of solar radiation.

Figure 32.2 illustrates the variation with latitude and time of year of the maximum solar radiation received (kW m^{-2}) on a horizontal surface at sea level. The global average value is about 200 W m^{-2}. The lack of symmetry between the two hemispheres is due to the slight ellipticity of the Earth's orbit, which results in the shortest distance to the Sun (perihelion) occurring on 4 January and the greatest distance (aphelion) on 5 July. At any particular location the solar radiation flux is determined by the time of day, the season, and the geographical latitude. Weather provides an unpredictable element. Under clear skies the radiant energy will be mostly direct, with only about 15% being diffuse, while under overcast conditions, obviously 100% of the radiant energy is diffuse. Australia is particularly lucky, with its solar radiant energy flux being more direct and constant than most countries, with peak maxima rising as high as 1.4 kW m^{-2}.

In order to exploit solar energy, designers of solar devices must contend with several factors:

- relatively low power density available
- variation in the availability of solar radiation
- low efficiency of various conversion techniques to other energy forms such as direct thermal or electrical.

In the rest of this chapter, we will examine some of the ways in which solar energy can be converted into other forms. These will primarily be by **photothermal** techniques concerned with the active and passive collection of solar radiation as heat, and **photovoltaic** techniques concerned with the direct conversion of solar radiation into electricity.

Questions

1 If the average solar radiation at a Western Queensland location is 600 W m^{-2}, calculate the total solar collector area needed at this site to generate 100 MW, assuming solar–electrical conversion efficiency of 9%.
2 Explain the interaction of the solar radiation with the Earth's atmosphere that allows a global average temperature equilibrium of 15°C.
3 Greenhouse gases include carbon dioxide, methane, and nitrous oxide. Research their predominant sources, both artificial and natural, and comment on their effect on global warming.
4 List reasons for the fact that about 30% of incident solar radiation is reflected back into space from the Earth.
5 Discuss any link that exists between these three statements of fact:
- The 1987 Montreal protocol called for a halving of chlorofluorocarbon (CFC) emissions by the end of the twentieth century.
- Stratospheric ozone blocks incident ultraviolet solar radiation that is less than 300 nm in wavelength and dangerous to living organisms.
- Since about 1975, marked ozone depletion has occurred in the stratosphere over Antarctica each spring — the 'ozone hole'.

6 Assume you are sunbathing at noon in summer when the direct solar radiation flux is 850 W m^{-2}. Estimate the total radiation incident on your body in 20 minutes. How would this change at a later time when the Sun is 50° above the horizon?
7 Using Figure 32.2, what is the difference in solar radiation received at a latitude of 60° on the date of 8 August in both hemispheres? What accounts for this difference?

PHOTOTHERMAL DEVICES 32.3

Architectural design

How often have you noticed the build-up of heat and general stuffiness of a closed room with sun shining into it through the windows? Have you ever used a magnifying glass to burn holes in paper? These two effects illustrate the **passive** and **active** aspects of solar heating. The term **photothermal** device is used to describe a device that passively or actively converts solar radiation into heat. Architects today are very interested in aspects of **passive solar design**, which uses the materials of the house itself as well as modifications to its surroundings to efficiently collect, store and distribute solar energy, thus allowing large electrical energy savings in the general heating and cooling systems. It represents a very natural approach.

Architects may also use **active solar design** systems, which typically use solar collectors for heating water, thermal masses such as rock-beds and Trombe walls, as well as forced ventilation of solar heated air. These will be discussed later in the chapter. The Australian CSIRO has been very active in solar design elements for general housing, as well as showing the way for larger industrial applications, such as hybrid solar/diesel power stations and solar power towers. At Highett, in Victoria, CSIRO maintains a low-energy-consumption home (LECH), which is used for solar principle demonstrations and practical research.

Consider Figure 32.3, which illustrates a range of passive solar design architectural aspects. Let us take each of the labelled parts in turn.

1 Considerably larger window area on the northern side This aims to catch sunlight in winter, and also, by incorporating verandahs, pergolas or overhang, to provide shade from the higher summer sun. In Queensland especially, even early colonial houses tended to have wide cooling verandahs. The roof design contains insulated fibreglass batts or loose fill in the ceiling with reflective aluminium foil against the undersides of the tiles. Note also the

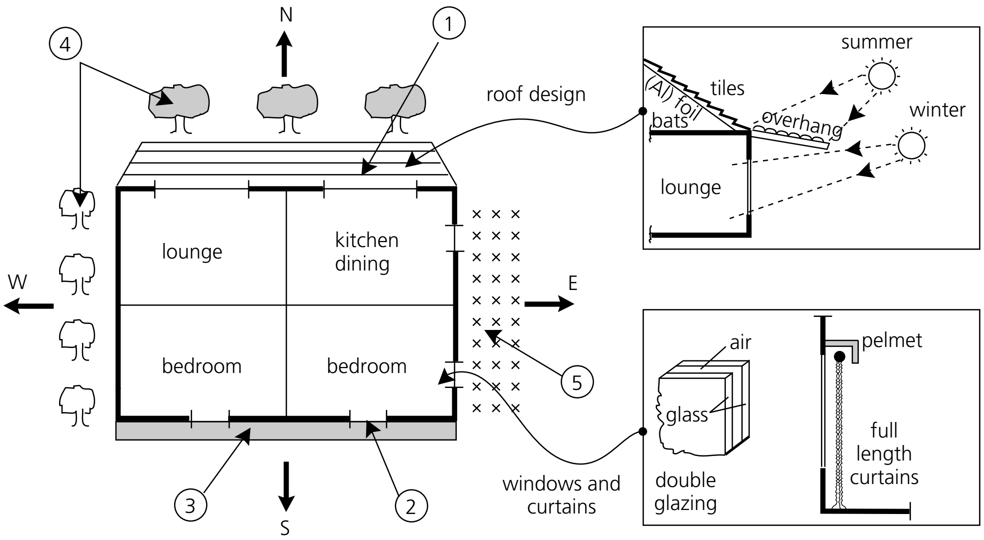


Figure 32.3
Passive solar design aspects for domestic housing.

general living areas of the house are on the northern side, while the bedrooms, for nightly warmth, are on the southern side.

2 Minimum window area on the southern side This will reduce energy loss in winter. Windows are double glazed with the use of heavy curtains and pelmets to prevent direct air flow against the cold windows.

3 Heavy concrete base for thermal stability and mass storage Brick walls often with a cavity (double brick) for air insulation are used. The large base thermal mass heats slowly on hot days and will not lose heat quickly on cold nights. Ideally there are no windows at all on the western side to block entry of afternoon sun and cold winter winds. The flooring of most rooms should assist thermal insulation, especially in southern states, with thick underlay and carpets.

4 External planting of evergreen trees and shrubs on the western side This provides sun and wind-breaks. Deciduous trees provide shade in summer at the front of the house, but lose leaves in winter to let in the light and heat from the Sun.

5 Minimum window area on eastern side Pergolas or awnings are also used. Ground cover plants or grass, rather than reflective concrete, are used to prevent morning light reflection, especially during summer months.

Consider Figure 32.4, illustrating the principles of active solar design. In part **(a)**, the house design incorporates a large thermal mass rock-bed base down into which is fan-forced hot air from the roof-top solar collector. During cold nights the rock-bed slowly releases its stored thermal energy. In part **(b)**, the house design incorporates a **Trombe wall**, which can be thought of as a solar operated storage heater. A thermally massive blackened wall is placed behind glazing on the north-facing side of the house. The diagram shows how both cool and warm air circulate through the house via ducts and shutter flaps, depending on whether house cooling or heating is required. In some systems the Trombe wall is filled with water to achieve the same degree of thermal storage capacity.

Solar hot water systems

Most household solar heating is achieved with **flat plate collectors,** as shown in Figure 32.5 and Figure 32.6. A number of collectors are placed onto the roof (north-facing) and attached to the hot water plumbing system. The maximum operating temperature for these collectors is about 80°C. If it is made to operate at higher temperatures, radiation and convective losses increase dramatically. CSIRO developed a Teflon strip system to help to reduce convective losses in these flat plate collectors. In this system a series of parallel vertical thin strips of Teflon film run up and down the slope of the solar collector about 5 mm apart between the glass cover and the absorber. This addition increased operating efficiency considerably.

PHYSICS FACT

The Australian National University (ANU) hopes to build the world's largest combined solar hot water and electricity system for one of its own buildings, called Bruce Hall. In the developing system, built in conjunction with Rheem Australia-Solarhart, Sun-tracking parabolic mirrors concentrate sunlight by a factor of about 30 times and shine it onto thermal solar receivers mounted with solar cells that convert the sunlight directly into electricity with about 20% efficiency. The system is called CHAPS (combined heat and power solar) and will completely supply the building with electricity and hot water.

Figure 32.4
Active solar design: **(a)** solar air heating and rock-bed; **(b)** Trombe wall design.

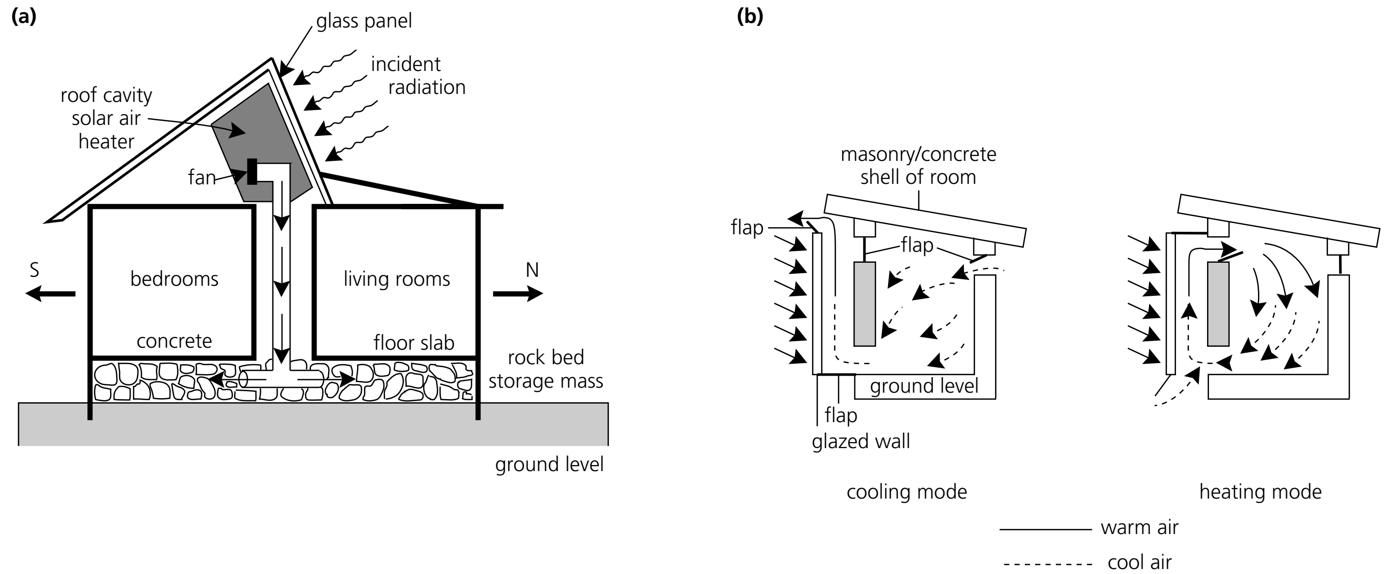


Figure 32.5
Flat plate solar collector — cross-section.

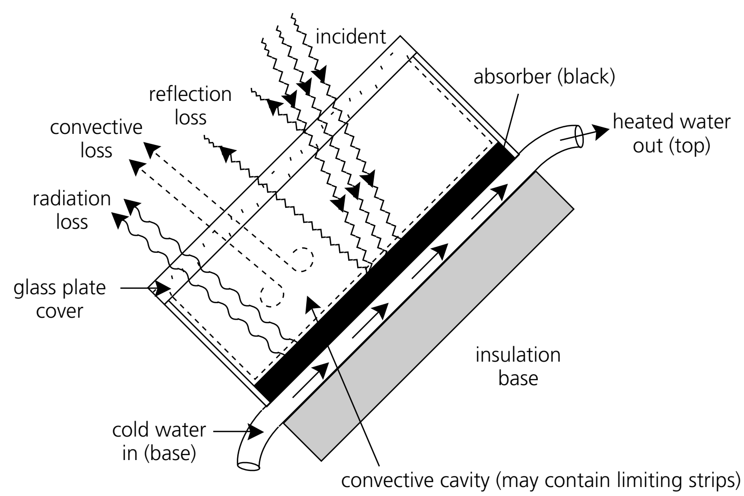
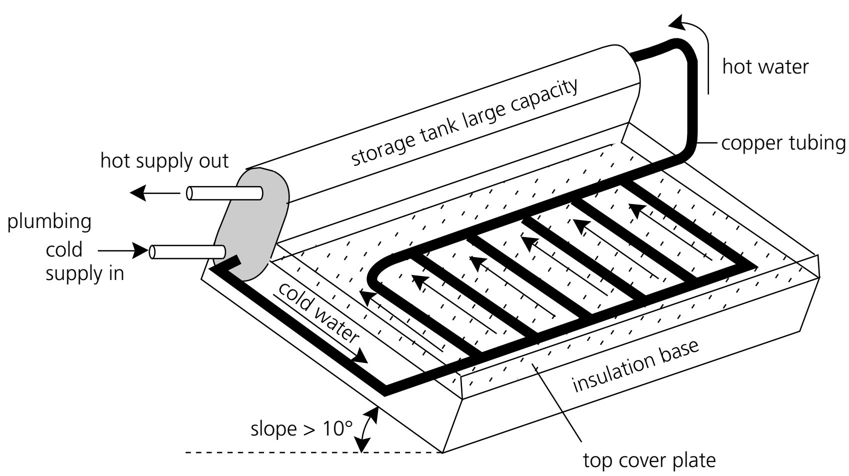


Figure 32.6
Thermosyphon solar water heater.



In Australia, the commonest design of solar hot water system uses the **thermosyphon** principle. A solar collector mounted on the roof is connected to a horizontal cylindrical storage tank. Cold water flowing into the collector pipes is heated and displaced upward by the cooler, more dense water. This natural convection flow produces a thermosyphon flow without the need for artificial pumps. The top of the collector sits below the bottom of the storage tank and the roof slope must exceed 10° for the thermosyphon effect to be efficient. In winter months a booster electric element situated in the top of the storage cylinder can be switched on to complete the heating of the top layers of water, if necessary. The solar collector should be angled to face the Sun. The angle is approximately equal to the latitude, so in Brisbane they are angled at 30°, while in Cairns the angle is 17°.

Although solar heating is becoming more efficient with design improvements, it is still only marginally economical in many countries where low levels of sunlight are received. Even in Australia, to power large systems with flat plate solar collectors is an unlikely future proposition. Larger industrial photothermal mechanisms that are in use in various parts of the world are called concentrating collectors and include power towers, solar farms, heliostat arrays and solar ponds. You might research some of the world's largest!

PHYSICS FACT — WORLD'S LARGEST SOLAR CELL

The Second World Conference on Photovoltaic Solar Energy Conversion, held in Vienna, July 1998, announced the production of the world's largest thin-film crystalline silicon solar cell.

It is manufactured by the Australian firm 'Pacific Solar' in conjunction with UNSW's Photovoltaics Special Research Centre.

Each solar module is 30×40 cm and has a planned efficiency of 15%, with an active cell thickness of 10 µm.

Pacific Solar hopes soon to be able to produce 1.0 m^2 solar cells.

32.4 PHOTOVOLTAICS

Probably the most efficient use of solar radiation is to convert it directly to electricity. Semiconductor materials can provide the direct conversion of the Sun's radiation to electrical energy under suitable conditions. **Photovoltaic cells** made from thin slices of silicon, gallium arsenide or other semiconductors were first developed in the 1950s for use in satellites as clean, lightweight, safe and reliable sources of electrical power. In 1954 the first 'solar battery', as it was called, was invented by D. M. Chapin, C. S. Fuller and G. L. Pearson, working as a team at Bell Laboratories in New Jersey, USA, as an extension of their work on transistors. Today, **solar cells**, as they are more commonly called, are manufactured by numerous companies and form the basis of many remote area electrical power installations.

Recall from Chapter 29 that the photoelectric effect required a vacuum tube and an external source of EMF in the photoemission circuit to allow the flow of a photocurrent. The advantage of the **photovoltaic effect** is that no vacuum environment is needed and the photovoltaic cell generates its own EMF. Early photographic light meters, such as the Weston cell, consisted of a selenium or cadmium sulfide layer deposited on metallic iron. Incident light photons passed through the selenium layer, promoting electrons from the iron metallic base into the selenium conductor and generating an EMF across the junction. The iron formed the positive electrode and the selenium formed the negative electrode of the photovoltaic cell. A sensitive galvanometer was used to display the electron current generated in the circuit and this was directly proportional to the intensity of the incident light. The galvanometer movement could be readily calibrated in terms of exposure values directly.

The silicon solar cell is the most widely used photovoltaic device today. Solar cells can be made in two ways. In one method, amorphous (solid) silicon is layered directly onto glass, with the rear being protected with a clear acrylic laminate. In the other method, monocrystalline silicon wafers are placed behind glass plates to protect them against physical shock as this form is particularly brittle. The amorphous type is more expensive but is more robust. The solar arrays used in the 1990 World Solar Challenge race by its eventual winner, the *Spirit of Biel Bienne* from Switzerland, were made of monocrystalline silicon cells developed by the Centre for Photovoltaic Devices and Systems at the University of NSW, Sydney, directed by Professor Martin Green. These 'green cells', made for the Spirit car, used laser grooved solar cells to maximise efficiency, which in the race peaked at 17%. In typical solar racing conditions the array averaged an output of 980 W, enabling the winning time from Darwin to Adelaide (3007 km) to be recorded at 46 hours 8 minutes.

On 22 August 1995, the Sandia National Laboratories in New Mexico confirmed that a 'thin' solar cell made at the University of New South Wales from crystalline silicon had achieved an efficiency of 21.5%. This was an improvement on the 1994 record of 17% set by

PHYSICS FACT

The University of New South Wales Photovoltaic Research Centre's recent work has led to development of the multilayer solar cell. This innovative cell structure has the potential to overcome the efficiency limits of amorphous solar cells while maintaining low processing costs. Such a development could make solar cells far more cost-competitive with conventional generators such as coal-fired power stations.

A conventional solar cell is 350 µm thick and consists of just one layer of P-type silicon and one layer of N-type silicon sandwiched together to form a PN junction. (A human hair is about 50 µm wide.) A multilayer solar cell only 15 µm in width, however, consists of up to 10 very thin alternate layers of P-type and N-type silicon deposited onto glass to give several PN junctions. It is this multilayer structure that enables moderate efficiencies to be achieved with low-quality material, since each light-generated charge carrier does not have very far to travel to reach a PN junction.

researchers at the Australian National University. To qualify as a thin device, the cell had to be less than 50 µm thick — about the thickness of a human hair. The record-breaking solar cell tested in New Mexico was about 47 µm thick. The UNSW team also holds the current record for the most efficient conventional solar cell at 24%. At over 400 µm thick, however, these cells require nearly 10 times as much silicon as the 25% 'thin' type.

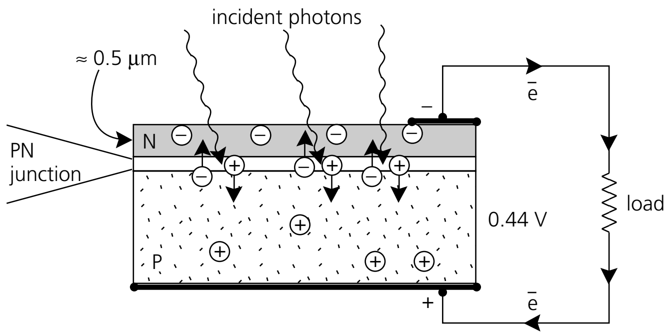


Figure 32.7
Action of a silicon solar cell.

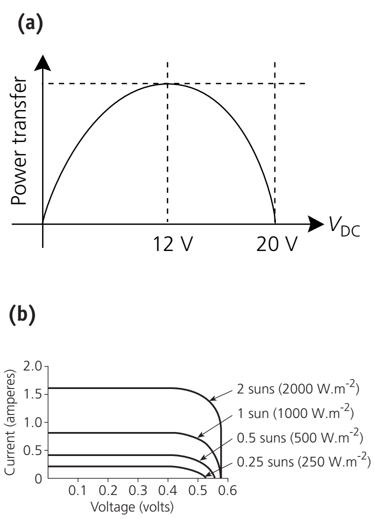
Refer to Figure 32.7, illustrating the typical photovoltaic action in a silicon solar cell. It represents a PN junction with an external circuit connected across it. Silicon atoms require incident photon wavelengths in the near infrared region of the spectrum ($\lambda = 10^{-6}$ m) to dislodge electrons in the crystal lattice. When light is incident on the exposed thin N-type surface, most pass through into the PN junction layer. The photon energy is transferred to the electrons that are dislodged from the atoms, producing electron-hole pairs in the junction region. The electric field in this region forces uncombined electrons into the N-type layer and equivalent holes are left in the P-type layer. This generates the EMF source of 0.44 V, which produces an electron flow in the external circuit. It should be noted that the layers of N-type and P-type semiconductor in the diagram could be reversed and the cell would operate just as efficiently.

A typical single-crystal silicon PV cell of 100 cm^2 will produce about 1.5 W of power at 0.5 V DC and 3 A under full summer sunlight (1000 W m^{-2}). The power output of the cell is almost directly proportional to the intensity of the sunlight. (For example, if the intensity of the sunlight is halved, the power will also be halved.) Figure 32.8**(b)** shows the current and voltage output of a solar cell at different light intensities.

An important feature of PV cells is that the voltage of the cell does not depend on its size, and remains fairly constant with changing light intensity. However, the current in a PV device is almost directly proportional to light intensity and size. When people want to compare different-sized cells, they record the current density, or amps per square centimetre of cell area.

Solar cells of this type can be constructed in panels that can be connected together in series to increase available voltage, as well as in parallel to increase the current capability. Practical solar panels are manufactured with useful voltage and power ratings. The power rating of a solar panel is directly related to its physical size. The rated voltage of a solar panel is not its open circuit voltage, however. For example, a 12 V panel usually has an open circuit voltage of between 18 V and 23 V, but delivers its output power most efficiently at 12 V (Figure 32.8**(a)**). If you search various electronics stores you'll find that most solar panels are rated at either 6 V or 12 V. The company BP Solar Australia, which sells solar cells to the biggest user in the world, Telstra, makes available a range of solar panels for serious electrical energy applications, such as the 32 cell 12 V 5.5 W unit through to the 36 cell 12 V 60 W unit covering 0.6 m^2. These panels, coupled with DC regulators, make ideal auxiliary charging systems for batteries in domestic, automotive and boating applications.

Figure 32.8
(a) Solar cell power curve.
(b) Current and voltage output of a single solar cell under varying light levels.



Questions

8 Explain the difference between a photothermal device and a photovoltaic device. Give an example of each.

9 Consider Figure 32.3. Explain any advantages or disadvantages of planting trees across the rear of the house.

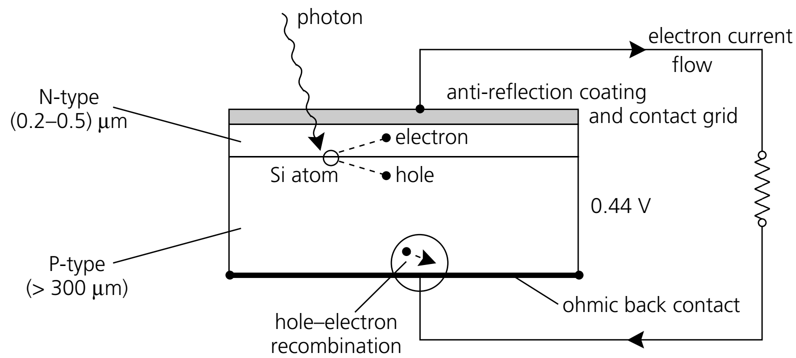


Figure 32.9
For question 10.

10 Figure 32.9 represents a cross-sectional diagram of an actual NP silicon solar cell. Explain each of the diagram labels shown in order to illustrate the principles of operation.

11 Photovoltaic action is a type of photoelectric effect. How do the two actions differ?

12 You are designing a solar car with a total roof area for solar cells of 6.4 m^{-2}. Calculate the electrical power available, assuming total cell efficiency of 17% and a constant solar flux of 980 W m^{-2}. Explain how this calculated power would vary in the actual operation of your car. What advantage would using the latest 'thin' cells provide?

32.5 ULTRAVIOLET RADIATION

UV spectrum

Australia has one of the world's highest potentials for the use of solar energy, but in recent years the abundance of this energy has led to concern over its medical problems as well. **Ultraviolet radiation** (UV) is electromagnetic energy that has wavelengths stretching from 400 nm, the wavelength of violet visible light, through to 1.0 nm, the wavelength of long X-rays. UV radiation can be harmful to living things, especially at wavelengths shorter than 300 nm. Ultraviolet sterilisation of surfaces uses wavelengths less than 310 nm because it kills bacteria and viruses. In humans, unprotected exposure to UV radiation can cause sunburn and eventually skin cancer, but it is not entirely harmful, as a large proportion of the necessary vitamin D that we need for good health is produced by skin irradiated with ultraviolet rays. Tanning of the skin is produced by gradual UV exposure. Delayed tanning is a result of melanin production in the skin as a response to UV radiation exposure. Photodamage, including premature wrinkling and aging of the skin, is a result of chronic exposure to UV radiation. No amount of tanning, however, will decrease the possibility of skin cancers developing, which is a common misconception held by the so-called 'bronzed Aussie' brigade.

Ultraviolet radiation is also used to erase programmable (EPROM) chips in computers as well as cause certain dyes and inks to fluoresce in applications such as 'black light' signature readers in banks and special effects lights at the local disco. Ultraviolet radiation is produced artificially mainly by vapour discharge tubes or electric arc lamps, whereas the major source of natural UV radiation is, of course, the Sun.

In terms of the solar spectrum, ultraviolet radiation is classified as:

- soft or **UV-A** radiation with wavelengths between 400 nm and 315 nm
- hard or **UV-B** radiation with wavelengths between 315 nm and 290 nm
- **UV-C** radiation with wavelengths less than 290 nm.

Although UV-B radiation is the most dangerous ionising form, as it can strip electrons from atoms in its path, it is generally considered today that both UV-A and UV-B can contribute to sunburn and more harmful skin cancer development. Hence, the development of 'broad-spectrum' sunscreens that will effectively block both bands for a given period of time under the right conditions. Ordinary window glass is opaque to a large portion of the UV spectrum, particularly short wavelengths. Special UV glass is transparent to the longer wavelengths.

The Earth's atmosphere protects living organisms from the majority of solar UV radiation. The **ozone layer** of the atmosphere absorbs most of the incident wavelengths, especially the shorter band. Ozone is a colourless gas present in the upper atmosphere. Although scientists are concerned about the ozone hole in the Earth's atmosphere, the gas itself is far less welcome at ground level. Ozone arises artificially from the interaction of vehicle exhaust gases, such as nitrogen oxides, hydrocarbons and carbon monoxide, with sunlight. Ozone can be detected by humans at about eight parts per billion (8 ppb), and it smells like weak chlorine. At 50 ppb it causes headaches, while above concentrations of 120 ppb, eye and mucous membrane irritation develops. Ozone is a very strong photochemical oxidant and it attacks and damages many materials including rubber, cellulose, dyes and organic paint binders. The US Environmental Protection Agency states that humans should not be exposed to ozone levels greater than 120 ppb (averaged per hour). Ozone levels often exceed 200 ppb, however, in many of the world's largest cities.

Ultraviolet monitoring

The monitoring and experimental analysis of incident solar UV radiation can be carried out with simple photodiode detectors such as the Vital Technology **BW-10 monitor**, which formed the basis of an Australia-wide monitoring network set up by the University of Canberra. Table 32.2 lists the instrument specifications.

Table 32.2 SPECIFICATIONS OF THE BW-10 MONITOR

Size	2.35″ × 3.5″ × 0.9″ outside dimensions
Weight	195 g
Battery life	8000 hours typical (9 V alkaline)
Frequency response	Weighted 290 nm to 365 nm (CIE human skin response)
Dynamic range	0.1 to 9.9 AES skin damaging UV units (CIE weighted)
LCD display type	2 digit transreflective — twisted nematic
A–D converter	Auto zero, temperature coefficient (1 ppm/°C)
Detector–diffuser	Extended range photodiode, Teflon diffuser

This simple device is calibrated in atmospheric environmental skin damaging units on a range of 0.0 to 9.9 with the highest corresponding to an incident UV radiation of 250 mW m^{-2}. The scale units can be converted using the figures in Table 32.3 after readings have been taken. Data collected on a permanent basis can be collated and examined for long-term trends over a wide area of Australia. A typical set of monitoring results is shown in Table 32.4, obtained by high school students in Wynnum, Queensland (latitude 27° 30′ S, longitude 153° E, elevation 20 m). Notice that combined UV, indirect UV, cloud cover and other atmospheric data are valuable in this type of research. Not only can this type of instrument be used for general UV monitoring but it can also form the basis of experimental design into such research topics as sunscreen testing, surface reflectivity, efficiency of sunglasses and effects of UV radiation on plant growth. Refer to the suggested activity and check with your teacher whether your school has a UV monitor.

Table 32.3 BW-10 CONVERSIONS

In the table below: Type I skin = Fair. Type VI = Olive-Dark. E.g. if the meter reads 5.0, a safe exposure time in the sun is 21 minutes for Fair, and 75 minutes for Dark skin.

AES #	MINUTES BY SKIN TYPE I	II	III	IV	V	VI	mW/m^2
0.2	520	700	960	1167	1427	1880	5
0.4	260	350	480	583	713	940	10
0.6	173	233	320	389	476	627	15
0.8	130	175	240	292	357	470	20
1.0	104	140	192	233	285	376	25
1.2	87	117	160	194	238	313	30
1.4	74	100	137	167	204	269	35
1.6	65	88	120	146	178	235	40
1.8	58	78	107	130	159	209	45
2.0	52	70	96	117	143	188	50
2.2	47	64	87	106	130	171	55
2.4	43	58	80	97	119	157	60
2.6	40	54	74	90	110	145	65
2.8	37	50	69	83	102	134	70
3.0	35	47	64	78	95	125	75
3.2	33	44	60	73	89	118	80
3.4	31	41	56	69	84	111	85
3.6	29	39	53	65	79	104	90
3.8	27	37	51	61	75	99	95
4.0	26	35	48	58	71	94	100
4.2	25	33	46	56	68	90	105
4.4	24	32	44	53	65	85	110
4.6	23	30	42	51	62	82	115
4.8	22	29	40	49	59	78	120
5.0	21	28	38	47	57	75	125
5.2	20	27	37	45	55	72	130
5.4	19	26	36	43	53	70	135
5.6	19	25	34	42	51	67	140
5.8	18	24	33	40	49	65	145
6.0	17	23	32	39	48	63	150
6.2	17	23	31	38	46	61	155
6.4	16	22	30	36	45	59	160
6.6	16	21	29	35	43	57	165
6.8	15	21	28	34	42	55	170
7.0	15	20	27	33	41	54	175
7.2	14	19	27	32	40	52	180
7.4	14	19	26	32	39	51	185
7.6	14	18	25	31	38	49	190
7.8	13	18	25	30	37	48	195
8.0	13	18	24	29	36	47	200
8.2	13	17	23	28	35	46	205
8.4	12	17	23	28	34	45	210
8.6	12	16	22	27	33	44	215
8.8	12	16	22	27	32	43	220
9.0	12	16	21	26	32	42	225
9.2	11	15	21	25	31	41	230
9.4	11	15	20	25	30	40	235
9.6	11	15	20	24	30	39	240
9.8	11	14	20	24	29	38	245
9.9	11	14	19	24	29	38	248

Table 32.4 UV MONITORING DATA

MARCH 1993	MON.	TUES.	WED.	THURS	FRI.	SAT.	SUN.
Time of day	noon	noon	noon	noon	noon	noon	noon
Combined UV (meter display)	5.9	5.4	5.7	6.3	5.5	5.6	4.8
Combined UV (mW m^{-2})	147.5	135	142.5	157.5	142.5	140	120
Indirect UV (meter-shade)	1.9	1.7	2.2	0.9	2.2	1.5	1.4
Indirect UV (mW m^{-2})	47.5	42.5	55	22.5	55	37.5	35
Cloud type (cumulus = 1 stratocumulus = 2)	1	nil	2	2	nil	1	1
Cloud cover northern (%)	25	0	20	30	0	50	25
Cloud cover southern (%)	75	0	20	10	0	50	80
Temperature °C	28	32	31	28	25.5	28	29
Relative humidity (%)	50	45	50	62	62	52	46
Pressure (hPa)	1026	1015	1013	1013	1012	1019	1022
Wind speed (km h^{-1})	4	8	9	6	<4	<4	<4
Wind direction	SW	W	SW	ENE			

Activity 32.1 USING THE UV MONITOR

Design an experiment, making use of a typical UV monitor, to gather conclusive data in order to answer any of the following questions:

1. Does the direct UV level (DUV) change more during the day in summer or winter?
2. How does the type of reflective surface affect the amount of UV radiation received by a person, for example? What is the difference between grass, concrete, asphalt or sand?
3. Do fluorescent lights pose a greater UV threat than incandescent lighting?
4. How does the screening effectiveness of materials such as shade cloth, polycarbonate sheeting, fibreglass sheeting compare as roofing cover on pergolas?
5. Is an SPF 15+ sunscreen more effective than an SPF 5+ sunscreen?
6. Does the degree of darkening of photochromic sunglasses have an effect on their ability to screen UV radiation?

Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***13** Explain the following terms related to solar physics: nuclear fusion; solar flux; corona; global warming; active solar design; Trombe wall; thermosyphon; flat plate collector; solar cell; soft UV rays; ozone.

***14** Explain why reflective foil insulation on the ceiling is better at reducing energy gain in summer than reducing energy loss in winter.

****15** Electric cars can be run off solar energy converted directly to electricity to run the electric motors and to be stored in batteries for later use.
- **(a)** If the solar cells are only about 15% efficient, where does all the remaining energy go?
- **(b)** Why do solar racing cars require very efficient battery systems?
- **(c)** What does it mean to say that the electric motor has an efficiency of 95%?
- **(d)** What maximum current is drawn from a solar array providing 65 V if the electric motor is rated at 1.2 kW DC?

***16** List the advantages and disadvantages of solar ultraviolet radiation for living organisms.

****17** Research five important skin cancer facts that all Australians should be aware of.

****18** Use Figure 32.8 to compare the open circuit voltage of a silicon solar cell array with its rated output voltage and power.

****19** Using the data of Tables 32.3 and 32.4, try to answer the following:
- **(a)** Check the conversions from meter display to readings in mW m^{-2}. Are they all correct?
- **(b)** List all meteorological features on the day of highest direct UV reading.
- **(c)** List all meteorological features on the day of lowest direct UV reading.
- **(d)** Does the incident UV reading appear to be influenced by cloud cover?
- **(e)** What are the limitations of this table of monitored data?

Extension — complex, challenging and novel

*****20** The CSIRO low energy consumption home (LECH) has a total solar air heater area of 19 m^2. (Refer to Figure 32.4.) If solar energy falls on this house at an average rate of 12 MJ m^{-2} per day. Calculate:
- **(a)** the total energy received per day;
- **(b)** the energy transferred to circulating air, assuming transfer efficiency of 65%;
- **(c)** the energy stored in the rock-bed during the day, assuming air transfer efficiency of 90%.

*****21** A solar hot water system receives solar energy at the rate of 10.5 MJ m^{-2} per day. If the collector area is 4.8 m^2, collector efficiency is 0.7 and the water volume is 325 L,
- **(a)** calculate the total water energy gain per day;
- **(b)** estimate the temperature rise of the water in the tank during the day.

*****22** Calculate the extra tilt support bracket length, l, that would be needed on a north-facing roof with a 15° pitch for a solar hot water system collector plate. Assume the house location is in Rockhampton, Qld (latitude 23°S) and the collector plate is 1.8 m long, with the tilt bracket at right angles to the roof.

*****23** Using Table 32.3, plot a curve of BW-10 AES reading versus exposure in minutes for skin type III. Can you deduce a formula linking these variables?

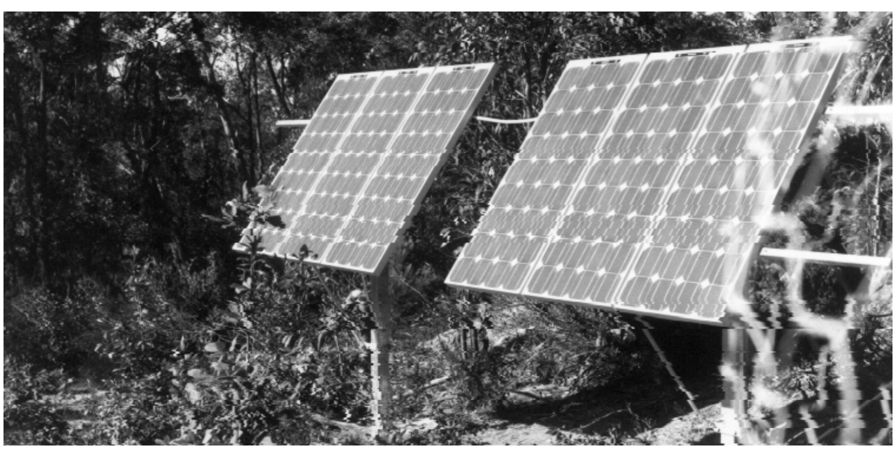


Photo 32.2
These solar modules are used to power a small house.

CHAPTER 33

Medical Physics

INTRODUCTION 33.1

Discoveries in physics have played a major role in the development of medicine, especially those branches of medicine that are concerned with the use of radiant energy and nuclear isotopes in the diagnosis and treatment of disease. Improvements in microscopy techniques, such as the higher resolution of electron based microscopes, have resulted directly from the practical applications of theoretical discoveries in quantum physics.

Ever since the Nobel prize-winning discovery of the X-ray by Bavarian physicist Wilhelm Conrad Roentgen in 1895, when he produced the first X-ray of his wife's hand, diagnostic radiology or medical imaging has improved in its ability to photograph and record the internal anatomy and physiology of the human body and those of other animals. Diagnostic radiology is the imaging and analysis of both the normal anatomy and physiology of the body as well as possible abnormal effects due to disease or injury. It is usually carried out using X-ray radiographs, tomographs or computerised axial tomographs (CT scans), but other diagnostic techniques such as ultrasonics, magnetic resonance imaging (MRI), or positron emission tomography (PET) are becoming widespread.

The radiologist uses direct observation of the image obtained or extra detail can be sought with the use of various contrast media that are administered to the patient just before the radiology. Examples include upper gastrointestinal examinations (GI series), intravenous pyelograms (IVP) for the kidney and bladder, barium enemas for colon examinations, arthrograms for skeletal joints and myelograms or angiograms for the spinal cord and blood lymph vessels. These procedures allow the radiologist to record movements of organ systems internally as the contrast material flows through them in real time. The image is viewed directly on a radiation-sensitive screen (fluoroscopy), computer monitor or by recording onto videotape.

Australia possesses many nuclear medicine departments in hospitals and private facilities that use a range of medical radioactive isotopes. These isotopes are produced mainly by neutron bombardment in reactors such as at the Australian Nuclear Science and Technology Organisation (ANSTO) at Lucas Heights in Sydney. The cyclotron particle accelerator device is also used to produce short-lived isotopes, at places such as the National Medical Cyclotron (NMC) at ANSTO and the Cyclotron and PET Centre at Melbourne University's Austin Hospital and School of Physics. These facilities provide isotopes for therapeutic radiology, which is the treatment of malignant disease with ionising radiation in conjunction with drug therapy, hyperthermia and psychological counselling.

Medical physicists are those specialists who work with radiologists, oncologists, physiologists and radiographers in providing numerous practical applications of physics in the medical sciences. The understanding of basic physical principles is a necessary prerequisite for all these fields of study. In this chapter let us briefly examine the underlying principles of these diagnostic and therapeutic tools. You may need to revise previous chapters on optics, electromagnetism, quantum physics and nuclear physics.

33.2 MICROSCOPY TECHNIQUES

A microscope's resolution, or ability to distinguish small detail in a specimen, is limited by the wavelength of the light used to illuminate the specimen. In an electron microscope (EM) a beam of electrons accelerated by a high voltage (50 kV) is used instead of visible light. The de Broglie wavelength of these electrons is about 100 000 times shorter than that of light photons and so an electron microscope greatly increases the possible resolving power. Modern **transmission electron microscopes** (TEM) can resolve details down to about 0.2 nm, compared with the best optical microscopes, which resolve down to about 200 nm, with magnifications up to ten million times. The limitation for the TEM is the ability of the electrostatic and magnetic lenses to maintain good focusing. Electron microscopes need the electron beam travelling through a vacuum in order to prevent scattering by air molecules. (See Figure 33.1.) The first types were built in the 1930s.

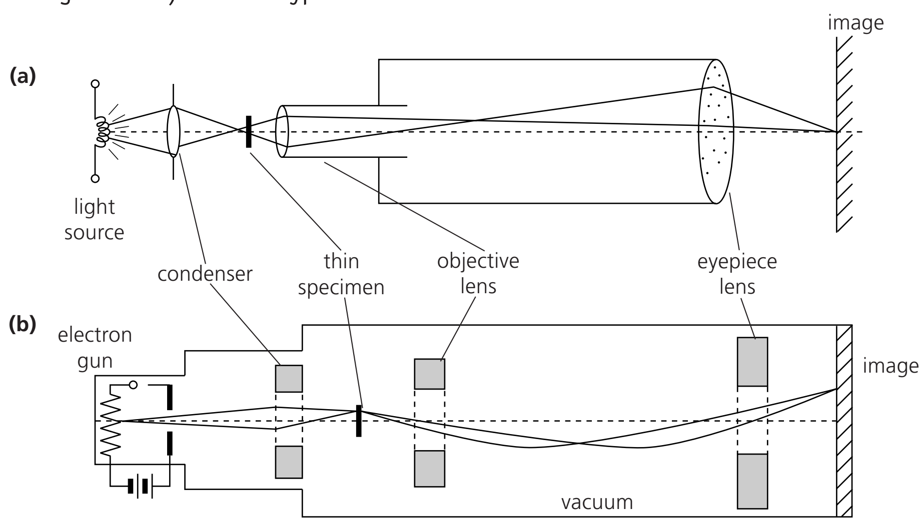


Figure 33.1
Comparison of a light microscope **(a)** with an electron microscope **(b)**.

A newer instrument is called a **scanning electron microscope** (SEM) and uses a well focused beam of electrons to scan the surface of a specimen. The first practical SEM was built in 1970 by the British-born American physicist Albert Victor Crewe. The instrument is capable of producing three-dimensional images and is not really a microscope at all. The spot beam is scanned backward and forward across a specimen by the scanning magnetic field. The incident electrons cause the ejection of secondary electrons with energies typically of a few electron-volts, which are collected to form a cathode ray tube (CRT) control grid current, as shown in Figure 33.2. The sweep or timebase of the CRT is in synchronisation with

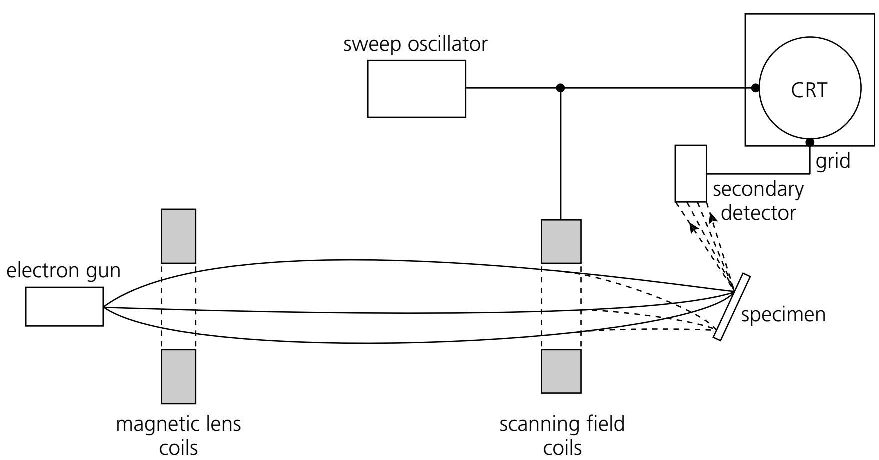
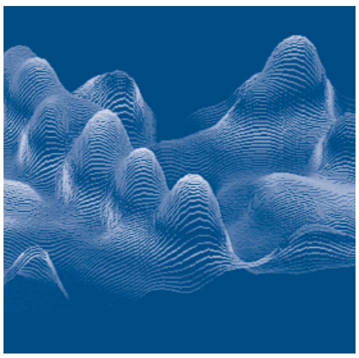


Figure 33.2
Scanning electron microscope.

the scanning beam, and the variations in electron collector current control the CRT sweep beam brightness. The specimen is usually placed at between 30° and 60° to the incident electron beam to improve secondary ejected electron current. The resolution of an SEM is less than for a TEM, with useful magnifications extending to about 50 000 times. With an SEM, thin slices of specimen are not needed as with a TEM, with even whole living specimens able to be observed. In certain circumstances, X-rays produced due to electron collisions can be used to obtain an elemental analysis of the specimen as well. The instrument is then referred to as an **electron probe microanalyser**.

Photo 33.1
A scanning tunnelling micrograph.



In 1986 the Nobel prize for physics was shared between Ernst Ruska, for his design of the first TEM in the 1930s, and Gerd Bennig and Heinrich Rohrer of the IBM research laboratory in Zurich, Switzerland, for designing a new kind of SEM called the **scanning tunnelling microscope** (STM). The device relies on the quantum tunnelling effect between a scanning metal tip probe and the surface of the specimen. As the probe is moved over the surface, the flow of tunnelling electrons is kept constant by varying the height of the probe above the specimen's surface. These fluctuations in height are used to produce topographical line scans from which 3D images can be constructed. Superconducting magnetic levitation principles are used to control the height of the probe. Bennig and Rohrer were able to obtain 10 angstrom separations with their first designs. Photo 33.1 is an STM image of the oblique surface of a crystal of tantalum diselenide obtained by Professor Dan Haneman at the University of New South Wales, showing the outer electron charge contours of the lattice array of atoms.

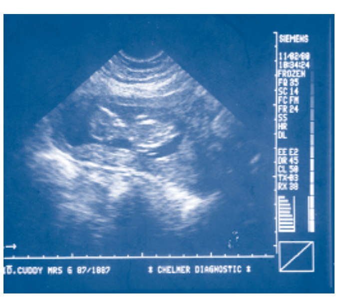
The most recent variation of the scanning probe microscope is called the **atomic force microscope** (AFM). The atomic force microscope does not use a tunnelling current, so the sample does not need to be able to conduct electricity. As the probe in an AFM moves over the surface of a sample, the electrons in the metal probe are repelled by the electron clouds of the atoms in the sample. As the probe moves along over the sample surface, the AFM adjusts the height of the probe to keep the force on the probe constant. An electronic sensing mechanism records the up-and-down movements of the probe, and feeds the data into a computer, which then constructs a three-dimensional image of the surface of the sample.

ULTRASOUND 33.3

Sound waves above the human audible frequency range, usually 20 kHz, are called **ultrasonic waves**. Modern ultrasonic generators can produce frequencies up to several gigahertz by transforming alternating voltages or currents into mechanical oscillations, through the use of **piezoelectric** crystals.

Ultrasonic waves have long been used by living organisms, such as bats and dolphins, for echo location, and similar sonar devices are used for underwater detection and communication by submariners and boaties. In physics and engineering, ultrasonics can be used in determining properties of matter, such as compressibility and elasticity, or for fault detection in industrial materials, such as sheet metal or cast components. High vibration rates caused by ultrasonic blasting is used to clean jewellery, produce photographic emulsions and even to homogenise milk. Ultrasonics in the gigahertz range can be used to produce an ultrasonic microscope able to resolve detail to about one micron.

Photo 33.2
Foetal ultrasound image.



In the medical field, ultrasound is used as a therapeutic tool to repair damaged tissue or to treat conditions such as bursitis, arthritis or muscular damage. These applications require the ultrasound probe to produce localised heating or diathermy as a result of tissue resistance to the transmission of the waves.

Ultrasound has been used to great advantage in destroying embedded kidney stones, reducing them to small fragments that can be easily removed by catheter or passed in the urine. As a diagnostic tool, ultrasound is often more revealing than X-rays in showing the subtle density differences in cancerous tissues. It is nowadays used widely to produce foetal images from the uterus. (See Photo 33.2.) Foetal ultrasound examination was first used by Dr Ian Donald of Glasgow, Scotland, in the early 1950s. The piezoelectric crystal is housed in a hand-held transducer unit pressed against the skin, using a surface gel, over the organ or part of the body being imaged. A narrow fan-shaped beam of 5 MHz ultrasound waves penetrates the surface and is partially scattered and absorbed. Reflected waves received by the transducer unit are again converted to electrical signals and sent to a computer for conversion to a two-dimensional video image in real time. Pure fluids in the body reflect very little sound, so a fluid image is black on the ultrasound scan. The ability of tissues to reflect sound waves to various degrees is called the tissue **echogenicity**. Tissues such as fatty masses and liver tissue image as white or light grey because of their high reflectivity, whereas tissues such as breast lymphoma image as dark grey because of their low reflectivity. Using high intensity, very fine ultrasonic beams, a surgeon can produce an ultrasonic scalpel for very delicate surgery in areas such as the brain or internal structure of the ear.

Doppler ultrasonography uses the Doppler effect of wavelength changes between the incident and back scattered waves to provide images of moving fluids within the body, such as blood flow (Figure 33.3). Blood cells travelling toward the transducer will cause reflected ultrasound waves of shorter wavelength than the incident waves. Again, a computer is necessary to convert the reflected wave energy into a comprehensible video image.

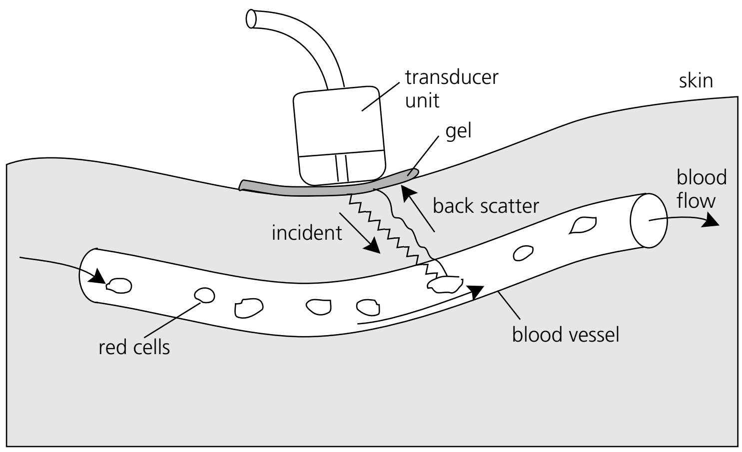


Figure 33.3
Doppler ultrasonography.

Colour Doppler allows the imaging to quickly indicate direction of blood flow. Blood flowing toward the transducer is coloured red, and blood flowing away is coloured blue. The colours are superimposed on the cross-sectional image, which gives the direction of blood flow. This technique is very useful in echocardiography studies and in identifying small blood vessels such as calf veins and kidney arteries.

Radiologists today have a wide variety of ultrasound probes. Those used for imaging body cavities and organs are called intercavitary scanners, such as are used for transvaginal scanning in the early stages of pregnancy up to about 12 weeks, and transrectal probes used for prostate gland examination. High-frequency and ultra-high-frequency (20 MHz) probes are now being developed for musculoskeletal applications and in the treatment of various skin disorders. The greatest advantages of ultrasound in medicine are its lack of ionising radiation, relatively low cost and ease of portability. Despite the possible destructive effects of ultrasound, medical imaging is now regarded as quite safe.

33.4 MEDICAL ISOTOPES AND RADIATION

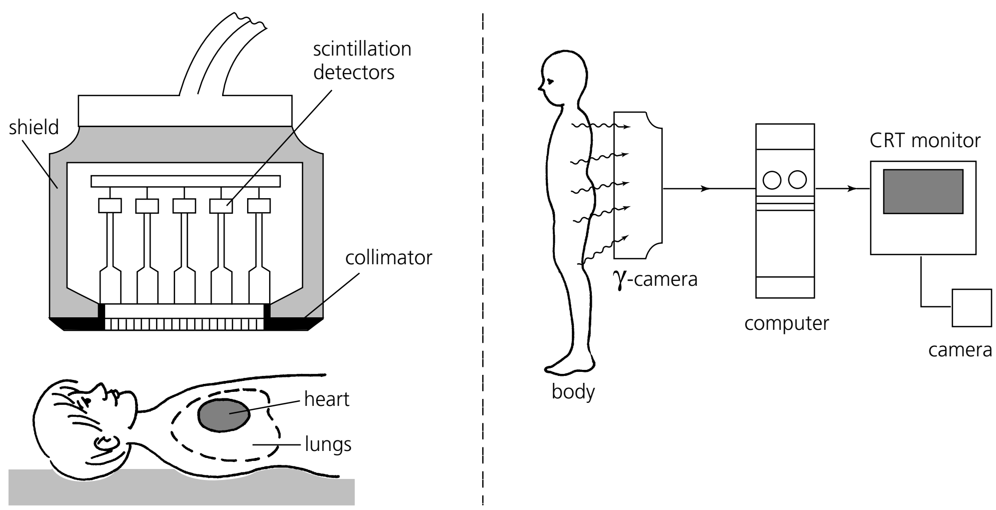
The history of nuclear medicine is closely interwoven with major discoveries in physics and chemistry. Radioisotope production techniques developed rapidly during the 1970s and 1980s. Today, the use of these isotopes has been combined with various imaging techniques and computer data analysis to become a powerful medical diagnostic tool.

Scintigraphy refers to the use of gamma (γ) radiation to form images following injection of a suitable **radiopharmaceutical** compound. The radio part refers to a radionuclide, which is the radiation emitter, such as the widely used technetium (^{99m}Tc), while the pharmaceutical part refers to the compound to which the radiation emitter is bound or attached, and which is injected into the body to be observed and analysed. The radioisotope technetium-99m is an isotope of the artificially produced element technetium and it has almost ideal features for nuclear medicine studies:

- It has a 6 hour half-life, which is long enough to adequately examine metabolic processes yet is short enough to minimise radiation dose to the patient.
- It decays by gamma rays and low energy electrons only.
- The low energy gamma rays escape the human body easily and are accurately detected by an external gamma camera.
- The chemistry of technetium is very versatile and it can be tagged onto a range of biomolecules that concentrate in different organ groups of the body.

Once the radiopharmaceutical compound is absorbed by organs or regions of the body, the gamma rays are imaged using an external gamma camera, which converts the absorbed energy of the radiation into an electrical signal for recording via a process called **scintillation**. When a gamma photon strikes a crystal of sodium iodide that has been doped with a small amount of thallium, the energy is absorbed and re-liberated by the crystal as a photon of visible light. This light is detected by a photovoltaic cell and converted into an electrical impulse that can be amplified and recorded. The gamma camera used in medical scintigraphy measures the radiation emitted by each spot in the body through the use of a multi-channel collimator. The camera contains numerous scintillation detectors corresponding to collimated channels. The outputs of the detectors are computer-combined into a single colour-enhanced image on a monitor screen. The gamma camera itself is housed inside a lead shield to protect the sensitive detectors from background radiation. A typical nuclear medicine camera is the General Electric 'Starcam 3000', or the General Electric 400 ACT or the Marconi Irix. (See Figure 33.4.)

Figure 33.4
Scintigraphy using a gamma camera.



The metastable atom ^{99m}Tc, in passing from the high energy state to the low energy state, releases a gamma photon with energy 140 keV. This makes it very suitable for use in imaging. Technetium-99 has a half-life of about 6 hours and is very versatile. If injected into the bloodstream together with a tin compound, for instance, it attaches to the red blood cells and can be used as a blood flow **tracer**. If administered as the compound ^{99m}Tc-methylene phosphorate, it is taken up by bone and can be used to detect early osteomyelitis much faster than the wait needed for early calcium deposition to be shown on X-ray images. As well as the technetium, other widely used diagnostic medical radioisotopes are gallium-67 (^{67}Ga), thallium-201 (^{201}Tl), iodine-123 (^{123}I) and indium-111 (^{111}In). Some therapeutic radioisotopes used in nuclear medicine include iodine-131 (^{131}I), phosphorus-32 (^{32}P) and strontium-89 (^{89}Sr). (See Table 33.1.)

Table 33.1 MEDICAL RADIOISOTOPES

ISOTOPE	HALF-LIFE	PRODUCTION	EMISSION (keV)	CLINICAL APPLICATIONS
^{99m}Tc	6 h	reactor	γ (140)	cerebral blood flow coronary artery disease oncology and renal function
^{67}Ga	78 h	cyclotron	γ (185)	lymph node cancer, infection
^{201}Tl	72 h	cyclotron	γ (168) X (83)	coronary artery disease oncology
^{18}F	2 h	cyclotron	γ (511)	glucose metabolism — neurology cardiology, oncology
^{131}I	8 d	reactor	γ (364) β (606)	hyperthyroidism and thyroid cancer
^{123}I	13 h	cyclotron	γ (159)	thyroid and kidney studies
^{111}In	67 h	cyclotron	γ (245)	protein, cells and antibody studies
^{32}p	14.3 d	reactor	β (1710)	blood disorders

Technetium has become the most widely used radionuclide for diagnostic nuclear medicine. It is formed from the decay of a parent radionuclide, molybdenum-99, which, through this parent–daughter process, can be provided in a convenient, readily available and mobile form, the technetium generator. Table 33.2 shows the available technetium-labelled compounds and their uses.

Table 33.2

RADIOPHARMACEUTICAL	SHORT FORM	CLINICAL USE
Technetium sulfur colloid	99mTcS/C	reticulo-endothelial system (liver, spleen and bone marrow scan)
Technetium macro aggregated albumin	99mTcMAA	pulmonary blood flow (lung scan)
Technetium diethylene triamino penta acetic acid	99mTcDTPA	renal blood flow, function and excretion (kidney scan)
Technetium methylene diphosphonate	99mTcMDP	skeletal studies (bone scan)
Sodium pertechnetate	$Na_2 99mTcO_4$	thyroid, salivary gland and gastric scans
99mTc red blood cells	99mTcRBC	cardiac function and blood pool scans
99mTc Sestamibi 99mTc Tetrofosmin	99mTcMIBI 99mTcTETRO	myocardial perfusion (heart muscle blood flow)
99mTc hexa methylene propylene amine oxime	99mTc HMPAO	brain scan and scans for infection

As can be seen from Table 33.1, the production of radioisotopes in Australia is both nuclear reactor and accelerator cyclotron based. Today, cyclotrons are the preferred method and a lot of research is currently being done to investigate the cyclotron production of technetium in Australia. Presently the technetium-99 and iodine-131 are produced at the Australian Nuclear Science and Technology Organisation (ANSTO) HIFAR reactor at Lucas Heights, Sydney.

Other reactor radioisotopes currently produced include:

- Cobalt-60: used for external beam radiotherapy.
- Iridium-192: supplied in wire form for use as an internal radiotherapy source.
- Iron-59: used in ferrokinetic studies of iron metabolism in the spleen.
- Selenium-75: used in the form of seleno-methionine to study the production of digestive enzymes.
- Ytterbium-169: used for cerebrospinal fluid studies in the brain.

ANSTO's National Medical Cyclotron facility (NMC) produces thallium-201 and gallium-67 for both myocardial (heart) and tumour studies at Australian hospitals. Australia's second medical cyclotron is housed at Melbourne's Austin Hospital, in the Cyclotron and PET Centre, which primarily produces positron emitting radioisotopes for positron emission tomography (PET). This is discussed in Section 33.5.

Other cyclotron produced radioisotopes include:

- Rubidium-81: as a gas source, this isotope can produce images of lung ventilation conditions such as asthma.
- Carbon-11 and nitrogen-13: used to study brain physiology and pathology, especially in conditions such as epilepsy and dementia.

In Australia, the National Health and Medical Research Council (NHMRC) maintains the standards for radiation protection. Radiation hazards occur to the body as a result of damage to cells caused by ionising radiation. This damage, as a result of the formation of chemically active ions inside the cells, can take various forms but usually involves a combination of temporary cell division inhibition, or genetic chromosome damage leading to mutations or even cell death. Those cells mostly at risk are the actively dividing ones, such as bone marrow, lymph glands or the gonads. The degree of damage varies according to radiation dose and dose rate, the irradiated volume of tissue and the type and duration of radiation.

Recall from Chapter 28 that the units for **absorbed dose** of radiation refer to the energy absorbed in a given mass of body tissue as a result of ionising radiation. The SI unit is the joules per kilogram (J kg^{-1}) and is referred to as the **gray (Gy)**, where 1 Gy = 1.0 J kg^{-1}.

Table 33.3 WEIGHTING FACTORS — DOSE EQUIVALENT

TYPE OF RADIATION	DOSE EQUIVALENT
Photons (X-rays and γ-rays)	1.0
Electrons (β particles)	1.0–2.0
Neutrons (fast or thermal)	5–20 depending on energy
Protons	5–10
Alpha particles	20
Heavy ions	20

Dose equivalent is a refined unit that takes into account the fact that some types of radiation can produce more damage in tissues than others, even though the absorbed dose is the same. This leads to the use of weighting factors for the different types of radiation, as shown in Table 33.3. The SI unit for dose equivalent is the joules per kilogram (J kg^{-1}) or the **sievert (Sv)**, where 1 Sv = 1.0 J kg^{-1} = (weighting factor) $\times$ absorbed dose.

Table 33.4 lists some typical absorbed doses administered to an adult patient for common X-ray procedures, and Table 33.5 lists the total body dose equivalent administered to an adult patient for some common nuclear medicine studies. The total average intake from natural background radioactivity is 1–2 mSv per year. The highest known level of background radiation is in the Kerala and Madras regions of India where a population of over 100 000 people receives an annual dose rate that averages 13 mSv. The dose from a normal X-ray is about 25 μSv, while the dose from a typical dental X-ray to the cheek is about 1.0 mSv. It is estimated that if 100 people are exposed to 1.0 Sv of radiation, then 5 of these people will develop a fatal cancer. A dose of 5–6 Sv over a short period of time leads to acute radiation sickness and death as a result of damage to bone marrow, the gastrointestinal system and the central nervous system.

Table 33.4 TYPICAL RADIATION DOSES ADMINISTERED TO ADULTS (X-RAY UNITS = mGy)

PROCEDURE	SKIN	BONE MARROW	OVARY
Abdomen AP	4.9	0.48	0.84
Chest AP	0.2	0.042	0.002
Pelvis AP	4.0	0.53	0.75
Kidneys IVP	5.2	0.47	0.53
Lumbar spine (lateral)	20.7	0.79	1.36

Table 33.5 TOTAL BODY DOSE EQUIVALENT USED IN SOME NUCLEAR MEDICINE PROCEDURES

NUCLEAR MEDICINE PROCEDURE	TOTAL BODY DOSE EQUIVALENT (mSv)
Thyroid scan ^{99m}Tc	1.16
Bone scan ^{99m}Tc.MDP	5.2
Lung scan ^{99m}Tc.MAA	1.8
Gallium scan ^{67}Ga	20.3

Radiation therapy

In medical terms, radiation therapy is the technique used to deliver a lethal radiation dose to a specific organ or site in the body while keeping the dose to surrounding tissues to a minimum. The most common methods involve using internal radioisotopes that target specific sites or external rotation techniques that allow concentration of radiation beams to very localised sites. Examples of isotopes used for these therapeutic purposes are cobalt-60 and caesium-137. The commonest use of therapeutic radioisotopes is in the treatment of cancerous tumours. The chances of recovery from the different forms of cancer are variable, depending on factors such as early detection. Lung and bowel cancers have the lowest survival rate, even following treatment. The methods of treating cancer involve a combination of surgery, chemotherapy (chemically based therapy), and radiation therapy. In radiation therapy the usual method is to give a total dose of about 120 mSv, split into a series of smaller doses of about 12 mSv over a span of 20 days. A newer technique involves delivering doses to specific targets, using radioactive elements that chemically bind to the DNA of the cancer cells. This method is highly localised and thus much more efficient.

In all forms of radiography and nuclear medicine, the protection of patients and staff is very important. In the techniques of diagnostic and therapeutic radiology the 'ALARA' principle is used. The probability of damage by all justifiable exposure to radiation is kept *A*s *L*ow *A*s is *R*easonably *A*chievable, which includes keeping doses to individuals, number of people exposed, and likelihood of others exposed, as low as possible. Staff are required to wear radiation monitors called thermoluminescent dosimeters (TLDs) in order to be checked monthly for exposure. Most operators return a 'below detectable limit' or BDL reading, but a high exposure might be 200 μSv for the month. These devices were discussed in Chapter 28. In Australia, radiation protection regulations are based on the International Commission on Radiological Protection (ICRP) guidelines, which provide a maximum permissible dose for occupational exposure of 20 mSv a year averaged over five years (total 100 mSv) with a maximum of 50 mSv in any one year. For public exposure the maximum dose is 1.0 mSv a year averaged over five years (total 5.0 mSv).

Hadron therapy

The term **hadron therapy** was first used in the early 1990s to describe radiation therapy using beams of heavy charged and uncharged particles such as protons, neutrons and heavy positive ions of carbon, neon and silicon. The name distinguishes this form of radiation therapy from its counterparts using X-rays (photons) and high energy electron beams (leptons) in the naming conventions of the standard model. One of the major problems with the therapy is the high cost of treatment machine facilities, which require cyclotrons and synchrotrons. However, in the small number of major centres of hadron therapy in the world, such as in the USA, Japan, Switzerland and Germany, clinical applications are showing excellent results.

The main advantage of hadron therapy over conventional radiation therapy that aims to kill tumour cells with beams of ionising radiation is that the hadron beams provide much better (higher) dose distributions to the tumour itself while limiting the doses to surrounding healthy tissue. This results from the way hadron beams are absorbed at the microscopic level within the tumour tissue. Generally speaking the neutron, proton or ion beams produce a greater number of secondary charged particle interactions along the beam path at the target site in the tissue. This means that they produce high energy depositions or high linear energy transfer (LET) characteristics in a very small area of tissue.

This is significant at the cellular level because it allows the energy to be directed at the cell's DNA double helix. In simple terms the hadron beams are much better at targeting and ionising the actual DNA, whereas both electrons and X-rays are quite poor at targeting such a small area. The biological DNA is the controlling centre of the tumour cell, so destroying the DNA kills the cells very effectively and leads to elimination of the cancer.

A second advantage with High LET radiation beam therapy is that cells irradiated by this method show far less tendency to change their cell division cycles, so far fewer cells are prone to therapy resistance as is often the case with electron or X-ray therapy.

A third advantage relates to the fact that typical cancer tumours are poorly supplied with blood vessels, and tumour tissues are therefore low in oxygen content ('hypoxic' conditions). Such tissues are resistant to conventional X-rays but can be destroyed more effectively by hadron beams. Since normal tissues surrounding the tumour are well oxygenated, irradiation with hadrons will kill more tumour cells than X-ray irradiation will.

The first use of neutrons in the treatment of cancer patients was in the USA at the University of California, Berkeley Cyclotron Laboratory in 1938. This cyclotron machine designed by E. O. Lawrence used a 16 MeV deuteron beam smashing into a thick beryllium target. The neutron beam itself had a mean energy of about 7.0 MeV. Today the most advanced neutron therapy accelerator is housed at the Wayne State University Gerhenshon Radiation Oncology Centre and uses a 48.5 MeV deuteron superconducting cyclotron coupled with multileaf collimators. The facility uses a single shielded room with the accelerator and its internal beryllium target mounted on a ring gantry that can be rotated 360 degrees around the patient; total mass is about 60 tonnes. (Refer to Figure 33.5.)

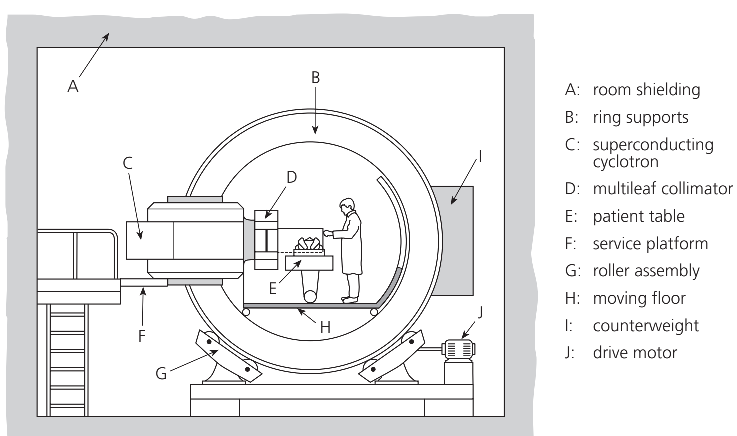


Figure 33.5
Neutron therapy superconducting cyclotron.

Proton beam therapy has been used since about 1961 and heavy ion (^{12}C) beam therapy has been used since 1975 at various centres around the world. Proton therapy tends to be the treatment of choice for paediatric cancers, as the risk of secondary radiation-induced cancer as a result of the treatment is far lower. Proton therapy may involve either 'proton radio-surgery', as used on brain lesions, or 'proton precision radiation therapy', used on the brain stem and spinal cord structures as well as prostate and cervical cancers. At present there are twenty active proton therapy centres in the world, twelve neutron and three heavy ion centres. Some of the new centres being planned will combine both proton and heavy ion machines. It is anticipated that in the next ten years hadron therapy will become a familiar tool in the armoury of oncology treatment around the world.

NEI Activity 33.1 RADIATION THERAPIES COMPARED

Use the material presented in this chapter, as well as Internet research, to compare and contrast the techniques of hadron, photon and lepton therapies in modern medicine. Outline what each contains, as well as the advantages and disadvantages of these methods. Present your report in such way as to convince the reader that more is available than just X-rays.

33.5 MEDICAL IMAGING TECHNIQUES

X-rays and tomography

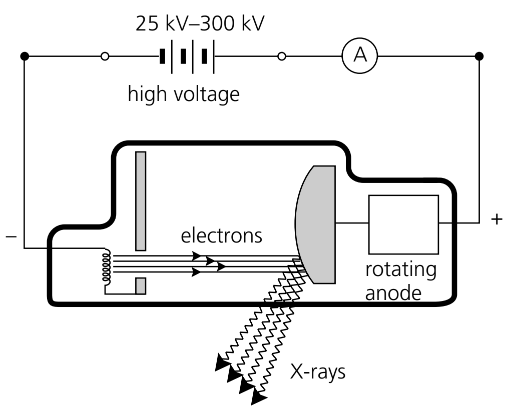


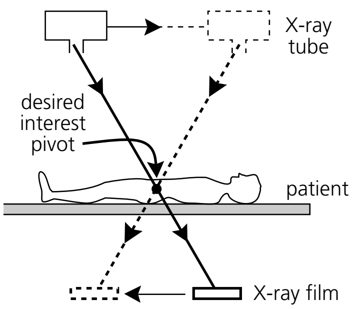
Figure 33.6
Basic X-ray tube apparatus.

X-rays are a form of EM radiation, as discussed in Chapter 28, with frequencies and energies much higher than those of visible light. X-rays are produced in an X-ray tube by focusing an electron beam onto a tungsten target. They are then able to be focused and pass through a patient's body and onto X-ray film, producing an image (Figure 33.6). The image is processed in much the same way as normal photographic film. As the X-rays pass through the body tissues they are absorbed by different amounts, resulting in a variation of densities on the final exposed and processed X-ray film. Five different densities are recognised by radiologists. Densities 1–5 in order are:
Density 1 — Air/gas: black; for example, lung, bowel, stomach
Density 2 — Fatty tissue: dark grey; for example, subcutaneous tissue layer or peritoneal fat
Density 3 — Soft tissue/water: light grey; for example, solid organs, heart, blood vessels, muscles
Density 4 — Bone tissue: off-white; for example, humerus bone
Density 5 — Contrast material/metal: bright white; for example, metal staples or pins holding a fracture.

Ch33-Walding 4th 25/8/04 1:52 PM Page 755

Organs are best seen with conventional X-ray film if they sit beside tissue of different density. For example, the right heart border is usually seen very well because it sits against air-filled lungs. Similarly, the psoas muscle is usually seen very well in abdominal X-rays because of the lower density fat tissue lying beside it. In the procedure called an intravenous pyelogram (IVP), or X-ray of the kidneys and bladder, a non-ionic iodine solution called Ioversol is injected into the patient's bloodstream to act as a contrast medium. This allows greater differentiation of the various tissues.

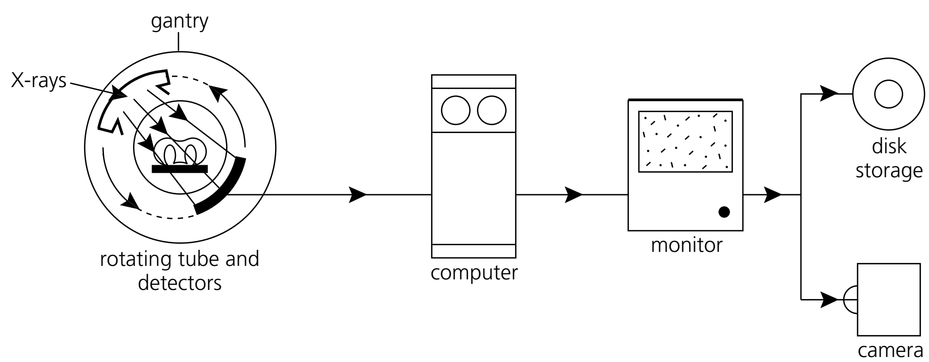
Figure 33.7
Conventional tomography.



Sectional radiography or conventional **tomography** (from the Greek *tomos* meaning 'slice') is used if the organ or structure being examined is obscured by overlying tissue as, for example, in radiography of the kidneys that are being obscured by bowel loops. In this process the X-ray tube and detecting film move about a pivot set at the desired plane of interest (Figure 33.7). Organs or structures above and below that being imaged are blurred by the motion of the X-ray tube. This technique is used today in conjunction with cross-sectional imaging techniques such as ultrasound and the newer CT or MRI scans. We will now take a look at these.

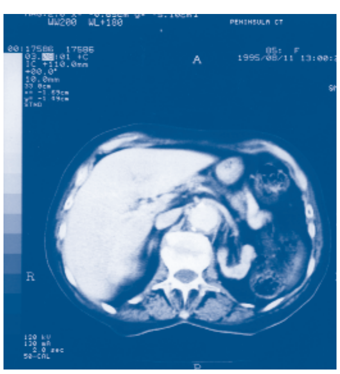
Computed tomography (CT)

Figure 33.8
Computed tomography (CT).



What used to be called the computed axial scanner (CAT) or body scanner was invented in 1972 by the British electronics engineer Godfrey Hounsfield, at the central research labs of EMI Ltd., reportedly with money made from sales of the company's Beatles records! The devices were in general use by 1979. A modern computed tomography scanner or **CT scanner**, such as the General Electric 'Pace', produces cross-sectional images with the use of X-rays. The patient passes through a gantry that rotates around the body at the level of interest (Figure 33.8). Information from the X-ray detectors is analysed by computer software and displayed as an image. These images are photographed to produce a series of slices through the body. Similar density differences are found in CT images as with conventional X-ray film but with much greater density control made available by the computer. Much greater differentiation is possible between solid organs as well as between organs and processes such as tumour or fluid collections. CT scans are also very sensitive to contrast material and minute amounts of calcium.

Photo 33.3
Typical CT scan of a body trunk.



Intravenous contrast material is used with CT scans for reasons such as differentiating normal blood vessels from abnormal masses, such as lymph nodes. Contrast material also makes tissue abnormalities more apparent. Oral contrast medium is used for abdomen CT scans to allow the distinction to be made between normal bowel loops and abnormal masses or fluid collections.

The computer software driving the CT scanners allows fine manipulation of the densities to display various tissues of the body where required. This is called altering the window settings and is especially used to view lung tissue and liver tissue in chest or abdominal CT scans. Photo 33.3 shows a typical body trunk CT scan.

Recently, CT scanners that allow continuous collection of data as a patient passes through the CT gantry have been developed. The tube and detectors rotate continuously around the body from head to toe in a spiral pattern. This is called a 'helical scan'. The software that operates the CT scanners is very complex but the major advantages of helical CT scanning are as follows:

- Increased speed of examination, a big advantage as patients undergoing CT scanning need to be kept very still.
- Rapid examination at optimal levels of intravenous contrast medium concentrations.
- Images can be retrospectively constructed from the computer data.
- High quality 3D images are possible.

Despite the tremendous complexity of CT scanners, they still involve the use of ionising radiation and cannot image many fine details in soft tissues. CT scans are also usually limited to transverse (head to toe) or axial (across the body) planes. For these reasons the technique known as magnetic resonance imaging is now becoming more widespread.

Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) has become accepted over the past ten years as a very powerful diagnostic imaging tool. The first MRI scanner was tested on 2 July 1977 by Brooklyn, NY medical researcher Ray Damadian, as a diagnostic tool that did not subject patients to X-rays. Britain introduced the technique in about 1974 and the scanner's commercial sale was approved in the USA in 1984. MRI uses the magnetic properties of the hydrogen atom to produce images, and as hydrogen is present in many biological compounds that make up the body tissues, many diagnostic applications of MRI have been developed.

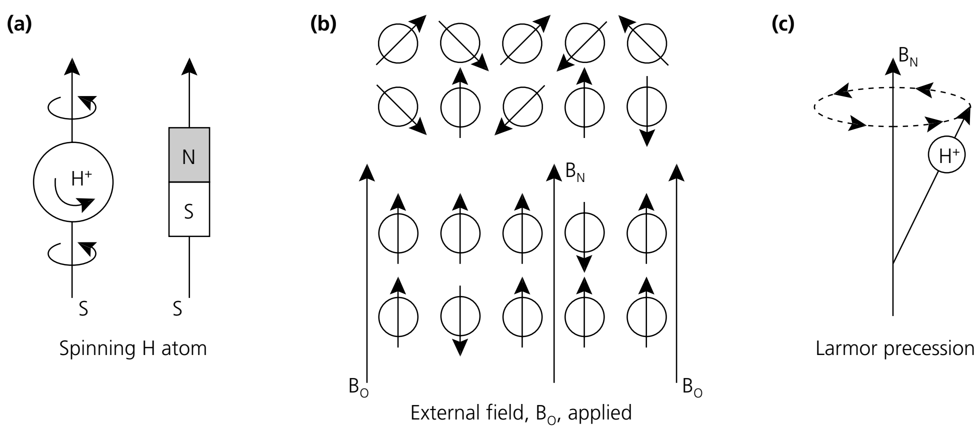
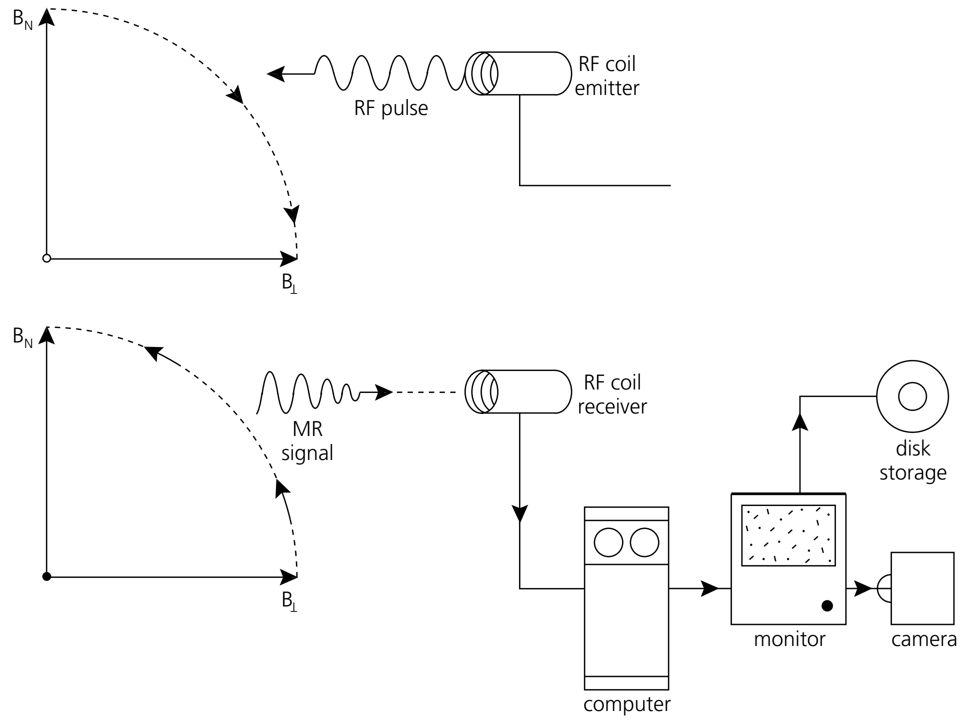


Figure 33.9 Magnetic resonance.

The single proton in the nucleus of a hydrogen atom may be thought of as a small spinning bar magnet with a north and south pole. (See Figure 33.9**(a)**.) If a very strong external magnetic field is applied to tissues containing hydrogen atoms, they will mostly align themselves in the direction of this applied field, rather than remaining randomly aligned (Figure 33.9**(b)**). Although now aligned in the direction of the applied field (B_0), the hydrogen nuclei do not remain motionless, but spin around the line of the field in a precessional motion, as shown in Figure 33.9**(c)**. The frequency of precession is an inherent property of the hydrogen atom in a magnetic field and is known as the **Larmor frequency**. The Larmor frequency changes in proportion to the magnetic field strength but is within the radio frequency range around 10 MHz. Hence the hydrogen atoms are radiating radio frequency (RF) energy.

A second magnetic field is now applied at right angles to the original external field. This second magnetic field is at the same frequency as the Larmor frequency and is called the RF pulse. It is applied by an electromagnetic RF coil (Figure 33.10). This RF pulse now causes the net magnetisation vector of the hydrogen atoms to turn toward a direction that is at right angles to the original external magnetic field. Thus, the applied RF pulse has added energy to the atoms. When it is switched off again the atoms relax in various ways, the net magnetisation vector returns to its original direction and in doing so emits an RF signal that is received

Figure 33.10
Generation of the MR signal.



by the RF coil. The signal induces small currents in the RF coil that are called the MR signal. It is these signals that are analysed by the computer software to produce an image.

Photo 33.4
MRI brain scan.

The MRI image is surprisingly similar to a CT image but, of course, has not been produced with X-rays. Photo 33.4 shows a typical head scan MRI image with all the soft tissue detail. CT scans depend on tissue density, and ultrasound scans depend on echogenicity, but much of the complexity of the MRI image is due to a variation of properties such as proton density, chemical environment, magnetic susceptibility and the relaxation time of the hydrogen atom in various biological compounds. Radiologists can alter the duration and amplitude of the applied RF pulse to provide different types of images designed to clarify anatomy details or pathology details.

The main advantages of MRI as an imaging diagnostic tool are that it allows:

- excellent soft tissue contrast
- imaging in any plane, being especially useful in scans of the musculoskeletal system
- no use of ionising radiation.

MRI is the radiologist's choice for most spine and brain disorders, but it has not replaced CT, ultrasound or endoscopy as the choice for thoracic and abdominal disorders. Unfortunately MRI is very expensive, with running and maintenance costs very high, and the instrumentation is obviously not very portable. The applications of MRI are being developed rapidly, with certain paramagnetic contrast materials that increase soft tissue detail even further being developed.

Further developments

The gamma camera as used in tomography produces only a two-dimensional image. In a technique known as **single photon emission computed tomography** (SPECT), the gamma camera moves around the body as with CT scans, but the computer is programmed to analyse data coming from a single depth within the patient. Cross-sectional scans, similar to those produced by plain tomography, are obtained. The main applications of SPECT are in bone scanning, thallium-201 cardiac scanning and in cerebral or brain studies in which colour-enhanced cross-sectional images are obtained using radioactive iodine-123 as the γ emitter attached to a variety of tracer compounds, such as amphetamines.

Ch33-Walding 4th 25/8/04 1:52 PM Page 758

Radioisotopes such as nitrogen-13 and oxygen-15 produce positrons that are very short-lived. As they are emitted from the nucleus, they collide with an orbital electron in adjacent atoms and are annihilated, producing energy in the form of two γ photons travelling in opposite directions. A technique called **positron emission tomography** (PET) uses a circular array of detectors around a patient to search for these γ pairs at coincident times. A target molecule, such as glucose, is tagged with the positron emitter, such as nitrogen-13. The tagged solution is injected into the patient's bloodstream. The data from this process are again computer processed to produce a colour-enhanced image. A PET scan can be done quite quickly so that it can provide information on rapidly changing internal processes, such as brain activity. It has had good success in the study of epilepsy, locating quickly the deep-seated focal point of the epileptic activity in the brain. The radionuclides used for PET scans are very short-lived, ^{15}O — 2 minutes, ^{13}N — 10 minutes, ^{18}F — 2 hours, and thus must be produced by a cyclotron at the hospital site. At present, PET studies in Australia are carried out primarily at the Austin Hospital PET Centre in Melbourne.

The biochemical properties of the commonly used positron emitting radioisotopes are generally superior to those of the single γ emitters for functional medical imaging because the elements used are in fact the principal elements of the human body. Several positron emitting isotopes are now produced in Australian cyclotrons. Glucose metabolism images taken using radioactive glucose (^{18}F-fluoro-2-deoxy-D-glucose or FDG) provide unique clinical information in cardiology, neurology, oncology and psychiatry. PET will also have a major role in biomedical research with its ability to radiolabel compounds enabling in situ studies of biochemical processes.

— Practice questions

The relative difficulty of these questions is indicated by the number of stars beside each question number: * = low; ** = medium; *** = high.

Review — applying principles and problem solving

***1** Describe how you would explain the difference between medical physics and nuclear medicine.

***2** Calculate the de Broglie wavelength of electrons accelerated by a potential difference of 55 kV in an electron microscope.

***3** What device takes the place of optical lenses within an electron microscope in order to focus the electron beam?

***4** Explain why a technician producing a specimen for scanning electron microscopy does not need to produce a thin cross-section.

***5** Place the following names into chronological order of their medical physics discoveries. Briefly outline the contribution made by each person: Ray Damadian, Ernst Ruska, Wilhelm Roentgen, Heinrich Rohrer, Ian Donald, Godfrey Hounsfield.

***6** Outline the differences between the echogenicity of normal fatty tissue and abnormal lymphoma tissue in the body. What imaging technique makes use of these differences?

***7** Define these terms associated with diagnostic ultrasound: piezoelectric, Doppler ultrasonography, echocardiography and intercavitary scanning.

***8** Make a list of the radioisotopes typically used in the clinical applications and diagnosis of coronary heart disease and cardiology. What radiation is emitted and how are the radioisotopes originally produced?

***9** A person is given a technetium-99 lung scan. How does the total body dose equivalent in this hospital procedure compare with that of a person who does not require this procedure? To what form of radiation is the scanned patient subjected?

***10** A radiographer refers to a typical density 5 area on an X-ray plain film. What object or part of the body is it likely to be?

***11** Explain how a CT scan differs from a plain X-ray image. Imagine you were trying to explain the procedure to an elderly family relative in order to allay their fears.

***12** Magnetic resonance imaging in medical physics or radiology depends on what type of energy emitted by the hydrogen atom?

****13** Patient A reports to a hospital with an acute kidney disorder and patient B reports with suspected damage to the nerve spinal cord at the base of the brain. Predict which mode of diagnostic imaging a consultant radiologist might order for each patient. Explain your reasons.

****14** Using sketches, outline the differences between an X-ray tube and a computed tomography scanner.

****15** Use Table 33.4 to determine which X-ray procedure to the body produces, on average, the highest and the lowest dosage rates. Explain why this might be so.

****16** Write a short report on diagnostic medical imaging techniques that allow monitoring of moving fluids in the body.

Extension — complex, challenging and novel

*****17** Ionising radiation interacts with body tissues basically by destroying cell components. Explain why the weighting factor for alpha particles might be much higher than for protons or electrons. Refer to Table 33.3.

*****18** Compile a report on one of the following medical topics **(a)** to **(e)**. You should include the following sections in your report, where appropriate to the topic:

- Overview of the ailment (who gets it and why)
- Physics principles underlying the procedure
- Dangers associated with the procedure
- Success rates and future possibilities.

(a) Breast cancer is second only to lung cancer as a cause of death from cancer among women in Australia. Women have a 1 in 10 chance of developing it during their lifetime. Radiation therapy achieves success in about 50% of cases. Discuss.

(b) X-rays can be used to treat malignant melanomas although surgical removal is the treatment of choice. Discuss the difference.

(c) Laser treatment seems to be effective in repairing detached retinas, removing portwine birthmarks and tattoos. Discuss.

(d) Fluoroscopes were used in suburban shoe stores throughout Australia in the 1950s to get X-ray images of feet in shoes. In retrospect, this was a dangerous procedure. Why?

(e) The following conditions often require the use of radiation therapy. Research and report on: prostate, bladder and testicular cancers; Hodgkin's disease; uterine, ovarian and cervical cancers; carcinoma of the lung; Ewing's sarcoma.

****19** Read the following material from information available on the website of the Austin Medical Centre in Melbourne, and complete the questions following. You may also need to refer to other tables in the chapter.

All radiology (X-ray) and nuclear medicine tests involve the administration of radiation. Whether this is by way of an X-ray tube or radioactive material is of no consequence; it is still radiation. The levels of radiation exposure, however, differ between the two. The following table [Table 33.6] compares the effective dose from a range of radiological and nuclear medicine tests with the equivalent number of chest X-rays. There is also a column listing the length of time one would need be exposed to natural background radiation to receive the same exposure.

Table 33.6

NUCLEAR MEDICINE INVESTIGATION	EFFECTIVE DOSE (mSv)	EQUIVALENT NUMBER OF CHEST X-RAYS (mSv)	EQUIVALENT PERIOD OF NATURAL RADIATION
Radiography (normal X-rays)			
Extremities (e.g. knee)	0.01	5	1.5 days
Chest	0.02	1	3 days
Skull	0.1	5	2 weeks
Cervical Spine	0.1	5	2 weeks
Thoracic Spine	1.0	50	6 months
Lumber Spine	2.4	120	14 months
Hip	0.3	15	2 months
Pelvis	1.0	50	6 months
Abdomen	1.5	75	9 months
Biliary Tract	1.3	65	7 months
Intravenous Pyelogram	4.6	230	2.5 years
CT examinations (X-ray scans)			
Brain	2.0	100	1 year
Cervical spine	3.0	150	1.5 years
Thoracic spine	6.0	300	3 years
Chest	8.0	400	4 years
Abdomen	8.0	400	4 years
Lumber spine	3.5	175	1.8 years
Pelvis	7.0	350	3.5 years
Nuclear medicine — 99mTc			
Bone imaging	3.6	180	1.8 years
Cerebral perfusion (blood flow)	4.5	225	2.3 years
Lung perfusion	1.0	50	6 months
Myocardial perfusion	5.0	250	2.5 years
Thyroid imaging	1.0	50	6 months
DTPA renogram (kidneys)	1.6	80	10 months
DMSA renal	0.4	20	8 weeks
Hepatobiliary	2.3	115	14 months
Liver sulfur colloid	0.7	35	4 months
Gastric emptying	0.3	15	2 months

Questions

1 What does the term mSv mean in the column heading for 'effective dose'?

2 You have a typical X-ray for a broken arm and your mate says that you received some dangerous levels of radiation. How would you respond to his worries?

3 Why would you think that CT scans to parts of the body are much higher in effective dose than normal radiographs?

4 Give a reason for the increased dose to the lumbar rather than the thoracic spinal column.

5 A patient requires investigation of her heart blood flow patterns. Which would be the technique of least effective dose for the patient — technetium gamma scan or CT scan? What other factors do you think the specialist might consider, especially if it were for a possible heart-attack victim?

6 Find out the meaning of the terms hepatobiliary, DTPA and DMSA, used in the table.

7 From the table calculate an average effective dose from natural radiation causes per year.

Answers to Selected Questions

See our Web page for worked solutions to three star (***) questions. The address is on the back cover.

Chapter 1

1 (a) luminous intensity, temperature; **(b)** ampere, second; **(c)** yard, year **2 (a)** 1.08×10^9 km h^{-1}; **(b)** 6.71×10^8 miles/h **3 (a)** 1030 cm; **(b)** 0.0125 m; **(c)** 11.20 m; **(d)** 143.367 m; **(e)** 1.8×10^{-3} m; **(f)** 1.4×10^{-3} m^2; **(g)** 4.8×10^{-6} m^3 **4 (a)** 172.72 cm; **(b)** 10.72575 kg **5 (a) (i)** is larger; **(b)** use scientific notation; **(c) (i)** 1×10^{-1} s, **(ii)** 1×10^{-5} s **6 (a)** 5.52×10^{-4}; **(b)** 7.3×10^7; **(c)** 1.5×10^6; **(d)** 2.50×10^{-4} **7 (a)** 2.64×10^{-7}; **(b)** 2.8125×10^{10} **8** 5×10^{-28} m^3 **9 (a)** 4; **(b)** 3; **(c)** 1; **(d)** 4; **(e)** 1; **(f)** 6; **(g)** 4; **(h)** 5; **(i)** 1; **(j)** 4; **(k)** 3 **10 (a)** 8.383×10^1; **(b)** 2.00×10^1; **(c)** 5; **(d)** 2.205×10^4; **(e)** 1×10^2; **(f)** $1.000\ 10 \times 10^2$; **(g)** 1.999×10^3; **(h)** 2.222 2; **(i)** 4×10^4; **(j)** 5.070×10^{-2}; **(k)** 2.00×10^{-7} **11 (a)** 2; **(b)** 3; **(c)** 3; **(d)** 4 **12 (a)** 4.20×10^2 m^2; **(b)** 7.6×10^6 m^2; **(c)** 7.2×10^1 cm s^{-1}; **(d)** 3.71 cm^2; **(e)** 4.0×10^{-7} **13 (a)** 45.6; **(b)** 22.611; **(c)** 3.3×10^4 m or 0.00034; **(d)** 5.4×10^{-2} or 0.054; **(e)** 2.35×10^6 or 2 350 000; **(f)** 3.5×10^{-2} or 0.035 **14** 0.2 kg cm^{-3} **15 (a)** 110 cm^2; **(b)** 2 cm^3; **(c)** 115.0 cm **16 (a)** 10^{22}; **(b)** 10^{13}; **(c)** 10^{-10}; **(d)** 10^{-14}; **(e)** 10^5; **(f)** 10^5; **(g)** 10^{-6}; **(h)** 10^{-3} **17 (a)** 10^3 or 10^4 if converted to OM first; **(b)** 10^{12}; **(c)** 10^3 or 10^4 if converted to OM first **18** 2200 ± 300 mm^3 **19 (a)** 25.5 ± 0.5 mm, 174.5 ± 0.5 mm; **(b)** 25.5 mm ± 1.96%, 174.5 mm ± 0.29%; **(c)** 200.0 ± 1.0 mm; **(d)** 4450 ± 100 mm^2 **20** 223 000 ± 5000 m^2, 1980 m ± 20 m **21 (a)** 27.6 ± 0.41; **(b)** 10.35 ± 0.06 **22 (a)** 5.3%; **(b)** 5.3%; **(c)** 5.3%; relative error remains the same even as the speed changes **23** 330 ± 20 ohms **24 (a)** 8.49 cm; **(b)** 9.8 mm **25** 3×10^8 mm **26 (a)** 4.00×10^7 m; **(b)** 1.08×10^{21} m^3; **(c)** 1.08×10^{12} km^3 **27** 66 m s^{-1} **28 (a)** 3.55876×10^3; **(b)** 40.00 (or 4.000×10^1; **(c)** 7.9×10^4; **(d)** $2.003\ 26 \times 10^5$; **(e)** 1.994×10^3; **(f)** 20.009 (or 2.0009×10^1; **(g)** 5.00×10^{-2}; **(h)** 2.5×10^6; **(i)** 8×10^{-7}; **(j)** 5×10^6 **29 (a)** 3.4×10^8; **(b)** 1.5×10^4; **(c)** 4.0×10^{-2}; **(d)** 3.0×10^9; **(e)** 5.3×10^{-11}; **(f)** 6.4×10^{15} **30 (a)** 10^8; **(b)** 10^8; **(c)** 10^5; **(d)** 10^{-4}; **(e)** 10^{-7}; **(f)** 10^7; **(g)** 10^{-6} **31 (a)** 3; **(b)** 4; **(c)** 2; **(d)** 3 **32 (a)** 2.40 V ± 0.8%; **(b)** 3.25 A ± 2%; **(c)** 25.4 mm ± 2%; **(d)** 0.0035 T ± 3%; **(e)** 325 cm ± 3% **33 (a)** micrometer ± 0.005 mm, vernier ± 0.05 mm; **(b)** 28.4 mm length, 16.444 mm diameter; **(c)** radius = 8.222 ± 0.005 mm (± 0.0608%), length = 28.4 ± 0.05 mm (± 0.176%); **(d)** 6030 mm^3 ± 0.2976% (± 17.9 mm^3); **(e)** 9.35×10^{-3} g mm^{-3} ± 0.648% **35** *Micro* means small wavelength; *pico* means smaller still **36** 1.8×10^{12} furlongs/fortnight **37** $1.1 \times 10^{-2} \pm 7 \times 10^{-4}$ g/mm **38** 3×10^{23} fermi **39** $690 (to 2 SF) **40** 3×10^7 words **42 (b)** 11 March; **(c)** 2 h 56 min **43** 3.25×10^{-6} m **44** True, 8.33×10^{24} molecules in glass, 5.21×10^{21} glasses in ocean **45** 5.97×10^{10} kg

Chapter 2

1 (a) 2 km N; **(b)** 8 km; **(c)** 6 km; **(d)** 4.5 km E26°N **2** 60 m East **3 (a)** 188 cm, 120 cm; **(b)** 377 cm, 0 cm; **(c)** 754 cm, 0 cm; **(d)** 94 cm, 85 cm **4 (a)** 20 m s^{-1}; **(b)** 9.3 m s^{-1} (33 km h^{-1}); **(c)** 4200 m; **(d)** 660 km; **(e)** 20 s; **(f)** 2 hours **5** 0.26 s **6** 4.8 s **7 (a)** speed 3.89 m s^{-1}; **(b)** velocity 2.78 m s^{-1} N53°W **8** 28 m s^{-1} **9 (b)** 2–4 s, 5–6 s; **(c)** 0–2 s, 4–5 s; **(d)** same at 0–2 s, 5–6 s, 6–9 s **10 (a) (i)** 6 m s^{-1}, **(ii)** 0 m s^{-1}, **(iii)** 4 m s^{-1}, **(iv)** 0 m s^{-1}, **(v)** –5.0 m s^{-1}; **(b)** 0; **(c)** 3.3 m s^{-1} **11** 6 m s^{-2} **12 a)** 8 m s^{-1}, 4 m s^{-2}; **(b)** 38 m s^{-1}, 1.0 s; **(c)** 0 m s^{-1}, –20 m s^{-1}; **(d)** –7 m s^{-1}, –2 m s^{-2}; **(e)** –5.65 m s^{-1}, –0.65 m s^{-1} **13** 6.7 m s^{-2} **14** 69 s **15** 0.54 s **16** 4×10^{-8} s **17 (a)** approx. 13 m s^{-1}; **(b)** 10 m s^{-1} **18 (a)** displacement 100 m, distance 200 m; **(b)** 0.5 m s^{-2}, –1 m s^{-2}, 1 m s^{-2}; **(d)** 0 s, 30 s, 50 s; **(e)** 0–20 s, 20–40 s, 40–50 s; **(f)** nowhere **19 (a)** 150 m; **(b)** 350 m; **(c)** 20–30 s, a = –1.5 m s^{-2}; **(d)** 0s, 30 s, 60 s; **(e)** 20–30 s, 30–40 s; **(f)** 40–50 s **20 (b)** 15.7 km **21** 11.25 m, 7.5 m s^{-1}; **(b)** 83 m s^{-1}, 34.7 m s^{-2}; **(c)** 131.25 m, 7.5 s; **(d)** 31.3 m s^{-1}, 14.9 s; **(e)** 18 m s^{-1}, 4 s; **(f)** 18.75 m, –2 m s^{-2}; **(g)** –10 m s^{-1}, –7.5 m s^{-2} **22 (a)** 6 m s^{-2}; **(b)** 36.75 m **23** 0.18 m s^{-2} **24** 6.4×10^{-4} s **25** 80 m **26 (a)** –22.3 m s^{-1}; **(b)** 2.23 s **27 (a)** 61.2 m; **(b)** 3.5 s; **(c)** same, 3.5 s **28 (a)** 4.9 s; **(b)** –44.5 m s^{-1} **29 (a)** 3.72 m s^{-1}; **(b)** 1.72 m s^{-1}; **(c)** 15 cm higher **30** constant acceleration of –10 m s^{-2} **31** A **33 (b)** slope at C = 57 cm s^{-1}, slope at F = 140 cm s^{-1}, same as calculated v; **(d)** area = 10 cm, about same as displacement at G; **(e)** slope = 1375 cm s^{-1} approx., about same as calculated acceleration **34 (c)** 90 cm s^{-1}; **(e)** same; **(g)** 1100 cm s^{-2}; **(i)** same; **(j)** area = 6 cm, about same as table 6.1 cm **35** 20 s, 300 m **36** No, v = 60 km h^{-1} **37 (a)** 50 m s^{-1}; **(b)** 33.3 km h^{-1}; **(c)** 10 200 m; **(d)** 715 m; **(e)** 20 s; **(f)** 55 846 h or 6.15 y **38 (a)** 18.77 km h^{-1}, 5.2 m s^{-1}; **(b)** 3940 s or 1 h 5 m 41 s **39 (a)** 4023 m; **(b)** 42.116 m s^{-1} or 151.6 km h^{-1} **40** 0.55 s **41 (a)** 76 m north; **(b)** 9 m s^{-1}; **(c)** 2.8 m s^{-1} N **42** 35 km h^{-1} **43 (a)** 15 m; **(b)** 40 m; **(c)** 0–1 s; **(d)** 1–4 s; **(e)** 0–1 s, 4–6 s; **(f)** –12.5 m s^{-1} **44 (a)** 17 m s^{-2}; **(b)** –82 m s^{-2}; **(c)** 244 m s^{-1}; **(d)** 6.9 s; **(e)** 6.7×10^{-3} s; **(f)** 2.4×10^{-3} m s^{-1} **45 (a)** 10.3 m s^{-2}; **(b)** 18.5 m s^{-1}; **(c)** 29 m; **(d)** 20.5 m s^{-1} **46 (a)** 162.5 m; **(b)** 392.5 m; **(c)** 50–60 s; **(d)** 0 s, 37 s, 60 s; **(e)** at no stage **47 (a) (i)** 3.375 m, **(ii)** 4.5 m s^{-1}; **(b) (i)** 286 m s^{-1}, **(ii)** 204 m s^{-2}; **(c) (i)** 735 m, **(ii)** 17.3 s; **(d) (i)** 40.3 m s^{-1}, **(ii)** 13.1 s; **(e) (i)** 17 m s^{-1}, **(ii)** 20 s; **(f) (i)** 44.5 m, **(ii)** –9.3 m s^{-2}; **(g) (i)** –72 m s^{-1} **(ii)** –0.19 m s^{-2} **48 (a)** 27.8 m; **(b)** 5.5 s **49 (a)** 10.2 m; **(b)** 2.1 s **50** 6.7 m s^{-2} **51** 64.9 m s^{-2} **52** 0.021 s **53 (a)** 396 m s^{-2}; **(b)** 1.89 s **54 (a)** 3 061 224 s (35.4 days); **(b)** 4.6×10^{13} m **55 (a)** –4 m s^{-2}; **(b)** 50 m; **(c)** 8.5 s; **(d)** –20 m s^{-1}; **(e)** –14.1 m s^{-1} **56 (a) (i)** 2 s, **(ii)** 5 m; **(b) (i)** +45 m s^{-1}, **(ii)** 9 s; **(c) (i)** +27.5 s, **(ii)** +38 m **57 (a)** –22.6 m s^{-1}; **(b)** 2.6 s **59 (a)** liberty; **(b)** none; **(c)** yes; **(d)** 40 km h^{-1} = 11.1 m s^{-1}, 70 km h^{-1} = 19.4 m s^{-1}; **(e)** price, mass, fuel consumption, warranty, accessories etc.; **(f)** not true — could depend on mass of car as well; **(g)** not supported — mass could have an effect **60** Chris is still faster **61** –14.3 m s^{-2} **62 (a)** 23.87 m s^{-1}; **(b)** 75.3 s **63 (a)** 0.04 s; **(b)** 0.17 s; **(c)** bottom is about 1/4 of time **64** Four times as high **65 (a)** –11.7 m s^{-1} **66** Acceleration most thrilling. Kitty: 46.9 m s^{-2}, Eli: 805 m s^{-2} **67** Both the same **68 (a)** $s_1 - s_2 = 10 - 10t_1$; **(b)** $v_1/v_2 = t_1/(t_1 - 1)$ **69** Case 1: a = 15.4 m s^{-2}, Case B: a = 11.8 m s^{-2}. The Case 1 rider probably reached top speed before the 400 m line. The rider in Case 2 was travelling at 305 km h^{-1} at the finish line but reached this after 5 s and held the speed constant **70** At current mass, a = 16.7 m s^{-2}; at heavier mass, a = 15.7 m s^{-2} **71 (a)** 30 knots = 30 nautical miles/h; **(b)** 54.9 km h^{-1} **73** 38.843 058 mph or 119.813 82 mph. See our Web Page for the solution **74** $s \propto t^2$ (correct), $v \propto s$ (incorrect, should be $v^2 \propto s$); **75** 160 s; **77** at 0 m, 4 m, 16 m, 36 m, 64 m

Chapter 3

1 (a) 12.2, 0.70, 35°; **(b)** 10.2, 0.61, 52°; **(c)** 17.9, 0.45, 27°; **(d)** 206, 0.25, 14° **2 (a)** 36 m, N56°E; **(b)** 34 m s^{-1}, W62°S; **(c)** 22.3 N, W71.6°N **3 (a)** 30 N, 53° to horizontal; **(b)** 3.1×10^{-3} N, E36°N; **(c)** 440 N, E8.2°N **4** Scalar: mass, height, time; Vector: velocity, acceleration, displacement **5** 29 N, S21.6°W **6 (a)** 50 m s^{-1} North; **(b)** 60 m s^{-1} E; **(c)** 43 m s^{-1} E35.5°S; **(d)** 53.8 m s^{-1} N22°W **7 (a)** 28 m s^{-1} up; **(b)** 39 m s^{-1} N39.8°E; **(c)** 28 km h^{-1} at E45°S **8** 40 m s^{-1} vertical **9** $\boldsymbol{F}_V$ = 50 N; $\boldsymbol{F}_H$ = 87 N **10 (a)** $\boldsymbol{F}_H$ = 75 N; **(b)** $\boldsymbol{F}_V$ = 27 N **11 (a)** $\boldsymbol{F}_P$ = 321 N; **(b)** $\boldsymbol{F}_\perp$ = 383 N **12 (a)** 8 m s^{-1} W; **(b)** 8 m s^{-1} E; **(c)** 4 m s^{-1} W **13** Increases their speed relative to the air **14** 7.06 h **15 (a)** 0.5 h; **(b)** 1.0 h **16 (a)** 4.9 m s^{-1} N24°E; **(b)** 44.4 m **17 (a)** heads west; **(b)** 33.3 h; **(c)** 18 km h^{-1} W; **(d)** must head south; **(e)** 44.7 h; **(f)** 13.4 km h^{-1} **18 (a)** treble W; **(b)** halve W **19 (a)** A; **(b)** D; **(c)** B; **(d)** C **21 (a)** 3.14 cm, 0.325 m/y, 6.0 m/s; **(b)** 44 cm, 1.625 m, 48 m; **(c)** 18.8 cm, 0.81 m, 18 m **22 (a)** 2 m west; **(b)** 7.1 m s^{-1} E45°S; **(c)** 5.8 m s^{-2} N56°W; **(d)** 9.2 m W22.4°N **23 (a)** 23.5 m, W12°S; **(b)** 50 m, E33°N; **(c)** 24.8 m S48°E; **(d)** 2 m north **24 (a)** 175 km h^{-1} rebound; **(b)** 75 km h^{-1} W37°S; **(c)** 27.0 m s^{-1} 80° to horizontal; **(d)** 451 m s^{-1} E53°N **25 (a)** 100 km N, 0 km E; **(b)** north component 43 m s^{-1}; east component 25 m s^{-1}; **(c)** north 19 N, east 16 N **26 (a)** 13 m s^{-1} N67°E; **(b)** 60 s; **(c)** 720 m **28** $H \propto I^2$ **29 (a)** 0–4 y, 23.6 cm/y; **(b)** 17–18 y, 52 kg/y; **(c)** sickness, –27 kg/y; **(d)** height, 10.3 cm/y; mass 8.71 kg/y; **(e)** height, 242.5 cm; **(f)** height 285 cm; **(g)** birth height and mass not zero **30** 16 m s^{-1} S; **(b)** 11.3 m s^{-1} W45°S; **(c)** 5.0 m s^{-1} W18°S **31 (a)** 87 N; **(b)** 50 N **32** 502.5 km h^{-1} E5.7°S **33 (a)** $A_P = A_Q/16$; **(b)** $V_P = V_Q/64$; **(c)** $m_P = m_Q/64$; **(d)** $\rho_P = \rho_Q$ **34 (a)** $r_1 = \sqrt[3]{35}\, r_2$; **(b)** r_2 = 15.3 cm **35 (a)** 600 N; **(b)** 800 N **36** 38° **37** 0.098 m, 100 Ω **38** n = 3 has sharper curve

Chapter 4

1 (a) 2.5 N N; **(b)** 101 N up; **(c)** 32 N, S72°W; **(d)** 16 N, N22°W **2 (a)** balanced; **(b)** Unbalanced. $\boldsymbol{F}_W$ causes book to accelerate downwards **3 (a)** Slowly: string A breaks because string A has to support the weight of the mass plus the pulling force; **(b)** String B breaks because of the inertia of the mass. String A would not experience the pulling force immediately **5** 0.123 kg **6** No! V = 1.76 m^3, m = 424 kg **7 (a)** 2941 N, 2.94 m s^{-2}; **(b)** 3.3 m s^{-1}, 1.67 m s^{-2}; **(c)** 100 kg, 3 s; **(d)** 5.5×10^{-4} N, -2.78×10^{-3} m s^{-2}; **(e)** 68 kg, 8.8 m s^{-1} **8** 9000 N **9** 96 000 N **10 (a)** false; **(b)** true; **(c)** true; **(d)** false **11 (b)** $\boldsymbol{F} \propto \boldsymbol{a}$ **12 (b)** $\boldsymbol{F} \propto \boldsymbol{a}$; **(c)** equal: m = 0.850 kg; **(d)** Steeper and wouldn't pass through origin. Need more $\boldsymbol{F}$ to produce the same acceleration; **(e)** Keep the hanging mass at 100 g and remove masses from trolley **13 (a)** Racquet pushes back; **(b)** Road pushes up and forward on horse; **(c)** Ground pushes up and forward on horse and log; **(d)** Ground pushes up on beetle **14 (a)** $\boldsymbol{F}_N = -\boldsymbol{F}_W$ = 25 N up; **(b)** $\boldsymbol{F}_N$ = 5000 N but table would collapse **15 (a)** $\boldsymbol{F}_W$ on Saturn is 1.07 times as great; **(b)** Because of Earth's rotation and shape, the acceleration will be affected **16 (a)** 33 N; **(b)** Use an inertia balance (or springs): convert mass to an equivalent Earth weight $\boldsymbol{F}_W$ = m × 9.8 N **17 (a)** 0; **(b)** 0; **(c)** 735 N; **(d)** 665 N **18 (a)** 246 N; **(b)** 172 N; **(c)** 5.73 m s^{-2} **19 (a)** T = 4 N; **(b)** 2 m s^{-2} **20** Figure (a), **(i)** 5 m s^{-2}; **(ii)** 15 N; Figure (b), **(i)** 2.5 m s^{-2}; **(ii)** 7.5 N **21** 0.55 **22** 35 380 N **23 (a)** 276 N; **(b)** 585 N; **(c)** 0.47 **24** 348 N **25** 56.7 m **26** 10.9 m **27** 13.1 m s^{-1} **28 (a)** 50 N; **(b)** 38 N; **(c)** 43.3 N; **(d)** 66 N **29** 0.47 **30** 0.17 **31 (a)** 25 N north; **(b)** 10.1 N up; **(c)** 36 N, S74°W; **(d)** 1.8 N, W 56° N **32** 202.5 g **33 (a)** 1st; **(b)** 3rd; **(c)** 2nd; **(d)** 1st **34** (a) **35 (a)** 15.36 N east; **(b)** 0.1 N; **(c)** 25 N; **(d)** 1.56 N; **(e)** 3×10^5 N **36 (a)** 0.84; **(b)** same force (2100 N) **37 (a)** 8.8×10^{15} m s^{-2}; **(b)** 1.8761×10^7 m s^{-1} **38 (a)** The retarding (friction) force; **(b)** 1.5 m s^{-2}; **(c)** 2 N, 3 N; **(d)** 1 kg **39** 3.2 N **40 (a)** Sodium 0.97 g cm^{-3}; potassium 0.86 g cm^{-3}; **(b)** No **41 (a)** 3rd law — Throw your jacket off behind you to accelerate; **(b)** 2nd law — The faster you throw, the greater the acceleration; **(c)** 1st law — To change direction, throw something to the side **42** Rolling and sliding friction is independent of area. Starting (static) friction varies with area **44 (b)** 3.3 kg **45 (b)** 0.25 m s^{-2}, 0.175 m s^{-2}, 0.11 m s^{-2}; **(d)** $m \propto 1/\boldsymbol{a}$; **(e)** 0.175 N in each case; **(f) (i)** 0.135 m s^{-2}, **(ii)** 2.72 s **46 (a)** combustion 60%, engine 21.4%, transmission 1.6%, accessories 2.2%, tyres 5.2%, air 4.9%, brakes 4.7%; **(b)** decrease i, ii, v, vi, vii; increase iii, iv **47** Steel plate drags along ground. Loss of control; dig up road **48 (a)** 2.3 N; **(b)** 0.28 **49** Wind force 2.1×10^{-3} N; tension 3.7×10^{-3} N **50 (a)** 2 m s^{-2} upwards; **(b)** 22 s **51** For one person, $\boldsymbol{a}$ = 2.72 m s^{-2}; for two people, $\boldsymbol{a}$ = 2.69 m s^{-2}; one person would go faster **52** 59 m s^{-1} (212 km h^{-1}) **53** 93 N **54** 15 m s^{-1} **55** 14 m s^{-1} **56** 15.3 cm **57** 70 N, 0.58 **58** 16.6 m s^{-1} **59** 15.3 m s^{-1} **61** 464 N

Chapter 5

1 (a) –125 m; **(b)** 125 m; **(c)** –56 m s^{-1} at 63° to horizontal **2 (a)** 4.5 s; **(b)** –46 m s^{-1} 80° to horizontal; **(c)** 36 m **3 (a)** u_V = 12.6 m s^{-1}, u_H = 27.2 m s^{-1}; **(b)** 7.9 m; **(c)** 2.52 s; **(d)** 68.5 m **4 (a)** 1.5 m s^{-1}; **(b)** 20 m s^{-1}, 4.3° (to horizontal); **(c)** 6.6 m; **(d)** 2.3 s; **(e)** 45.8 m **5 (a)** 5.0 s; **(b)** 47 m s^{-1}, 72.7° (to horizontal); **(c)** 70 m **6** 7.3 m s^{-1} **7 (a)** could clear the ferris wheel by 7.0 m; **(b)** 68 m **8 (a)** 12.5 m s^{-1}; **(b)** 23 750 kg m s^{-2} **9 (a)** 13.3 m s^{-2}; **(b)** 94.2 s **10 (a)** 2.4×10^9 m; **(b)** 1023 m s^{-1}; **(c)** 2.7×10^{-3} m s^{-2}; **(d)** 2×10^{20} kg m s^{-2} **11 (a)** Centripetal force is provided by the force of attraction between the water and the cloth. As long as this force is greater than mv^2/r, the water will remain on the clothes; **(b)** No, you need to increase speed; **(c)** Probably not — the force between the water and the clothes is too strong; **(d)** moves through the cloth **12** 27.8 N **13** 7.74 m s^{-1} **14 (a)** 2550 N; **(b)** 24 m s^{-2}; **(c)** No; **(d)** 50 N; **(e)** 77.4 m s^{-1} **15** 22.5 N **16 (a)** 0.52 rad s^{-1}; **(b)** 0.1 m s^{-1} **17 (a)** 24 rad s^{-1}; **(b)** 3.8 rev/s or 229 rpm **18** Rotation of the Earth; molten core not taken into account **19** 5.48 N m^{-1} **20** Chair 12.5 kg, astronaut 54.4 kg **21 (a)** 1.644 s; **(b)** 1.647 s **22 (a)** 11.3 m s^{-2}; **(b)** 1.02 s; **(c)** 3.3 m s^{-2} **23 (a)** into his hands; **(b)** fall behind him; **(c)** fall to side of him **24 (a)** 31.25 m; **(b)** 37.5 m; **(c)** 29.1 m s^{-1} 59° to horizontal **25 (a)** 4.33 m s^{-1}; **(b)** 17.8 m s^{-1} 76° to horizontal **26 (a)** u_H = 17.2 m s^{-1}, u_V = $^{+}$24.6 m s^{-1}; **(b)** $^{+}$30.2 m; **(c)** 4.92 s; **(d)** 84.6 m s^{-1} **27 (a)** $^{+}$5.3 m s^{-1}; **(b)** 28.2 m s^{-1} at 6° to horizontal; **(c)** $^{+}$5.26 m; **(d)** 2.1 s; **(e)** 59.2 m **28 (a)** 31.1 m s^{-1}; **(b)** 45°; **(c)** $^{+}$24 m; **(d)** 4.4 s **29 (a)** 10 m s^{-2}; **(b)** 22 500 N; **(c)** 18.8 s **30** 3 '*g*' or 30 m s^{-2} **31** 9.0×10^{22} m s^{-2}; **32 (a)** 57.3°; **(b)** 487°; **(c)** 1.57 rad; **(d)** 0.52 rad s^{-1}; **(e)** 15.9 rev s^{-1}; **(f)** 6.28 m; **(g)** 30 m s^{-1} **33 (a) (i)** 4.1 m s^{-1}, **(ii)** a_c is the same as the ferris wheel is rigid and it is assumed that v is constant; **(b)** at top 383 N, at bottom 917 N **34** 64.7 rpm **35 (a)** 209 rad s^{-1}; **(b)** 136 m s^{-1} **36 (a)** 10 N m^{-1}; **(b)** 0.888 s; **(c)** 1.12 Hz **37 (a)** 1.346 s, **(b)** 1.347 s **38 (a)** 0.42 m; **(b)** 1.3 m s^{-2} **41 (a)** true; **(b)** true; **(c)** true **42** 0.18 m **43 (a)** 4.58 s; **(b)** 45.8° to horizontal; **(c)** 150.7 m **44** He was 0.328 m off maximum (96.5%) **45 (a)** –233 m s^{-1} at angle of 53° to vertical; **(b)** 837 m; **(c)** 263 m s^{-1} at 44.8° to horizontal **46 (a)** 7.2 m s^{-1}; **(b)** 1225 N **47 (a)** 18 m s^{-2} (1.8 g); **(b)** will remain conscious; **(c)** 2240 N **49** 2.2 s **50 (a)** 80 m; **(b)** 76°; **(c)** 12.8 m, 104 m

Chapter 6

2 1.05×10^4 km s^{-1} **3** 3.4×10^4 km s^{-1} **4** 9.8×10^9 y **5** 1.01×10^{-6} m **6 (a)** 1.53×10^{27} watts; **(b)** 3.9×10^{10} **7** Red has longer λ therefore lower temperature **8** False. Will give off radiation with $\lambda = 1.06 \times 10^{-5}$ m **9** False. Twice λ and twice energy but not twice rate **10** 10^{30} **11** expansion **12** least favoured **13** last 10^{37} years **14** true **16 (a)** 2.64×10^9 s; **(b)** 83.5 years **17** 4.51×10^{12} m **18** 4.34×10^6 s (50.2 days) **19** 3.54×10^{22} N **20** 2.2×10^{16} N **21 (a)** 60 m s^{-2}; **(b)** 45 000 N; **(c)** 11.2 min **22 (a)** 0.034 m s^{-2}; **(b)** 2.2 N; **(c)** 465 m s^{-1} **23 (a)** 6100 N; **(b)** 5590 m s^{-1}; **(c)** 4 hours **24 (a)** $\boldsymbol{F}_{g2} = 1/4\, \boldsymbol{F}_{g1}$; **(b)** $\boldsymbol{v}_2 = 0.5\, \boldsymbol{v}_1$; **(c)** $T_2 = 4\, T_1$ **25 (a)** 7279 m s^{-1}; **(b)** 1.8 h **26 (a)** 2.37 km s^{-1}; **(b)** 8900 m s^{-1}; **(c)** 59 500 m s^{-1}; **(d)** 618 000 m s^{-1} **27 (a)** 86.8 N; **(b)** 17.4 m s^{-2} **28 (a)** 7328 N; **(b)** 25.4 N; **(c)** 7302 N; **(d)** 750 kg **29** 0.61 m s^{-2} **30** 3 Earth radii above surface **31** Graph shows $\boldsymbol{g} \propto 1/d^2$ **32** 9.27×10^{21} km **33** 0.0427 c **34** 11.0 km s^{-1} Mpc^{-1} **35 (a)** age $\propto 1/H_0$ $\therefore$ H_0 of 50 is younger; **(b)** 19.6 billion years **36** 9.38×10^{-7} m **37** 32 days **38 (a)** 7.4×10^{-12} N; **(b)** 3.56×10^{22} N; **(c)** 9×10^{-56} N **39** 8.2 m **40 (a)** 0.31 m s^{-2}; **(b)** 708 N; **(c)** 10 670 s (2.96 h) **41 (a)** 14.7 N; **(b)** 9.53 m s^{-2}; **(c)** nil $\boldsymbol{g}$ = 0 **42 (a)** 6254 N; **(b)** 6064 m s^{-1}; **(c)** 3.17 h **43 (a)** 36 176 N; **(b)** 7810 m s^{-1}; **(c)** 5277 s (1.5 h) **44** 640 m s^{-1} **45 (a)** 1.6 m s^{-2}; **(b)** 1.37 s **46 (a)** 9.8 m s^{-2}; **(b)** 1.57 m s^{-2}; **(c)** 0.61 m s^{-2}; **(d)** 1.6 m s^{-2}; **(e)** 275 m s^{-2} **47 (c)** graph shows $\boldsymbol{F} \propto 1/d$ or $1/d^2$; **(d)** confirms $\boldsymbol{F} \propto 1/d^2$; **(e)** slope = 8.35×10^{-12} N m^2 hence $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}; **(f)** F = 9.2×10^{-10} N **48 (a)** third; **(b)** be flat; **(c)** yes – elastic potential energy; **(d)** magnetic boots, velcro; **(e)** no, only that relying on gravity **49** 6400 km **50** Some possibilities **(a)** half the mass but same radius; **(b)** double the radius, keep mass constant; **(c)** double the radius, keep mass constant **51** 0.94 **52** Linear, therefore $r^3 \propto T^2$ **53 (a)** 3.4×10^8 m; **(b)** 0.90 of the distance to the moon; **(c)** No. In circular orbit you would experience free-fall and feel weightless **54 (a)** 9.8011 m s^{-2} — No noticeable difference; **(b)** Same as on Earth's surface — no difference **55 (a)** 9.2×10^{11} m s^{-2}; **(b)** 2.8×10^8 m s^{-2} **56** Spring clock not affected by gravity therefore it will keep time. Pendulum won't. They won't agree with each other **57 (a)** 3.54×10^7 m; **(b)** 410 m s^{-1}; **(c)** 0.030 m s^{-2}; **(d)** 1.79 N; **(e) (i)** towards centre of Earth, **(ii)** perpendicular to Earth's axis through Brisbane, **(iii)** towards centre of Earth **58** 4.9×10^{10} m s^{-2} (can't escape, $v > c$) **59** 6.5×10^{23} kg **60 (a)** 3.8×10^{-7} m; **(b)** Procycon (bigger surface area); **(c)** Same

Chapter 7

1 1282 Pa **2** 12 000 Pa **3** 1.25×10^6 Pa **4 (a)** 1270 N; **(b)** 5.3×10^5 Pa; **(c)** 120 m **5 (a)** 9690 N; **6 (a)** 8.97 kg, **(b)** 89.7 N, **(c)** 78 N **7** 0.6 g cm^{-3} **8 (a)** 0.8 N; **(b)** 80 g **9** 1670 N **10** 3.8×10^6 N **11** 400 kg **12** 103.36 kPa **13** 307.3 kPa **14** 20 635 Pa **15** 219 600 Pa **16 (a)** 11 000 Pa; **(b)** 112.3 kPa **17 (a)** 401 200 Pa; **(b)** 197 N; **(c)** 269 m **18 (a)** 25 N; **(b)** 4.1 g cm^{-3} **19 (a)** 0.8; **(b)** 800 kg m^{-3} **20 (a)** 150 g; **(b)** 10.05 kg m^{-3} **21** 2.80 **22 (a)** 240 N; **(b)** 11.52 N; **(c)** 228.48 N (= 22.8 kg) **23** 0.02 cm^{-3} (20 kg m^{-3}) **24** 5.5 cm **25 (a)** 0.94 g cm^{-3}; **(b)** 92% submerged **26** Fresh water less buoyant therefore floated at lower level **27** Glass bulb expands and floats higher. Real density would be a lower value **28 (a)**

downward force becomes less therefore rises higher; **(b)** water level in pond will fall; **(c)** perhaps use ball bearing inside a test tube floating in a measuring cylinder **29** Equalise pressure on liquid inside can otherwise the low pressure stops liquid flowing out **30** 1. Weigh in a vacuum (use rigid container); 2. Alternatively weigh a rigid container in air, add the 1 L H_2 and reweigh **31 (a)** 0.58; **(b)** rises **32 (a)** 7.38×10^5 N; **(b)** 74 tonnes **33 (a)** ii; **(b)** ii; **(c)** ii; **(d)** all the same **34** Top and bottom holes — 34.6 cm out from base; middle hole — 40 cm from base **35** No, the water level is the same. The ice cubes have bigger volume than liquid water but the cubes project over the top of the glass **36** H_2 balloon could only lift 1.14 times the load of the He balloon **37** Equal. The weight of water lost from glass equals weight of block **38** Cork will sink. **39** 1087 kg m^{-3} **40** 19.4 kPa **41** 0.5 m **42** 1.47 g cm^{-3} **43** Water level still higher. The water column is 13.6 times the height of the mercury column.

Chapter 8

1 0.92 m from 2.5 kg end **2** 4.61×10^6 m from Earth **3 (a)** 1.6 kg m s^{-1}; **(b)** 30 000 kg m s^{-1}; **(c)** 1.8×10^{29} kg m s^{-1} **4** 10 kg m s^{-1} west **5** 2.57 s **6** 79 200 N s **7** 1.67 m s^{-1} W **8** 0.1 m s^{-1} backward **9** 6.25 m s^{-1} **10** 7.2 m s^{-1} **11** 4.1 m s^{-1} in same direction **12 (a)** 227 m s^{-1}; **(b)** mass of rifle **13 (a)** 11 020 N; **(b)** 53 m s^{-1} **14 (a)** 1.81 m s^{-1} forward; **(b)** −1.86 m s^{-1} (backwards) **15** 31.25 kg **16** 17.39 m s^{-1} at 135° to either neutron **17** 26.9 m s^{-1} S42°E **18** 2.5 m s^{-1} E37°S **19** 3.53 m s^{-1} N38°E **20** 36 N m **21** 5.6×10^{-4} kg m^2 s^{-1} **22** 4×10^{-3} kg m^2 s^{-1} **23** 0.89 m from 20 kg end **24 (a)** 2 kg m s^{-1} E; **(b)** 1.7 kg m s^{-1} N35°E; **(c)** 800 kg m s^{-1} N **25** 1×10^{-22} kg m s^{-1} **26** 66 000 N s **27 (a)** 10.0 N s; **(b)** 45.5 m s^{-1} N **28** 3.25 cm s^{-1} **29** 9 m s^{-1} east **30** 2.25 m s^{-1} **31** 8.8×10^{33} kg m^2 s^{-1} **32** 4.6 m s^{-1} E55°S **33 (a)** air table minimises friction; **(b)** to the right; **(c)** blood pumping in opposite direction (from auricle to ventricle); **(d)** 0.025 m s^{-2}; **(e)** 0.021 kg m s^{-1} **34 (a)** zero; **(b)** 2.09×10^{-20} kg m s^{-1}; **(c)** 53 470 m s^{-1} **35** 34 m s^{-1} at 11° below horizontal **36** 3.5 m s^{-1} S38°E **37** 13 750 m s^{-1}

Chapter 9

1 (a) 87.5 J; **(b)** 170 J; **(c)** 50 000 J **2 (a)** 18 000 J; **(b)** by the piano **3 (a)** Horse A 8400 J, Horse B 6000 J; **(b)** 14 400 J **4 (a)** 400 J; **(b)** 40 000 J; **(c)** 280 J **5** $\boldsymbol{v}_A = 0$, $\boldsymbol{v}_B = 5$ m s^{-1} W **6** 0.25 m s^{-1} **7 (a)** 6 m s^{-1}; **(b)** 432 kJ **8 (a)** 62.5 J; **(b)** 62.5 J **9** 3.3 m **10** 48 200 J **11** 3000 kW **12 (a)** 600 W; **(b)** 3420 W **13 (a)** $E_{k(init)}$ = 110 000 J, $E_{k(final)}$ = 247 500 J; **(b)** 137 500 J; **(c)** 6.875 kW **14** 178 571 N **15** 746 W **16** 13 m s^{-1} **17 (a)** 6384 J; **(b) 6.9** m s^{-1} **18** 150 m s^{-1} **19** 15.3 J **20 (a)** 4 J; **(b)** 200 N m^{-1}; **(c)** 2.89 J **21** 0.04 m **22 (a)** 30%; **(b)** heat; not lost, just transferred **23** 740 J **24** 48 J **25** 1.92 J **26** 210 000 J **27** $E_{k(init)}$ = 80 J, $E_{k(final)}$ = 58.5 J. Not elastic **28** 700 J **29** 557 W **30** 13 m s^{-1} **31** 4.7 J **32 (a)** 0.36 J; **(b)** 22 N m^{-1}; **(c) (i)** 0.11 J, **(ii)** 0.44 J **33 (a)** run-up 1, take-off 2, flight 3–6; **(b)** at beginning of run $E_k = 0$ and increases; **(c)** no, this is GPE of centre of mass; **(e)** no, lying on back has lower GPE **34 (a)** 1.9 m s^{-1}; **(b)** E_K initial and final is 31.7 J therefore elastic; **(c)** no, can show that $\boldsymbol{v}_1 = -3.9$ m s^{-1} and $\boldsymbol{v}_2 = +3.8$ m s^{-1} and balls can't jump over each other **35** 913 m s^{-1} **36** $55 per second; $4.75 million per day **37 (a)** 0.28 km; **(b)** 5.94×10^4 W. **38** 10.7 J **39** 254 N m **40** 1.2 kg **41** 6.0 m s^{-1} (21.7 km h^{-1}) **42** 5 m s^{-1} **44** 31.7 tonnes **46 (a)** 7 m s^{-2}, friction; **(b)** 100 000 N

Chapter 10

2 (a) neither; **(b)** steam **3** Thermal energy **4 (a)** the same; **(b)** 100 mL **5 (a)** 293 K; **(b)** 123 K; **(c)** 793 K; **(d)** 201 K; **(e)** not possible **6 (a)** −223°C; **(b)** 5°C; **(c)** 727°C; **(d)** not possible **7** 1.16×10^5 J **8** 6.3×10^4 J **9** 2.5×10^3 J kg^{-1} K^{-1}; methylated spirits **10** 1.17×10^3 J kg^{-1} K^{-1} **11** 112 s **12** 2.1×10^3 J kg^{-1} K^{-1} **13** 1.6×10^5 J **14** 1.2×10^5 J **15** 8.3×10^5 J **16** 58 min **17** 31°C **21** 2.5°C, 1°C, 0.5°C **23 (a)** 563 K; **(b)** 248 K; **(c)** 332.2 K **24 (a)** −204°C; **(b)** 1103°C; **(c)** 72.6°C **25** 36.1°C **27** 84.3°C **28** 64°C **31** 26 g **32** 1.5×10^6 J kg^{-1} **36 (a)** 80°C; **(b)** 150°C; **(c)** 1.3×10^5 J; **(d)** 1.3×10^6 J; **(e)** 1×10^7 J kg^{-1}; **(f)** 4.3×10^3 J kg^{-1} K^{-1} **39** 29.3°C **40** 1.9 cm **41** 57 g

Chapter 11

1 920 kPa **2** 50 m^3 **3** 50 cm^3 **4** 2.0 kg **5** 35 cm^3 **6** 180 K **7** 8.8 L **8** 76 cm^3 **9** 5028 balloons **10** 330 kPa **11** 39.8 h **12** 9.7×10^5 Pa **13** 2.1×10^{22} molecules **14** 0.012 moles **15 (a)** 1.0×10^7 Pa; **(b)** 1.0×10^7 Pa **16** 4.8×10^2 m s^{-1} **17** 1.2×10^3 m s^{-1} **18 (a)** 1:1; **(b)** 1: $\sqrt{10}$ **19** 1.6×10^5 Pa **20** 0.90 mm for contraction, 0.75 mm for expansion **21 (a)** 1.5×10^{-2} mm; **(b)** 8.5×10^{-2} mm; **(c)** rings **22** 50°C **23** 4.8 cm **24** 1.1 L **25** 1.7 L **26** 497.8 mL **27** 1.0 m^3 **28** 99.5 h **29** 174 cm Hg **30** 5.2 atm **31 (a)** 1/6 of original volume; **(b)** twice the original **32** 5.9×10^{-2} m^3 **33** 8×10^{23} molecules **34** 175.5°C **35** 4.9×10^{22} molecules **36** 4.8×10^{23} molecules **37** 4.0×10^{-17} J **38** 6.1×10^{-21} J **39** 2.1×10^{-22} J **40** 1999 m **41** 5.5 mm **42** 13.4 g cm^{-3} **43** 17.5×10^{-6} m° C^{-1} **45** 8.3×10^2 J **46** 30°C **47 (a)** 52°C; **(b)** 61°C **48 (a)** 20 m; **(b)** $V_2 = 3.2\, V_1$ **49** 1.1×10^5 Pa **50** 9.5 atm **51** °M = °C + 303

Chapter 12

1 (b) steel, iron, brass, aluminium, and copper **3 (a)** 1.6 kW; **(b)** 5.8×10^6 J **4** 1.3×10^5 J **7** iron **11** 85 W **12** 4.1×10^2 W **13** 9.6×10^6 J **22 (a)** the can of Coke **36** 2.7×10^6 J

Answers-Walding p761 recto 25/8/04 10:03 AM Page 765

Chapter 13

1 1.0 m s^{-1} **2** 6×10^{14} Hz **4** 1.5×10^{3} m s^{-1}, 3.3×10^{-3} s **5** B down, G down, P up, D up, and M up **8** **(a)** 20 cm; **(b)** 0.60 s; **(c)** 1.7 Hz **9** **(a)** 80 cm; **(b)** 15 cm **(c)** 1.0 Hz **10** **(a)** 20 cm; **(b)** 40 cm; **(c)** 2.5 $\times 10^{2}$ cm s^{-1} **11** Speed increases **13** **(a)** transverse **14** **(a)** transverse; **(b)** longitudinal **17** **(a)** 2.0 Hz, 0.50 s, 1.0 m s^{-1}; **(b)** **(i)** frequency and wavelength, **(ii)** speed of the wave; **(c)** change the tension in the spring **18** **(a)** second is heavier than the first; **(b)** transmitted in phase **20** **(a)** 20 cm; **(b)** 10 cm **21** 0.30 m s^{-1} **22** 1.1 m **23** 600 m to 10 m **24** **(a)** 0.62 m; **(b)** 550 Hz; **(c)** 2.6 m **25** **(a)** 4 cm; **(b)** 20 cm; **(c)** 0.10 s; **(d)** 2.0 m s^{-1} **26** **(a)** P longitudinal, S transverse, L transverse; **(b)** 300 Hz **27** **(a)** A down, B down; **(b)** A to left, B to right **28** 'A' heavier than 'B', 'C' heavier than 'B' **29** **(a)** between A and E **(b)** A, C, E, G, I, K; **(c)** B, D, F, H, J; L **(d)** 4 m **30** 0.40 s **31** **(a)** λ = 16 cm, f = 0.10 Hz, amplitude = 4 cm, v = 1.6 cm s^{-1}; **(b)** **(i)** H,C, **(ii)** A,E,F, **(iii)** B,D,G; **(c)** 4.0 cm above where it is now **32** **(a)** 4.0 m; **(b)** 40 m s^{-1}; **(c)** A and E, B and F **35** **(b)** at an undisturbed position; **(c)** 8.0 cm **36** **(a)** 4.0 cm; **(b)** 2.5 Hz **(c)** 2.0 cm; **(d)** 10 cm s^{-1} **37** 333 m s^{-1} **38** $\lambda = \dfrac{4l}{2n - 1}$ *where n = 1, 2, 3, . . .* **39** 36 m **41** **(a)** B; **(b)** 1.0 m s^{-1}

Chapter 14

3 0.10 m s^{-1} **4** 0.10 m s^{-1} **5** **(a)** 0.05 s; **(b)** 1.25 cm; **(c)** 4 waves; **(d)** 0.625 cm **9** **(a)** different depths of water; **(b)** 5.0 Hz; **(c)** 5.0 Hz; **(d)** 15 cm s^{-1} **10** 1.9 cm **11** **(a)** Region (i); **(b)** 4:3; **(c)** 4:3; **(d)** 1:1; **(e)** refraction **12** **(a)** the dotted line; **(b)** 2.0 cm; **(c)** **(i)** constructive, **(ii)** destructive, **(iii)** constructive, **(iv)** constructive **15** **(a)** 7.5 cm s^{-1}; **(b)** 0.17 s; **(c)** 1.25 cm; **(d)** 6.0 Hz **16** 5 cm s^{-1}, 4.2 Hz **18** 62.5 cm s^{-1}, 25 Hz **19** **(a)** 1.3 Hz; **(b)** 4.5 cm **21** **(a)** they become circular; **(b)** no change; **(c)** no change; **(d)** no change **23** No change **25** **(a)** path difference = 6.0 λ; **(b)** 2.5 λ; **(c)** $n\,\lambda$ **26** **(a)** the number of nodal lines will decrease; **(b)** increase in the number of nodal lines; **(c)** decrease in the number of nodal lines **28** **(a)** 1:1; **(b)** 6:10 **29** v_d = 80 cm s^{-1}, v_s = 60 cm s^{-1} **31** A **32** C **33** A — nodal line, B — antinodal line, C — nodal line **34** **(a)** 10 Hz; **(b)** 2.0 cm; **(d)** 8; **(e)** 'a' is doubled, 'b' is halved, 'd' is doubled **35** **(a)** destructive; **(b)** destructive; **(c)** the second order antinodal line; **(d)** the fourth order nodal line **36** **(a)** 20 cm s^{-1}; **(b)** 20 cm s^{-1}; **(c)** 2.0 cm; **(d)** destructive interference; **(e)** the first nodal line; **(f)** point X lies on the first antinodal line **37** **(b)** Constructive interference occurs when the path difference = (n – 1/2) λ, and destructive when the path difference = $n\,\lambda$

Chapter 15

1 **(a)** 2.4 cm; **(b)** 6.5 cm; **(c)** 1.6 cm **2** 700 nm **3** **(a)** first-order = 0.53 cm, second-order = 1.06 cm; **(b)** first-order = 0.79 cm, second-order = 1.58 cm; **(e)** blue = 440 nm, red = 660 nm **4** 5.8 cm **5** **(a)** 1.3 cm; **(b)** 2.6 cm; **(c)** 2.6 cm; **(d)** 1.3 cm; **(e)** central maximum is twice as wide as others **6** 0.20 mm **7** **(a)** 22.5 seconds of arc; **(b)** 32.5 seconds of arc; **(c)** Blue light has greater resolving power **8** **(a)** 2.6 m; **(b)** 3.9 m **9** **(a)** 23.6°; **(b)** 36.9°; **(c)** 5 **10** λ = 532 nm, green **11** **(a)** Infra red 1080 nm or Green 540 nm; **(b)** Infra red 720 nm or Blue 432 nm **12** Blue 432 nm **13** 500 nm **14** 7.0×10^{-5} m **15** 1.1×10^{-5} m **18** **(a)** 8.7 mm; **(b)** 5.2 cm **19** 2.0×10^{-6} m **20** **(a)** 1.5 cm; **(b)** 5.1 mm **21** 667 nm **22** **(a)** 0.125 m from P; **(b)** double crest, maximum **23** **(b)** 1.4 cm; **(c)** 11.3 mm **24** Green light (568 nm) **25** 480 nm **26** 141 **27** **(a)** **(i)** 3×10^{8} m s^{-1}, **(ii)** 3×10^{8} m s^{-1}, **(iii)** 3×10^{8} m s^{-1}; **(b)** **(i)** 4.8×10^{14} Hz, **(ii)** 6.4×10^{14} Hz, **(iii)** 3.0×10^{17} Hz; **(c)** **(i)** 4.8×10^{14} Hz, **(ii)** 6.4×10^{14} Hz, **(iii)** 3.0×10^{17} Hz **29** 2.6 m **30** 75 km **32** **(a)** 1.5 cm; **(b)** **(i)** 2.25 cm, **(ii)** 1.5 cm; **(c)** 4.5 cm **33** **(a)** radio waves; **(b)** microwaves; **(c)** infra-red waves; **(d)** X-rays, gamma rays; **(e)** visible light; **(f)** X-rays; **(g)** X-rays; **(h)** microwaves; **(i)** ultraviolet light **34** FM radio **36** **(a)** **(i)** 0 λ, **(ii)** 1/2 λ **(iii)** 2 λ; **(b)** **(i)** destructive, **(ii)** constructive; **(c)** 440 nm **39** 0.80° **40** 338 nm **41** 674 nm **42** Green (580 nm) Indigo-Violet (414 nm) **43** 4.0×10^{-5} m **45** all false

Chapter 16

4 1 Hz $\rightarrow$ 7 Hz **5** 8.8×10^{-3} s, 0.34 m **6** Pipe by 0.021 s **7** 336 m s^{-1} **8** 170 m **9** 255 m from you **10** Minimums 1 m apart **11** 333 m s^{-1} **13** **(b)** fish are 725 m below **14** 103 kHz **16** **(b)** 297 Hz **17** f_o = 340 Hz, third harmonic = 1020 Hz **18** f_o = 1133 Hz, 2 f_o = 2267 Hz, 3 f_o = 3399 Hz, 4 f_o = 4532 Hz **19** **(a)** 2550 Hz; **(b)** 5100 Hz; **(c)** 850 Hz **21** **(a)** 3.0 Hz; **(b)** 4.0 Hz; **(c)** 6.0 Hz **22** 50 dB **23** 47 dB **24** 3.16×10^{-4} W m^{-2} **26** **(a)** 1264 Hz; **(b)** 1142 Hz **27** **(a)** 1109 Hz; **(b)** 911 Hz; **(c)** same, 1000 Hz **28** In the direction of motion **29** **(a)** 2.4×10^{3} Hz; **(b)** 336 m s^{-1} **30** 3.4 km **32** 0.22 m **33** 840 m **34** 2550 Hz **35** 425 Hz **36** **(a)** 340 Hz; **(b)** 0.50 m **37** **(a)** 1.16 m **(b)** second overtone = 884 Hz, third overtone = 1179 Hz **38** 343 m s^{-1} **39** **(b)** 336 m s^{-1} **40** **(d)** 200 Hz, 350 Hz, 400 Hz **41** **(a)** 1063 Hz; **(b)** **(i)** 16 cm from the top, **(ii)** 4 cm from the top, **(iii)** 2.7 cm from the top **43** 439 Hz **44** 248 Hz **45** 437 Hz **46** 484 Hz **47** 8 km h^{-1} **48** 990 m away perpendicular to line of microphones; and 24 m from microphone 1 towards microphone 2

Chapter 17

3 **(a)** 50°; **(b)** 70°; **(c)** 65° **4** 4 m s^{-1} **7** **(a)** Convex mirror; **(b)** concave mirror **8** 23 cm in front of the mirror **9** **(a)** 3.3 cm behind the mirror; **(b)** virtual, upright, smaller; **(c)** 1.32 cm **10** **(a)** 60 cm in front of the mirror; **(b)** 3:1; **(c)** 3.0 cm; **(d)** real, inverted, magnified **11 (b)** 11 cm behind the mirror; **(c)** 0.88 cm **12 (b)** 2.4 cm behind the mirror; **(c)** 1.2 cm high **13 (a)** u = 30 cm, v = 15 cm; **(b)** u = 15 cm, v = 30 cm **16** 4.0 m **17** Minimum length = 80 cm **19 (a)** diverging; **(b)** diverging; **(c)** parallel; **(d)** converging; **(e)** Parallel; **(f)** Diverging; **(g)** parallel; **(h)** none reflected **20 (b)** real, inverted, magnified; **(c)** 23 cm in front of the mirror; **(d)** 2.3 cm **21** 4 cm **22 (b)** virtual image; **(c)** 8.6 cm behind the mirror **24 (a)** 13.3 cm; **(b)** 10 cm; **(c)** real; **(d)** inverted **30** 60° **31 (a)** concave; **(b)** at the focus; **(c)** use a parabolic dish **35** A, B, D **36 (a)** concave dish; **(b)** at the focus **37 (c)** 2; **(d)** 22.5 cm; **(e)** 15 cm **38** 15 cm **39** 4 cm in front of the convex mirror **40 (a)** u = 10 cm, v = 20 cm behind the mirror; **(b)** u = 30 cm, v = 60 cm in front of the mirror

Chapter 18

2 (a) 1.15; **(b)** 1.67; **(c)** impossible **3 (a)** 2.0×10^8 m s^{-1}; **(b)** 1.25×10^8 m s^{-1}; **(c)** 2.3×10^8 m s^{-1} **4** 1.81 **5 (a)** 1.46 **(b)** 28.7° **(c)** 64.1° **(d)** undefined **6** 1.22 **7 (b)** sin i/sin r = 1.52; **(c)** 1.52; **(d)** 32° **8** 1.14 **9 (a)** 26.5°; **(b)** 24.2°; **(c)** 38°C; **(d)** 1.09 **10 (a)** 1.41; **(b)** 1.14; **(c)** 1.25; **(d)** 1.37 **11** θ_w = 28°, θ_g = 23° **12** (b), (d) **13** True depth = 0.9975 m **14 (a)** no; **(b)** yes; **(c)** no; **(d)** yes; **(e)** yes; **(f)** yes **15** 58.8° **16** 1.2 **18 (a)** 60.5°; **(b)** 1.76×10^8 m s^{-1} **19** 46.4° **20 (a)** towards; **(b)** away; **(c)** towards; **(d)** away; **(e)** away; **(f)** away **21 (a)** 33.6°; **(b)** 2.05×10^8 m s^{-1} **23** 2.6×10^8 m s^{-1} **24 (a)** 1.25; **(b)** 1.67; **(c)** 1.1; **(d)** impossible **26 (a) (i)** 1.69×10^8 m s^{-1}, **(ii)** 2.5×10^8 m s^{-1}, **(iii)** 1.43×10^8 m s^{-1}, **(iv)** 2.11×10^8 m s^{-1}; **(b) (iii)** Highest refractive index **27 (a)** 1.1; **(b)** 43.3° **28** 422 nm **29 (a)** 66.5 cm **30** 88.9° **31** 86.6 m^2 **32** 1.52 **33** 0.67° **34 (a)** 1.19; **(b)** 1.09; **(c)** 1.41; **(d)** impossible **37** 29.6° **39** 1.43 cm **40** Ray passes into prism unrefracted, then retracts at 30.5° into water from upper surface of prism and strikes water surface at 14.5°, refracting into air at 19.5° **41** R, O refract out of right side at different angles; Y, G, B, I, V undergo total internal reflection and head as a single beam towards bottom **42** A spectrum diagonally across page: violet (top left) to red (bottom right) **42 (a)** Colour, don't use plastic, different R.I, degradable plastic

Chapter 19

4 Real inverted image 60 cm from the lens **5** The image is 7.1 m on the object side of the lens **6 (a)** Virtual, upright magnified **(b)** v = –6.67 cm; **(c)** H_i = 4 cm **7 (a)** v = –7.5 cm; **(b)** H_i = 1.25 cm; **(c)** virtual, upright, diminished **8 (a)** virtual, upright, larger, 20 cm on the same side as the object; **(b)** virtual, upright, smaller 4.5 cm on the same side as the object; **(c)** real, inverted, same size, 50 cm on the other side **9 (a)** HM = 2; **(b)** u = 15 cm **10 (a)** M = 0.25; **(b)** u = 40 cm; **(c)** f = 8.0 cm; **(d)** real, inverted, smaller **11** f = 103.7 cm **12 (a)** –4D; **(b)** –0.5D; **(c)** 5D; **(d)** 1D **13 (a)** concave lens f = 20 cm; **(b)** convex lens f = 10 cm; **(c)** convex lens f = 4.0 cm; **(d)** concave lens f = 2.0 cm **14 (a)** convex; **(b)** convex; **(c)** concave; **(d)** concave; **(e)** convex **15 (a) (ii)** 20 cm; **(b) (ii)** 2.25 cm **16 (a)** just inside the focal length; **(b)** H_i = 4 mm **17 (a)** v = 48 cm on the opposite side; **(b)** H_i = 3.7 cm; **(c)** real, inverted, smaller; **(d)** 16.7 cm on the same side as the object **18 (a)** v = –10 cm; **(b)** m = 0.50; **(c)** 5.0 cm **19 (a)** convex; **(b)** virtual, upright, larger; **(c)** –33.3 cm; **(d)** 3.3 mm **20 (a)** 50 mm; **(b)** 56 mm **22** Ray 2, 4 **24 (b)** 25 cm; **(c)** 38 cm; **25 (a)** convex; **(b)** inside the focal length **26** 5.4 cm to the right of the concave lens **27 (a)** two images **(b)** 64.3 cm to the right of the lens, 23.7 cm to the right of the lens **28** f = 20 cm **29** f = 35.5 cm **30** f = 43.8 cm **31 (a)** u = 16 cm from lens; **(b)** u = 24 cm **32** f = 20 cm **33** f = 40 cm

Chapter 20

5 (a) closer **(b)** f_{11} and larger **6 (b)** 2.5 cm **(c)** F4 **7 (a)** 12 cm; **(b)** 17.1 cm; **(c)** cannot be done **8 (a)** 40, 410 cm; **(b)** 2.0 m, 205 cm; **(c)** 5 cm, 305 cm; **(d)** 1.98 m, 99 **10 (a)** 1.25; **(b)** 12.5; **(c)** 0.25; **(d)** 5.0 **16 (a)** cornea and lens; **(b)** ciliary muscles; **(c)** iris; **(d)** rods **20** 40 cm **21 (a)** 1.0 cm; **(b)** 0.50 cm; **(c)** 2.5 cm **23 (a)** convex; **(b)** concave; **(c)** convex; **(d)** convex; **(e)** convex **26 (a)** minimum distance = 30 cm; **(b)** maximum distance = 2.55 m; **(c)** 1.95 m^2 **27** c, d, f **31** 6.67 cm from the lens **33** 50 cm

Chapter 21

1 (a) glass positive, silk negative; **(b)** rubber negative, wool positive; **(c)** gold negative, cat fur positive; **(c)** **4** 4.8×10^{-19} C, the atom has lost three electrons **5** each +2 μC **6** A = +4 μC, B = +1 μC **7** 2.3 N repulsive **8 (a)** repulsive; **(b)** 1.5×10^{-4} N; **(c)** 2.4×10^{-3} N; **(d)** 6.0×10^{-4} N **9** 4.1×10^{-1} N up the page **11** 1.35×10^6 N C^{-1} radially inwards **12** 4.1×10^5 V **13 (a)** 24 V; **(b)** 1.3×10^{-4} J **14 (a)** 1.0×10^4 V m^{-1}; **(b)** 6.0×10^{-2} N upwards; **(c)** 1.2×10^{-3} J **16** Perspex positive, silk negative **17** Excess electrons conducted through the body to earth **18** 3.1×10^{-6} C positive **19** 5.0×10^7 N C^{-1} **20** 1.4×10^{-17} C **21** 7×10^{-9} C **22** 1.8×10^6 V **23** 4.6×10^4 V **24 (a)** 1.2×10^5 V m^{-1}; **(b)** 9.6×10^{-13} N upwards **25** 3000 eV, 4.8×10^{-16} J, 3.2×10^7 m s^{-1} **26 (a)** 1.6×10^{-17} N down page; **(b)** 1.67×10^8 m s^{-2}; **(c)** 8.3×10^{-6} s; **(d)** 5.7×10^{-3} m; **(e)** towards the bottom right **27 (a)** point A; **(b)** zero; **(c)** 200 V; **(d)** 9.0×10^{-7} J; **(e)** 1.8×10^{-6} J **32 (a)** 2.1×10^{-9} s; **(b)** 9.7×10^{-4} m upwards; **(c)** 1.06×10^7 m s^{-1}, angle = 5°

Chapter 22

2 5.3 V **3** 3.1×10^{-4} m s^{-1} **4** 1.5 V **5 (a)** 0.6 A; **(b)** 2.16×10^{3} C **6** 30 mV, it will not be electrocuted **8 (a)** 1.08 Ω; **(b)** 0.72 Ω; **(c)** assuming that conductor B is copper, R = 0.8 Ω **10** Voltmeter 11.8 V, ammeter 0.39 A **11** 0.96 A, current through each 0.48 A **12** Current through R_5 = 1 A, voltage across R_5 is 5 V voltage drop across resistors 1–4 is 5 V and current through each is 0.5 A **13** 144 W; **14 (a)** 1152 Ω; **(b)** 58 Ω; **(c)** 580 Ω **15** 2.2×10^{6} J **16 (a)** 8.9 Ω; **(b)** energy is 18.9 kWh **17 (a)** 4 A; **(b)** 0.25 A; **(c)** 1.1 A **18** 2.2×10^{2} J **19** 8.0×10^{-7} Ω m **21** series **22** The 60 W bulb has a comparatively lower resistance and thus a thicker diameter filament, assuming an equivalent length to the 25 W bulb **23 (a)** R_{tot} = 10 Ω, A = 1.2 A, V = 7.2 V; **(b)** R_{tot} = 15 Ω, A_1 = 1.33 A, A_2 = 0.66 A, V_1 = 6.6 V; **(c)** R_{tot} = 20 Ω, A_1 = 0.9 A, V = 9 V **25** Z = 4 A, Y = 8 A, X = 16 A, V = 64 V, EMF = 160 V **26 (a)** 4 A; **(b)** 60 Ω **27 (a)** 20 bulbs in series; **(b)** 20 W; **(c)** 12 V; **(d)** 83 mA, R = 145 Ω **30 (a)** A is the ohmic resistor; **(b)** 14.5 V; **(c)** 15 V; **(d)** the same **35** 5.1 hours **36** 1.25 Ω **37 (a)** zero potential difference between the opposite ends of the bridge; **(c)** R_x = 1.0 kΩ **38 (a)** 92 Ω; **(b)** 45 V across the 60 Ω resistor **39** R = 232 Ω or 43 Ω

Chapter 23

2 Adjust vertical amplifier to a smaller value in volts/division and decrease the timebase period **3** V_{PP} = 44 V **4** P = 30 W, Yellow–purple–brown **6** Plastic film or greencaps, electrolytics, ceramic capacitors **7** τ = 12 μs **9** time = 12.5 ms; I = 2.4 A **11** V_P = 10.6 V **13 (a)** 20 V; **(b)** 40 V; **(c)** 0.0 V; **(d)** 14.1 V; **(e)** 25 Hz **14 (a)** 5.6 A; **(b)** 3.96 A; **(c)** 60 Hz **16** 75 V, I = 13.3 mA **17** V_1 = 4.1 V, V_2 = 5.9 V, I = 72 mA **20** τ = 47 s, full charge after 141 s, $W = 1.9 \times 10^{-2}$ J **22** 3.2 W **23** V_o = 6.1 V, V_{av} = 3.9 V **24** Decreasing the load resistance increases the current drawn from the supply, with an increase in ripple voltage or hum **29 (a)** $\tau = 2 \times 10^{-2}$ s; **(b)** $\tau = 1 \times 10^{-2}$ s **30 (a)** $I = I_L + I_Z$; **(b)** $V_{in} = V_R + V_Z$; **(c)** $R = \dfrac{V_{in} - V_Z}{I_L + I_Z}$; **(d)** P = 57 mW; **(e)** I_L = 4 mA **(f)** R = 470 Ω preferred **31 (a)** 1.4×10^{-9} C and 5.6×10^{-9} C; **(b)** 2.8×10^{2} V; **(c)** 2.4×10^{-8} J

Chapter 24

3 I_C = 3 mA, I_E = 3.015 mA **6** I_B = 25 μA, V_C = 1.4 V **7** β = 100 **9** Z_{in} = 10 kΩ **11** Square wave **13** Binary levels are two state; ON = high (5 V), OFF = low (0 V) **15** f_0 = 200 Hz **16** phototransistor — input transducer, loudspeaker — output transducer **20** β = 86, or β = 116 **22** A_V = 94 **25 (a)** I_C = 100 mA; **(b)** R_C = 30 Ω; **(c)** R_B = 5.2 kΩ nearest preferred **26 (a)** R_B = 280 Ω; **(b)** 0.4 mW in R_B, 0.13 W in R_L; **(c)** 5.4 V; **(d)** 0.7 V **27 (b)** I_B = 7.15 μA, I_C = 0.86 mA, V_{CE} = 1.4 V

Chapter 25

1 A substance that can be magnetised. Yes it does! **2** A magnetic pole is permanent, an induced magnetic pole is temporary **3** An experiment based on repulsion **4** The way in which they are influenced by magnets **5** Development of new magnetic materials with greater strength **7** Magnetic flux is the total number of field lines passing through an area, whereas magnetic flux density is the magnetic flux per unit area **8** 2.2 T **10** Declination is the angle between the Earth's magnetic axis and it's geographical axis. Inclination or dip is the angle of the field lines to the horizontal **11 (a)** 7.3×10^{-6} T; **(b)** 3.3×10^{-5} T, direction for both depends on position around the wire being considered. Use the screw rule **12** 0.75 T **13** 75 A current from A to B **14** 4×10^{-5} T up, 1.6×10^{-4} T down **15** 1.4×10^{-4} N to the right **16** 0.05 N up the page **17 (a)** down the page; **(b)** to the left **18 (a)** side AB downwards, side DC upwards; **(b)** to allow coil to freely rotate; **(c)** 0.18 N **19 (a)** 3.15×10^{-14} N; **(b)** acceleration = 5.6×10^{12} m s^{-2} **20 (a)** X and Y positive, Z negative; **(b)** particle Y has the greatest mass as it's radius of curvature is greatest **23** 1.1×10^{-5} Wb **24 (a)** A: N, B: S; **(b)** A; N, B: S; **(c)** both A and B not magnetised **25** Force is attractive, 1.2×10^{-4} N m^{-1} **26** 2.0×10^{-2} N downwards **27** 7.5 A, towards the east **28** 1.3 mT into page **31** Field strength is too small to affect the watch **33 (a)** Magnetic field directed into the page; **(b)** 3.2×10^{-17} J; **(c)** $m = \dfrac{qB^2r^2}{2V}$; **(d)** 1.0×10^{-30} kg **34** Current direction is Y to X, 0.98 A **35 (a)** 67 A; **(b)** 5.4 N m **36 (a)** ${}^{4}_{2}He^{2+}$: b, ${}^{4}_{2}He^{+}$: d, ${}^{3}_{2}He^{2+}$: a, ${}^{3}_{2}He^{+}$: c; **(b)** in order left to right, ${}^{200}_{80}Hg^{2+}$, ${}^{204}_{80}Hg^{2+}$, ${}^{200}_{80}Hg^{+}$, ${}^{204}_{80}Hg^{+}$, ${}^{400}_{160}Hg_2^{+}$, ${}^{404}_{160}Hg_2^{+}$, ${}^{408}_{160}Hg_2^{+}$

Chapter 26

2 (a) 0.24 V; **(b)** 0.2 V **3 (a)** $V_{AB} = V_{DC}$ = 0 V, $V_{AD} = V_{BC}$ = 0.06 V; **(b)** no current will flow as V_{AD} opposes V_{BC} **6** A very large output AC voltage could be induced even for a small input voltage **8** 1600 μV **9** Step-up, turns ratio is 52 **10** 140 V peak **11** Losses due to lower voltage distribution **12 (a)** appliances used to heat water and cook breakfast meals; **(b)** use of artificial heating in homes; **(c)** appliances used to cook evening meals as well as general heating **13** At 11 kV, power loss is 3.6×10^{5} W, at 66 kV, 9.9×10^{3} W only **14** 2.4 V **15** 90 V **16** 267 V **17** Because without it, the law of conservation of energy would be violated **19** 53 turns **20** Because a changing magnetic flux is needed for induction **23** The magnet will, at first, be attracted downwards into the right solenoid and then pushed upwards again **24** 240 V_{RMS} @ 50 Hz **26** Factors: input AC voltage, turns ratio, laminations;

energy appears as heat **27** 40 mA **28 (b)** 12 mV **29** Induced current will flow from right to left through the ammeter and then the solenoid **30 (a)** 5.1 kW; **(b)** 218 V at hospital, hence could not use satisfactorily; **(c)** voltage would drop considerably; **(d)** voltage loss about 1.0 V only **31** 10 Hz.

Chapter 27

2 (a) 8080 V m^{-1}; **(b)** 1.274×10^{-18} C; **(c)** 8 e^- **5** electrometer **6** filament **9** False–neutron not discovered until 1930. **12** ${}^{12}_{6}\mathrm{C}$, ${}^{14}_{6}\mathrm{C}$ **13** ${}^{3}_{2}\mathrm{He}$ (1n, 2p, 2e); ${}^{4}_{2}\mathrm{He}$ (2n, 2p, 2e) **14** Neutrons are neutral and aren't repelled by electron cloud or positive nucleus **15** protons, neutrons: **(a)** 1,1; **(b)** 6,6; **(c)** 8,9; **(d)** 11,12; **(e)** 16,16; **(f)** 47,60; **(g)** 53,74; **(h)** 92,146 **16** 28.103 **17 (a)** 8.48 MeV, 2.83 MeV/nucleon; **(b)** 7.71 MeV, 2.57 MeV/nucleon; **(c)** 104.4 MeV, 7.46 MeV; **(d)** 127.6 MeV, 7.97 MeV **18** 341.324 MeV **19 (a)** 5.5×10^{3} V m^{-1}; **(b)** 1.92×10^{-18} C, 12 e^- **20 (a)** 82, 125; **(b)** 17, 18; **(c)** 7, 8; **(d)** 85, 130; **(e)** 83, 133 **22 (a)** ${}^{40}_{20}\mathrm{Ca}$, ${}^{42}_{20}\mathrm{Ca}$, ${}^{43}_{20}\mathrm{Ca}$, ${}^{45}_{20}\mathrm{Ca}$; **(b) (i)** 20, **(ii)** 40; **(i)** 22, **(ii)** 42; **(i)** 23, **(ii)** 43; **(i)** 25, **(ii)** 45 **23 (a) (i)** 0.320 802 u, **(ii)** 1.915 055 u, **(iii)** 1.034 425 u, **(iv)** 0.042 131 u; **(b)** 8.5 MeV/nucleon, 7.58 MeV/nucleon, 8.5 MeV/nucleon, 5.6 MeV/nucleon **25 (a) (i)** to produce e^-, **(ii)** anode is +ve, **(b) (i)** more e^- produced, **(ii)** faster acceleration **26** 2.44×10^{-15} kg **27** 22 **28** D **29** F

Chapter 28

1 The air is ionised and the electroscope attracts opposite charge from the air **3 (a)** 88, 138; **(b)** 1, 0; **(c)** 93, 146 **4 (a)** radium; **(b)** hydrogen; **(c)** Bk; **(d)** Sr **5 (a)** ${}^{214}_{84}\mathrm{Po}$; **(b)** ${}^{239}_{92}\mathrm{U}$; **(c)** ${}^{4}_{2}\mathrm{He}$; **(d)** ${}^{45}_{21}\mathrm{Sc}$; **(e)** ${}^{58}_{28}\mathrm{Ni}$; **(f)** ${}^{230}_{92}\mathrm{U}$ **6 (a)** ${}^{32}_{16}\mathrm{S}$; **(b)** ${}^{0}_{-1}\mathrm{e}$; **(c)** ${}^{22}_{10}\mathrm{Ne}$; **(d)** ${}^{0}_{-1}\mathrm{e}$; **(e)** ${}^{1}_{0}\mathrm{n}$ **7 (a)** ${}^{210}_{83}\mathrm{Bi}$; **(b)** ${}^{214}_{83}\mathrm{Bi}$; **(c)** ${}^{222}_{86}\mathrm{Rn}$; **(d)** ${}^{210}_{84}\mathrm{Po}$ **8 (a)** ${}^{14}_{6}\mathrm{C} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{14}_{7}\mathrm{N}$; **(b)** ${}^{24}_{11}\mathrm{Na} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{24}_{12}\mathrm{Mg}$ **(c)** ${}^{32}_{15}\mathrm{P} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{32}_{16}\mathrm{S}$ **9 (a)** ${}^{23}_{11}\mathrm{Na} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{22}_{10}\mathrm{Ne}$; **(b)** ${}^{18}_{9}\mathrm{F} \rightarrow {}^{0}_{+1}\mathrm{e} + {}^{18}_{8}\mathrm{O}$; **(c)** ${}^{19}_{10}\mathrm{Ne} \rightarrow {}^{0}_{+1}\mathrm{e} + {}^{19}_{9}\mathrm{F}$ **(d)** ${}^{199}_{82}\mathrm{Pb} \rightarrow {}^{0}_{+1}\mathrm{e} + {}^{199}_{81}\mathrm{Tl}$ **10 (a)** 0.02235 s^{-1}, 1.34 min^{-1}; **(b) (i)** 5 g, **(ii)** 1.25g, **(iii)** 3×10^{-5} g **11 (a)** 89.7 min **(b)** 34.46 min **12 (a)** 0.0693 min^{-1} (or 0.00115 s^{-1}) **(b)** 8×10^{13} Bq **(c) (i)** 1.245×10^{12} Bq **(ii)** 1165 Bq **(d)** 7.7 hours **13** 2.62 minutes **14 (b)** 4.38×10^{-3} min^{-1}; **(c)** 158 minutes **15** 11 400 y **16** 686 years old, hence 1302 AD; previously believed to be genuine and hence should have a date of about 32 AD when Jesus died **17 (a)** 2.03×10^{-11} J; **(b)** 5.21×10^{13} J/kg **18 (a)** 2.82×10^{-12} J; **(b)** 5.24×10^{-13} J; **(c)** 3.02×10^{-12} J. Most energy comes from reaction (c) **24 (a)** U; **(b)** H; **(c)** Ra; **(d)** P **25 (a)** ${}^{24}_{12}\mathrm{Mg}$; **(b)** ${}^{22}_{10}\mathrm{Ne}$; **(c)** ${}^{206}_{82}\mathrm{Pb}$; **(d)** ${}^{32}_{16}\mathrm{S}$ **26 (a)** ${}^{0}_{-1}\mathrm{e}$; **(b)** ${}^{222}_{86}\mathrm{Rn}$ **27 (a)** ${}^{b}_{a}\mathrm{X} \rightarrow {}^{4}_{2}\mathrm{He} + {}^{b-4}_{a-2}\mathrm{Y}$ **(b)** ${}^{b-4}_{a-2}\mathrm{Y} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{b-4}_{a-1}\mathrm{Z}$ **28 (a)** ${}^{1}_{0}\mathrm{n}$; **(b)** ${}^{24}_{11}\mathrm{Na}$; **(c)** ${}^{25}_{12}\mathrm{Mg}$; **(d)** ${}^{39}_{19}\mathrm{K}$; **(e)** ${}^{27}_{13}\mathrm{Al}$; **(f)** ${}^{9}_{4}\mathrm{Be}$ **29** ${}^{96}_{38}\mathrm{Sr} \rightarrow 4{}^{0}_{-1}\mathrm{e} + {}^{96}_{42}\mathrm{Mo}$ **30 (a)** ${}^{14}_{6}\mathrm{C} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{14}_{7}\mathrm{N}$; **(b)** 0.000 381 u = 6.32×10^{-31} kg; **(c)** 5.69×10^{-14} J **31 (a)** ${}^{116}_{49}\mathrm{In} \rightarrow {}^{0}_{-1}\mathrm{e} + {}^{116}_{50}\mathrm{Sn}$; **(b)** 5.297×10^{-30} kg; **(c)** 4.767×10^{-13} J **32** No, need something that exchanges carbon dioxide **33** 34 670 y **34** 13.8 million years **35 (a)** 353.5 GBq; **(b)** 9.95 half-lives, **36 (a)** 148 days; **(b)** 285 days **37 (a)** 0.0859 $days^{-1}$; **(b)** 5.0×10^{16} atoms; **(c)** 126 days **38** emit as β **40** 1.00 min **42 (a)** 49.6 y; **(b)** 15 **43 (a)** 4.123×10^{-12} J; **(b)** 6.16×10^{14} J **44 (i)** 8.145×10^{13} J/kg; **(ii)** 5.72×10^{14} J/kg, reaction **(ii)** produces more energy per kilogram; **(b)** reaction i is fission, ii is fusion **45** $X_1 = {}^{13}_{7}\mathrm{N}$; $X_2 = {}^{13}_{6}\mathrm{C}$; $X_3 = {}^{14}_{7}\mathrm{N}$; $X_4 = {}^{15}_{8}\mathrm{O}$; $X_5 = {}^{15}_{7}\mathrm{N}$; $X_6 = {}^{12}_{6}\mathrm{C}$; **(b)** sum: $4{}^{1}_{1}\mathrm{p} \rightarrow 2{}^{0}_{+1}\mathrm{e} + {}^{4}_{2}\mathrm{He}$ **46 (a)** 4.3333×10^{6} kg/s; **(b)** 7.3×10^{15} y **47** 8.97×10^{13} J; 1.8 million times greater **48 (a)** 172 800 kg; **(b)** 4.84×10^{14} kg **49** Half life = 2.5 days **50** Make alloy of Au-198 with Al and make a saucepan. Use this for cooking and scan brain of patient **51** Number of particles in 10 s = 3.7×10^{14}; energy = 74 J; dose = 0.74 J; absorbed dose = 0.011 Gy **52** ${}^{238}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{239}_{92}\mathrm{U}$ ${}^{239}_{92}\mathrm{U} \rightarrow {}^{239}_{93}\mathrm{Np} + {}^{0}_{-1}\mathrm{e}$ ${}^{239}_{93}\mathrm{U} \rightarrow {}^{239}_{94}\mathrm{Pu} + {}^{0}_{-1}\mathrm{e}$.

Chapter 29

1 Electromagnetic force acts on electrons, Strong and weak nuclear forces act on nucleons, whereas the gravitational force acts on all matter. Gravity has the biggest range and the weak nuclear force has the smallest range **2** E = 2.85×10^{-19} J, λ = 697 nm (red) **3** Violet photons have greater energy and shorter wavelengths than red photons **4** $W = 3.3 \times 10^{-19}$ J, $v = 6.7 \times 10^{5}$ m s^{-1} **5** 253 nm ultraviolet photons **8** Lyman series of ultraviolet photons: 122 nm, 103 nm, 98 nm **10** 760 nm **11** 9.7×10^{-10} m **12** An electron's position around the nucleus can only be stated with a certain probability **13** Under conditions applying to the sub-atomic domain **17** 3.62×10^{-19} J = 2.3 eV **18 (a)** 500 nm (green); **(b)** 3.98×10^{-19} J; **(c)** 1.66×10^{-19} J; **(d)** 2.32×10^{-19} J; **(e)** 7.1×10^{5} m s^{-1} **19** 3.0×10^{-12} m, order of magnitude equivalent **20 (a)** 2.03×10^{-18} J; **(b)** 2.17×10^{-18} J; **(c)** 3.14×10^{15} Hz **21** 1.64×10^{-13} J **22 (a)** leptons, β decay; **(b)** nucleons, hadrons; **(c)** all matter **28** 1.69×10^{-27} kg **29** x = 0.6 eV, y = 0.8 eV **30** 4.5×10^{-11} m **31 (a)** 8.16×10^{-19} J; **(b)** 3.36×10^{-19} J and 4.96×10^{-19} J; **(c)** 6.2×10^{-20} J

Chapter 30

1 0.54c **2** 7.8×10^{-9} s **3 (a)** 0.06c; **(b)** 2.85×10^{8} m s^{-1}; **(c)** 2.8×10^{14} km; **(d)** 317 ly **4** still 40.0 m **5** 0.97c **6 (a)** 5.44 y; **(b)** 3.26 y; **(c)** 2.61 ly **7 (a)** 0.40c; **(b)** 0.909c **8** 0.54c **9** 0.96c **10** 4.9×10^{-28} kg **11** 0.0003 kg **12** 7.1×10^{-16} J **13** 3.3×10^{-13} J **14** on the car **15** 1.0 c **16** 0.87c **17 (a)** 0.867c; **(b)** 1.1×10^{-7} s **18** Not perceptibly **19** 181.4 m **20** 0.866c **21** 14.4 ly **22 (a)** 0.447c; **(b)** 0.966c **23** c **24** 0.866c **25 (a)** 0.999 999 218c; **(b)** 1.875 m **26 (a)** 9.0×10^{7} J; **(b)** 9.0×10^{5} kg **27** To you nothing would change. To observers on Earth, mass would increase and any length in the direction of motion would decrease **28 (a)**, **(b)**, **(c)**, **(g)** dependent; **(d)**, **(e)**, **(f)** independent **29** yes **30** 0.0153 m **31 (a)** 8.92 m long, no change to height; **(b)** 13 s; **(c)** v is the same (0.760c) **32** 0.8c **33 (a)** 9.336 ly; **(b)** 21.0 years **34 (a)** 5.9×10^{19}J; **(b)** 4.7% **35** for example: 70.7 kg at 0.9999 c **36** his car 5.90 m; your car 6.25 m **37 (a)** 5.28 y; **(b)** 3.168 y; **(c)** 2.535 ly 39 **(a)** 0.51 MeV **(b)** 1.64×10^{-13}J **(c)** 0.94c

Chapter 31

1 (a) $1.6 \times 10^3\ \Omega$; **(b)** $3.1 \times 10^3\ \Omega$; **(c)** 3.2 mA; **(d)** 31^0 **2 (a)** 356 Hz; **(b)** 0.33 A; **(c)** V_R = 50 V, V_C = +74 V, V_L = –74 V; **(d)** 16.5 W; **(e)** 1.5 **3** 560 Ω nearest, input voltage = 12 V **4** voltage divider bias **5** R_E = 100 Ω, R_C = 360 Ω, R_1 = 22 kΩ, R_2 = 3.9 kΩ, C_1 = 1 μF, C_2 = 47 μF **8 (a)** OR; **(b)** AND; **(c)** NOR; **(d)** 3 input AND; **(e)** NOT **11 (a)** 660 Ω; **(b)** 810 Ω; **(c)** 7.4 mA; **(d)** 55° **13** R_E = 130 Ω, R_C = 560 Ω, R_1 = 33 kΩ, R_2 = 5.6 kΩ, C_1 = 10 uF, C_2 = 47 uF A_V = 50 **15 (b)** I_P = 12 mA **17** Component listing is LDR, LED, single pole switch, variable resistor, inverter gate, OR gate, resistor, connecting wires **18 (a)** Z_{AB} = 42.1 Ω; **(b)** Z = 53 Ω; **(c)** I_{RMS} = 0.1 A; **(d)** V_{AB} = 4.2 V

Chapter 32

1 1.85×10^6 m^2 **6** 5×10^5 J, at 50° reduces to 3.8×10^5 J **7** 330 W m^{-2}. The northern hemisphere is in summer during August **8** Photothermal devices convert solar energy to heat (solar heater) whereas a photovoltaic device converts solar energy directly to electricity (solar cell) **11** Photovoltaic action generates its own EMF and does not require a vacuum environment as does the photoelectric effect **12** 1.1 kW. Actual power output depends primarily on the solar radiation flux change with time, as controlled by such variables as cloud cover **15 (a)** heat losses; **(b)** to obtain maximum power to weight ratio; **(c)** ratio of energy output to energy input is 0.95; **(d)** 18.5 A **18** Open circuit voltage (no load) is 20 V DC, whereas at maximum power transfer the output voltage is 12 V DC **19 (a)** Combined reading for Friday is incorrect **20 (a)** 228 MJ; **(b)** 148 MJ; **(c)** 133 MJ **21 (a)** 34.3 MJ; **(b)** 25°C **22** Bracket length is 250 mm

Chapter 33

2 $\lambda = 5.2 \times 10^{-12}$ m **3** Magnetic deflecting coils **5** Roentgen, Ruska, Rohrer, Donald, Hounsfield and Damadian **6** Diagnostic ultrasound; fatty tissue images dark grey while lymphoma images light grey **8** Radioisotopes: ^{99m}Tc, gamma emitter, reactor; ^{201}Tl, X and gamma emitter, cyclotron; ^{18}F, gamma emitter, cyclotron **9** 1.8 mSv, subjected to gamma radiation that is about the same as the yearly background dose **10** It is probably a metallic object **12** MRI depends on RF radiation emitted by hydrogen **15** Highest dosage — lumbar spine film, lowest dosage — chest AP film **17** Alpha particle's mass is considerably higher, hence more damage ability

Glossary

Absolute sound intensity (I), the energy carried by the waves per second through an area of one square metre, as opposed to a relative decibel rating (dB).

Absorbed dose, radiation that deposits one joule of energy per kilogram of tissue. It has the units of one gray (Gy).

Acceleration, the rate at which the velocity changes with time.

Accuracy, a measure of how close a measurement is to an accepted value. The terms absolute error and relative error are used.

Active device, any semiconductor component that can change the form of electrical signals it receives — such as diodes, transistors and integrated circuits.

Activity, the number of radioactive disintegrations per second within a radionuclide.

Alternating current (AC), an electric current that reverses direction of flow of charge during its cycles.

Ammeter, an instrument for measuring electric current.

Amplitude, the maximum displacement of an oscillating system from equilibrium.

Analog, an electrical signal whose magnitude is continuously variable.

Angle of incidence, the angle between the incident ray and the normal to the surface.

Angle of refraction, the angle between the refracted ray and the normal to the interface.

Angular velocity, the velocity of an object travelling in a circle and expressed in degrees (or radians) per unit of time.

Antinodes, result from the intersection of two crests or two troughs producing super crests and troughs.

Archimedes' principle, when an object is wholly or partially immersed in a fluid, the upthrust on the object is equal to the weight of the fluid displaced.

Astigmatism, occurs if a person sees objects or parts of an object blurred in a particular direction while other parts are in focus.

Atomic mass, the number of protons and neutrons in the nucleus of an atom. Sometimes called mass number.

Atomic number, the number of protons in the atomic nucleus of an atom.

Avogadro's number (N_A), the number of particles in a mole of a substance and is equal to 6.02×10^{23} particles.

Barometer, an instrument used to measure atmospheric pressure often by use of a mercury column or an evacuated metal box (Bourdon gauge).

Beats, periodic variations in loudness heard when sound waves of slightly different frequencies occur together. The constructive and destructive interference of these sound waves causes a sound that represents the difference in frequency.

Biasing, the process of providing the correct DC operating voltages for semiconductor devices such as diodes and transistors.

Binding energy, the energy converted from mass when a nucleus is formed from its constituent nuclear particles, all initially in their free state.

Black hole, describes a region of space that contains matter so dense that even light cannot escape its gravitational force.

Bosons, the fundamental gauge particles of the Standard Model together with hadrons and leptons. The gauge bosons carry the fundamental forces of nature such as the electromagnetic photon and the gravitational graviton.

Boyle's law, states that for a fixed mass of gas at constant temperature the volume of the gas varies inversely with the pressure.

Capacitor, component used to store electric charge in circuits. Types include electrolytic, plastic film and ceramic.

Cathode ray oscilloscope (CRO), instrument used to display rapidly varying voltage wave shapes in electronic circuit testing.

Central maximum, the antinodal line through the centre of an interference pattern. All points on this line are equidistant from the two sources.

Centre of curvature, of a mirror is the centre of the sphere that the mirror forms a part of.

Centre of mass, the point at which the whole mass of an object is considered to be concentrated for the purpose of applying the laws of motion.

Centripetal force, the force directed inward that keeps any object moving in a curve.

Chain reaction, one in which the products of the reaction initiate further reactions. These can be controlled chain reactions such as in nuclear fusion reactors or uncontrolled chain reactions as in the nuclear bomb.

Charles' law, states that for a confined gas where the pressure remains constant, the volume of the gas is directly proportional to its Kelvin temperature.

Circuit, a closed pathway for the flow of electric current containing a source of EMF, conductors and load devices.

Coefficient of friction, the frictional force divided by the normal contact force for a given surface type.

Coefficient of linear expansion (α), of a solid is the change in length of a one metre length of a solid due to a temperature change of one degree Celsius.

Coefficient of volume expansion (β), of a liquid is the change in volume per cubic metre of a liquid due to a temperature change of one degree Celsius.

Coherent, another term for waves that are in phase or sources producing wave crests at the same time.

Compressional wave, or longitudinal wave, occurs when the particles of the medium vibrate in the same direction as the direction of propagation of the wave.

Computed tomography (CT), a modern medical imaging process producing X-ray slices of the body by rotating an X-ray scanner about the body. Often called a CAT scan.

Conduction, the process where heat energy is transferred through a medium by the vibrating particles of the medium, but without the particles moving with the heat energy transfer.

Conductor, a substance that passes the flow of electric current, as opposed to an insulator.

Control rod, a substance, such as cadmium, used to absorb neutrons within a nuclear reactor.

Convection, the process where heat energy is transferred through a medium by the particles of the materials, which actually move with the heat flow.

Coolant, a liquid that circulates through the reactor core to remove excess heat energy and stop it from overheating. Water and heavy water (deuterium oxide D_2O) are often used.

Coulomb's law, states that the magnitude and direction of the electrostatic force between two charges depends directly on the sign and magnitude of the charges and inversely on the square of the distance between them.

Critical angle, the angle of incidence that produces total internal reflection within an optical system.

Critical mass, the minimum mass required to sustain a chain reaction.

Decay series, a series of radioactive decays; for example, ^{238}U decays by a series of alpha and beta decays to ^{206}Pb.

Diffraction, the bending of waves as they pass through a slit or around the edges of objects.

Digital, an electrical signal whose magnitude exists in discrete steps from zero to some fixed value.

Diode, a semiconductor PN junction component used to control the direction of flow of DC current.

Dioptre, a lens maker's unit that defines the optical power of a lens. It is equivalent to the reciprocal of the focal length (in metres).

Direct current (DC), an electric current that has one direction of flow of charge continuously.

Dispersion, the separation of white light into its component wavelengths or colours.

Displacement, the change in position of an object in a given direction.

Doping, the process of adding chemical impurities to semiconductor crystalline materials to increase conductivity.

Dose equivalent, to quantify the potential of radiation to damage and ultimately kill cells, physicists use the sievert (Sv) unit. For example, for X-rays, γ-rays and β-particles 1 Gy = 1 Sv.

Effective dose, the dose that is obtained by summing the equivalent doses in all tissues and organs of the body weighted by their sensitivity to radiation.

Efficiency, a measure of the useful energy output compared with the energy input.

Elastic collision, one in which kinetic energy is conserved.

Elastic potential energy, the energy stored in a spring or other elastic body by virtue of its distortion, or change in shape.

Electric current, a flow of electric charge defined as conventional current (positive charge flow) and electron flow (negative charge flow).

Electric field, a zone of influence where a force acts on any electric charge brought into it. Represented with electric field lines.

Electric potential difference, the work done per unit charge as the charge is moved between two points in an electric field. Measured in volts.

Electrical resistance, the opposition to the flow of electric current in any conductor. The larger the resistance (ohm), the smaller the current flow for a given voltage.

Electromagnetic induction, the process of inducing an EMF in a conductor by a changing magnetic flux.

Electromagnetic relay, a switching device using a small current through a solenoid to control a much larger current via an electromagnetic field.
Electromagnetic waves, those that require no medium for transmission and travel at the speed of light in a vacuum. Include long wavelength radio waves through to short wavelength gamma rays, called the electromagnetic spectrum.
Electromotive force (EMF), the energy per unit charge supplied by a source of electric current.
Electroscope, one of the earliest instruments used to detect electrical charge or ionising radiation. An electroscope contains two thin pivoted metal foils inside a protective container.
Energy level diagram, shows the discrete energy level series that characterises the allowed electron energies in a particular atom's excited states.
Enrichment, the process of concentrating nuclear fuel to about 5% of the U-235 isotope so as to have a self-sustaining chain reaction.
Entropy, a measure of the orderliness of the particles of a system.
Error, discrepancy between an instrumental measurement of a quantity and its actual value. The causes of error may be random or systematic.
Escape velocity, the velocity needed to escape a planet's gravitational pull. For Earth it is 11.2 km s^{-1}.
Extrapolation, a graphical method that involves extending a line past the first or last data points.

Ferromagnetic, the class of material that is strongly attracted by magnets, as opposed to diamagnetic and paramagnetic materials.
Flat plate collectors, used in a common method of collecting energy from the sun in order to produce hot water. Utilises flat blackened absorbing surfaces.
Fluid, a substance that can flow, such as liquids, gases and plasma.
Fluorescence, the emission of light during the absorption of radiation from other sources.
Fraunhofer lines, dark absorption lines in the solar spectrum caused by atoms present in the low pressure solar atmosphere.
Frequency, the number of oscillations of a wave source per second, measured in hertz (Hz).
Friction, a force that resists motion between two surfaces in contact. May be either dynamic or static.
Fundamental frequency, the lowest natural frequency produced by an object or musical instrument.

Galvanometer, an instrument used to measure the magnitude and direction of electric currents.
Gauge pressure, the difference between atmospheric pressure and the pressure in a connected vessel.
Gravitational mass, a measure of the pull of gravity on an object. A spring balance is often used to measure gravitational mass.
Gravitational potential energy, the stored energy of an object by virtue of its position above a reference surface.

Hadrons, those nuclear constituents of matter influenced by the strong nuclear force. They include baryons and mesons. The proton and the neutron are the most well known examples.
Half-life, the time taken for half the radioactive atoms in a sample to decay.
Heat, term used to describe the internal energy transferred throughout the heating process.
Hooke's law, states that the displacement of a spring is proportional to the force applied.
Hydrostatics, the science of fluids at rest, as opposed to hydrodynamics.

Independent variable, the one in which a change is made to determine the effect on the dependent variable.
Inductor, an electronic component used to store energy in a magnetic field. Types include solenoids, relay coils and choke coils.
Inertial mass, a measure of resistance to motion. If a known force is applied to different objects, then the resultant acceleration is related directly to mass.
Inertial reference frame, one in which Newton's first law of inertia is valid. Rotating or otherwise accelerating frames of reference are non-inertial frames.
Instantaneous speed, the speed as measured over a very small period of time.
Integrated circuit, an active miniaturised circuit component that contains complex internal circuits fabricated from a body of semiconductor material. Types include linear and digital.
Interference, occurs with multiple wave sources when crests and troughs interact to either reinforce or cause cancellation at a point.
Internal energy, energy associated with the random vibrations or motions of electrons, atoms and molecules within an object; for example, chemical, heat and electrical energy.
Interpolation, determining the value of a measurement in between two or more other measurements. The usual method is by using a graph.
Ionising radiation, radiation that causes an atom to absorb so much energy that an electron completely escapes from the atom and a positive ion is produced.
Isotopes, nuclei with the same number of protons in their nuclei and the same chemical symbol, yet differing in the number of neutrons in their nuclei. Their atomic numbers will be identical but their atomic masses will be different.

Kinetic energy, energy possessed by virtue of the motion of an object.

Kirchhoff's laws, the circuit laws relating to current and voltage within electrical networks. The current junction law states that the sum of all currents entering any circuit junction is equal to the sum of all currents leaving that junction point. The voltage loop law states that the algebraic sum of all voltage changes encountered in any complete closed circuit loop is equal to zero.

Law of conservation of energy, energy is not lost, it just gets transferred from one place to another.

Law of conservation of momentum, in a closed system, the change in momentum is zero.

Leptons, those subatomic particles influenced by the weak interaction force, with the most common examples being electrons and neutrinos.

Light-year (ly), the distance travelled by light in one year. Numerically it is equal to 9.5×10^{15} m.

Logic circuits, a basic class of digital electronic circuits or ICs that act as switches, latches, counters and timers.

Luminescence, the property of emitting light, such as from paint containing radioactive radium in the pigment.

Magnetic flux density, the total number of magnetic field lines or flux passing per unit area in a magnetic field. Measured in teslas (T).

Magnetic resonance imaging (MRI), a diagnostic medical imaging technique using the magnetic properties of the hydrogen atoms present in the various body tissues.

Manometer, instrument for measuring liquid and gas pressures of moderate range in the laboratory. A U-tube manometer is partially filled with a liquid such as mercury or water.

Mass, a characteristic of a body's resistance to motion. Also called inertia.

Mass defect, the mass that has been converted to binding energy, that is, energy that binds the nuclear particles together.

Mass spectrometer, an instrument used to separate gaseous ions or isotopes in a magnetic field according to their mass differences.

Matter waves, the wavelike behaviour of matter as exhibited by subatomic particles. Originally suggested by Louis de Broglie.

Mechanical waves, waves that require an elastic medium for the transfer of energy.

Microscopy, the technique of using optical or electron waves to view very small objects. Primary types are the light microscope, SEM, TEM and STM.

Moderator, a substance that slows neutrons around the reactor core. Examples are carbon (graphite), water, heavy water (D_2O), or liquid sodium.

Momentum, a vector quantity, being the product of a scalar (mass) and a vector (velocity).

Monochromatic light, light consisting of only one wavelength. Light may also be polychromatic.

Motor principle, a conductor carrying an electric current has a force exerted upon it if it is placed within a magnetic field. Basis of electric motors, where the force is actually a torque produced on a rotor assembly.

Mutual induction, the induction of an AC EMF in one solenoid by the varying magnetic field of a closely positioned second solenoid.

Node, is created by the interacting of a trough of one wave and a crest of another, producing a point of zero displacement.

Normal reaction force, the force pushing on an object normal (perpendicular) to the surface.

Nuclear fission, the division of a nucleus into roughly two equal parts, at the same time emitting one, two or three neutrons.

Nuclear fusion, the formation of a single nucleus of higher mass by the combination of lighter nuclei. The reaction may also produce high energy, neutrons, protons, etc. In the sun four hydrogen nuclei are fused into a single helium nucleus in a three step process called the proton-proton cycle.

Nuclide, used to describe a particular atomic species, for example C-12.

Ohm's law, states that the current flowing through a conductor is directly proportional to the potential difference applied across the ends of the conductor, provided that temperature and other physical factors are kept constant.

Operational amplifier, a linear IC representing a very high gain voltage amplifier designed to amplify signals over a wide frequency range. An example is the 741 Op Amp.

Overtones, integer multiples of the fundamental vibration frequency. Also called harmonics, they play a large part in musical instrument sound quality.

Parallel connection, circuit connections that allow multiple circuit paths that branch and join, with a common voltage drop across each component.

Pascal's principle, pressure applied at any point to a fluid in a closed vessel is transmitted equally to every other point in the fluid.

Periodic motion, one in which the object travels over the same path in a repetitive manner; for example, a pendulum, with a given period measured in seconds.

Photoelectric effect, the process whereby electrons are emitted by a metal when illuminated by light of sufficiently high photon energy.

Photovoltaic cells, semiconductor materials that provide direct conversion of the Sun's radiation into electrical energy. Most common are the various types of silicon solar cells.

Polariser, a device that allows only one component of the electromagnetic field to pass through. Crossed polarisers will not allow any energy through.

Potentiometer, a variable resistor, usually containing three connection points, that is used as a voltage divider.

Power, a measure of the rate of energy output. It has the units joules per second or watts.

Presbyopia, the inability to focus on distant or close objects — a mixture of hypermetropia (long-sightedness) and myopia (short-sightedness).

Pressure, the force per unit area acting on a surface, measured in N m^{-2} or pascals (Pa).

Proper time, time of an event as measured by observers for which the event occurs at one place. Contrast this with dilated time (t) or relativistic time. Dilated time is longer than proper time.

Quantum mechanics, the mathematical model including wave equations, electron positional probability and the uncertainty principle, that is the basis of the modern description of the atom.

Quarks, the fundamental building blocks of hadron particles held together with gauge boson force particles called gluons. The Standard Model allows for eight gluons and six quarks.

Radiation, energy travelling through space. It can be transmitted in the form of waves, or as energetic particles.

Radioactive decay, the break-up of nuclei by either natural or artificially induced means.

Range, the horizontal displacement of a projectile moving in a gravity field.

Rare earth magnets, those modern materials made from rare earth element alloys that are very strongly ferromagnetic; for example, neodymium iron boride.

Reactance, the AC equivalent of resistance, a property that is dependent on circuit frequency. Types include both capacitive and inductive.

Rectification, a process using semiconductor diodes to convert AC voltages to DC voltages.

Refracting telescope, consists of two convex lenses. The objective lens has a long focal length and the eyepiece lens has a short focal length. These lenses are set up so their focal points coincide.

Refraction, the bending of waves at the boundary or interface, as they go from one medium to another.

Regular or specular reflection, occurs when parallel rays incident on a surface are reflected parallel. Reflection may also be diffuse.

Resonance, occurs when a body vibrates at its natural frequency, or a circuit oscillates at one particular frequency.

Rest length, the distance as measured by an observer at rest to the measuring instrument.

Reverberation time, the time it takes for the sound intensity to fall to one millionth of its original intensity, that is, to fall by 60 dB.

RMS, the root mean square method of measuring AC voltage magnitudes, which represents the equivalent DC rating in terms of resistance energy dissipation.

Rotational inertia, is different from mass because for a rotating object not all the mass is travelling at the same speed — the outside goes faster than the inside. How the mass is distributed in that object will determine how difficult it is to start or stop the object rotating.

Scalar quantities, those that do not include a direction; for example, mass.

Schwartzschild radius, the minimum distance from which it is still just possible to escape from a black hole. The boundary is called the event horizon.

Scientific notation, a shorthand means of expressing numbers and is often called exponential notation; for example, 3.5×10^{20}. Order of magnitude is the nearest power of ten.

Scintigraphy, the process of using gamma radiation to form images of body organs and tissues following injection of a suitable radiopharmaceutical compound. The process uses a gamma camera.

Scintillation, impacts of charged subatomic particles on a fluorescent screen or within a crystal in which flashes of light occur.

Semiconductor, crystalline material made from doped silicon in either P-type or N-type modes. These materials show a decrease in resistance as temperature rises.

Series connection, circuit connections that allow only single path current flow with the same current through each component.

Significant figures, those integers of a measurement known with certainty plus the next integer.

Simple harmonic motion, periodic motion in which the displacement is proportional to the force but in the opposite direction.

Solar design, the principles of dwelling construction that increase the use of solar energy efficient design. Include both passive and active design elements.

Sound, a form of energy produced by the vibrations of objects and carried by longitudinal mechanical waves. Characteristics include pitch, quality, timbre and frequency range.

Special relativity, the laws of physics have the same form in all inertial reference frames, and that light propagates through empty space with a definite speed (*c*) independent of the speed of the source or observer.

Specific gravity (SG), defined as the ratio of the mass of an object in air compared with the mass of an equal volume of water. Also called relative density.

Specific heat capacity (*c*), defined as the amount of energy required to raise the temperature of one kilogram of a substance by one degree Celsius or by one kelvin.

Specific latent heat of fusion, the amount of energy required to melt one kilogram of a substance at its melting point.

Specific latent heat of vaporisation, the amount of energy to change one kilogram of a substance from a liquid to a gas, at its boiling point.

Spherical aberration, the inability of a convex lens or a concave mirror to focus light to a precise point.

Split-ring commutator, a special contact assembly on a rotating coil shaft that allows the direction of current to be reversed every half-cycle in a motor or a generator.

Square wave clock, a digital IC producing a square wave voltage signal output used in precision timing processes.

Standard Model, summarises the known constituents of matter as well as the interactions between them. Consists of two parts called the electroweak theory and quantum chromodynamics.

Stroboscope, an instrument, which in its most common form consists of a xenon flash tube similar to that found in a camera flash. It can be made to flash at variable rates from about one hertz to many kilohertz.

Strong force, that which binds adjacent nucleons together. It is a very short-ranged force, which decreases rapidly as nuclear separation increases.

Superposition, the process of creating a resulting waveform by adding the displacements, from the equilibrium positions of two interacting waves.

Temperature, of an object is a measure of the average kinetic energy of its particles.

Terminal velocity, when the speed of an object falling through a fluid becomes constant.

Thermal conductivity (*k*), of any material is a measure of the rate of flow of heat through one square metre of a one metre thick layer of the material which has a temperature difference of one kelvin between each side.

Thermal energy, the sum of the kinetic and potential energies of all the particles of a substance.

Thermal neutrons, or slow neutrons, as distinct from high energy neutrons, which have to be slowed down by a moderator if they are to initiate any further nuclear reactions.

Time constant, the feature of any RC circuit that determines the time for the voltage across the capacitor to reach 63% of its final value. Also occurs in LC circuits.

Time dilation, the slowing down of time because of the effects of gravity or of high speed motion. 'Moving clocks are measured to run slowly.'

Total internal reflection, occurs when light travelling from a more optically dense to a less optically dense medium is reflected from the interface rather than transmitted.

Transformer, an electrical device using mutual induction between a primary and a secondary coil to achieve AC voltage level changes.

Transistor, an active three terminal semiconductor device used to switch and/or amplify electrical signals.

Transverse waves, waves where each point of the wave vibrates perpendicularly to the direction the wave is travelling.

Ultrasonography, process of using ultrasonic sound waves to probe the bodies of patients in medicine or properties of materials in engineering and physics.

Ultraviolet radiation, electromagnetic energy whose wavelengths stretch from 400 nm down to 1.0 nm. Classified into 3 bands, UV-A, -B, -C. The cause of serious sunburn or skin cancer.

Uncertainty, a measure of how confidently a measurement or result can be stated. It is a direct result of the limitations of an instrument.

Uncertainty principle, Heisenberg's quantum mechanical principle, which states the impossibility of very accurately measuring both the position and the momentum of a fundamental particle at the same time.

Unified atomic mass unit (u), one twelfth the mass of an atom of the carbon isotope of atomic mass 12.0000 (that is, ^{12}C). The unified mass unit includes the masses of the six electrons of the carbon atom: $1\ u = 1.6606 \times 10^{-27}$ kg.

Vector quantity, one that requires both magnitude and direction for its specification.

Velocity, the rate of change of displacement with time.

Voltage divider, a circuit type using a potentiometer or various discrete resistors to divide a source of voltage into smaller values.

Voltage regulator, a circuit that achieves very high stability of a DC output voltage. Often uses a three terminal linear IC.

Voltmeter, an instrument for measuring potential difference or voltage.

Wavelength (λ), the minimum distance between two points on the wave that are in phase.

Wave particle duality, the fact that electromagnetic radiation can be simultaneously considered as either a wave or as a particle stream.

Weight, a measure of the force of gravity acting on a mass, which will vary depending on what gravitational forces it is being subjected to.

Work, defined as the product of the force and the distance moved in the direction of the applied force. It is measured in joules.

Appendices

Appendix 1 SYSTEME INTERNATIONAL (SI) STANDARD UNITS OF MEASUREMENT AND THEIR DEFINITIONS

PHYSICAL QUANTITY	NAME	UNIT SYMBOL
Length	metre	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd
Density	kilogram per cubic metre	kg m^{-3}
Velocity	metre per second	m s^{-1}
Acceleration	metre per second squared	m s^{-2}
Momentum	kilogram metre per second	kg m s^{-1}
Frequency	hertz	Hz (1 Hz = 1 s^{-1})
Force	newton	N (1 N = 1 kg m s^{-2})
Pressure	pascal	Pa (1 Pa = 1 N m^{-2})
Energy, work	joule	J (1 J = 1 N m)
Power	watt	W (1 W = 1 J s^{-1})
Electric charge	coulomb	C (1 C = 1 A s)
Electric potential	volt	V (1 V = 1 J C^{-1})
Electric resistance	ohm	Ω (1 Ω = 1 V A^{-1})
Moment of force	newton metre	N m
Heat capacity	joule per kelvin	J K^{-1}
Specific heat capacity	joule per kilogram per kelvin	J kg^{-1} K^{-1}
Specific latent heat	joule per kilogram	J kg^{-1}
Electric field strength	volt per metre	V m^{-1}
Electric resistivity	ohm metre	Ω m
Celsius temperature	degree Celsius	°C
Pressure	standard atmosphere	1 atm (1 atm = 101 325 Pa)
Power of lens	dioptre	D (1 D = 1 m^{-1})
Electric energy	kilowatt hour	kWh

Appendix 2 METRIC PREFIXES AND THEIR ORIGINS

PREFIX	ABBREVIATION	MEANING	ORIGIN
exa	E	10^{18}	Greek *exa* — out of
peta	P	10^{15}	Greek *peta* — spread out
tera	T	10^{12}	Greek *teratos* — monster
giga	G	10^{9}	Greek *gigas* — giant
mega	M	10^{6}	Greek *mega* — great
kilo	k	10^{3}	Greek *khilioi* — thousand
hecto	h	10^{2}	Greek *hekaton* — hundred
deca	da	10^{1}	Greek *deka* — ten
baseunit	—	10^{0}	
deci	d	10^{-1}	Latin *decimus* — tenth
centi	c	10^{-2}	Latin *centum* — hundred
milli	m	10^{-3}	Latin *mille* — thousand
micro	μ	10^{-6}	Greek *mikros* — very small
nano	n	10^{-9}	Greek *nanos* — dwarf
pico	p	10^{-12}	Italian *piccolo* — small
femto	f	10^{-15}	Greek *femten* — fifteen
atto	a	10^{-18}	Danish *atten* — eighteen

Appendix 3 PHYSICAL CONSTANTS

QUANTITY	SYMBOL	APPROXIMATE VALUE
Speed of light in a vacuum	c	3.00×10^{8} m s^{-1}
Gravitational constant	G	6.67×10^{-11} N m^{2} kg^{-2}
Coulomb's constant	k	9.00×10^{9} N m^{2} C^{-2}
Charge on electron	$-e$	-1.60×10^{-19} C
Charge on proton	e	1.60×10^{-19} C
Electron mass	m_e	9.11×10^{-31} kg
Proton mass	m_p	1.673×10^{-27} kg
Neutron mass	m_n	1.675×10^{-27} kg
Atomic mass unit	u	1.660×10^{-27} kg
Boltzmann's constant	k	1.38×10^{-23} JK^{-1}
Planck's constant	h	6.63×10^{-34} Js
Avagodro's number	Na	6.02×10^{23} mol^{-1}
Magnetic constant	k	2.0×10^{-7} N A^{-2}
Electron volt	eV	1.6×10^{-19} J

Appendix 4 THE GREEK ALPHABET

Α	α	alpha	Ι	ι	iota	Ρ	ρ	rho
Β	β	beta	Κ	κ	kappa	Σ	σ	sigma
Γ	γ	gamma	Λ	λ	lambda	Τ	τ	tau
Δ	δ	delta	Μ	μ	mu	Υ	υ	upsilon
Ε	ε	epsilon	Ν	ν	nu	Φ	φ	phi
Ζ	ζ	zeta	Ξ	ξ	xi	Χ	χ	chi
Η	η	eta	Ο	ο	omicron	Ψ	ψ	psi
Θ	θ	theta	Π	π	pi	Ω	ω	omega

Appendix 5 USEFUL CONVERSION FACTORS

Mass 1 tonne = 10^3 kg 1 a.m.u. = 1.6606×10^{-27} kg	**Velocity** 1 km h^{-1} = 0.2778 m s^{-1}
Length 1 km = 10^3 m 1 cm = 10^{-2} m 1 mm = 10^{-3} m 1 μm = 10^{-6} m 1 nm = 10^{-9} m	**Angle** 1 rad = 57.295° 1° = 0.0175 rad
Volume 1 litre (L) = 1000 cm^3 = 10^{-3} m^3 1 gallon (US) = 3.7854 L 1 gallon (UK) = 4.5461 L	**Frequency** 1 Hz = 1 cycle s^{-1} 1 rev min^{-1} = 0.0167 Hz
Density 1 g cm^{-3} = 10^{-3} kg m^{-3}	**Force** 1 N = 1 kg m s^{-2}
Time 1 min = 60 s 1 h = 60 min = 3600 s 1 solar day = 24 h = 1440 min = 86 400 s	**Pressure** 1 Pa = 1 N m^{-2} 1 atm = 1.013×10^5 Pa = 760 Torr = 760 mmHg

Appendix-Walding 4th 25/8/04 10:09 AM Page 779

Appendix 6 PERIODIC TABLE OF THE ELEMENTS

Key

6 ← Atomic number C ← Symbol Carbon 12.01 ← Atomic mass

I A	II A	III B	IV B	V B	VI B	VII B	VIII	VIII	VIII	I B	II B	III A	IV A	V A	VI A	VII A	
1 H Hydrogen 1.008																	2 He Helium 4.003
3 Li Lithium 6.941	4 Be Beryllium 9.012											5 B Boron 10.81	6 C Carbon 12.01	7 N Nitrogen 14.01	8 O Oxygen 16.00	9 F Fluorine 19.00	10 Ne Neon 20.18
11 Na Sodium 22.99	12 Mg Magnesium 24.31											13 Al Aluminum 26.98	14 Si Silicon 28.09	15 P Phosphorus 30.97	16 S Sulfur 32.06	17 Cl Chlorine 35.45	18 Ar Argon 39.95
19 K Potassium 39.10	20 Ca Calcium 40.06	21 Sc Scandium 44.96	22 Ti Titanium 47.90	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 58.70	29 Cu Copper 63.55	30 Zn Zinc 65.38	31 Ga Gallium 69.72	32 Ge Germanium 72.59	33 As Arsenic 74.92	34 Se Selenium 78.96	35 Br Bromine 79.90	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 95.94	43 Tc Technetium (97)	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 Cesium 132.9	56 Ba Barium 137.3	57[1] La Lanthanum 138.9	72 Hf Hafnium 178.5	73 Ta Tantalum 180.9	74 W Tungsten 183.9	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 204.4	82 Pb Lead 207.2	83 Bi Bismuth 209.0	84 Po Polonium (209)	85 At Astatine (210)	86 Rn Radon (222)
87 Fr Francium (223)	88 Ra Radium 226.0	89[2] Ac Actinium (227)	104 Rf Rutherfordium (257)	105 Db Dubnium (260)	106 Sg Seaborgium (263)	107 Bh Bohrium (262)	108 Hs Hassium (269)	109 Mt Meitnerium (268)	110 Ds Darmstadtium (271)	111 Uuu (272)	112 Uub (285)	113 Uut	114 Uuq	115 Uup	116 Uuh		

Lanthanoid Series[1]	58 Ce Cerium 140.1	59 Pr Praseodymium 140.9	60 Nd Neodymium 144.2	61 Pm Promethium (145)	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 173.0	71 Lu Lutetium 175.0
Actinoid Series[2]	90 Th Thorium 232.0	91 Pa Protactinium 231.0	92 U Uranium 238.0	93 Np Neptunium 237.0	94 Pu Plutonium (244)	95 Am Americium (243)	96 Cm Curium (247)	97 Bk Berkelium (247)	98 Cf Californium (251)	99 Es Einsteinium (254)	100 Fm Fermium (253)	101 Md Mendelevium (256)	102 No Nobelium (254)	103 Lr Lawrencium (257)

Appendix 7 RELATIVE ATOMIC MASSES

NAME	SYMBOL	ATOMIC NUMBER	RELATIVE ATOMIC MASS	NAME	SYMBOL	ATOMIC NUMBER	RELATIVE ATOMIC MASS
* Actinium	Ac	89	(227)	Mercury	Hg	80	200.6
Aluminium	Al	13	27.0	Molybdenum	Mo	42	95.9
* Americium	Am	95	(243)	Neodymium	Nd	60	144.2
Antimony	Sb	51	121.8	Neon	Ne	10	20.2
Argon	Ar	18	39.9	* Neptunium	Np	93	(237)
Arsenic	As	33	74.9	Nickel	Ni	28	58.7
* Astatine	At	85	(210)	Niobium	Nb	41	92.9
Barium	Ba	56	137.3	Nitrogen	N	7	14.0
* Berkelium	Bk	97	(247)	* Nobelium	No	102	(259)
Beryllium	Be	4	9.0	Osmium	Os	76	190.2
Bismuth	Bi	83	209.0	Oxygen	O	8	16.0
Boron	B	5	10.8	Palladium	Pd	46	106.4
Bromine	Br	35	79.9	Phosphorus	P	15	31.0
Cadmium	Cd	48	112.4	Platinum	Pt	78	195.1
Caesium	Cs	55	132.9	* Plutonium	Pu	94	(244)
Calcium	Ca	20	40.1	* Polonium	Po	84	(209)
* Californium	Cf	98	(251)	Potassium	K	19	39.1
Carbon	C	6	12.0	Praseodymium	Pr	59	140.9
Cerium	Ce	58	140.1	* Promethium	Pm	61	(145)
Chlorine	Cl	17	35.5	* Protactinium	Pa	91	(231)
Chromium	Cr	24	52.0	* Radium	Ra	88	(226)
Cobalt	Co	27	58.9	* Radon	Rn	86	(222)
Copper	Cu	29	63.5	Rhenium	Re	75	186.2
* Curium	Cm	96	(247)	Rhodium	Rh	45	102.9
Dysprosium	Dy	66	162.5	Rubidium	Rb	37	85.5
* Einsteinium	Es	99	(254)	Ruthenium	Ru	44	101.1
Erbium	Er	68	167.3	Rutherfordium	Rf	104	261
Europium	Eu	63	152.0	Samarium	Sm	62	150.4
* Fermium	Fm	100	(257)	Scandium	Sc	21	45.0
Fluorine	F	9	19.0	Selenium	Se	34	79.0
* Francium	Fr	87	(223)	Silicon	Si	14	28.1
Gadolinium	Gd	64	157.3	Silver	Ag	47	107.9
Gallium	Ga	31	69.7	Sodium	Na	11	23.0
Germanium	Ge	32	72.6	Strontium	Sr	38	87.6
Gold	Au	79	197.0	Sulfur	S	16	32.1
Hafnium	Hf	72	178.5	Tantalum	Ta	73	180.9

Appendix 7 Cont'd

NAME	SYMBOL	ATOMIC NUMBER	RELATIVE ATOMIC MASS	NAME	SYMBOL	ATOMIC NUMBER	RELATIVE ATOMIC MASS
Helium	He	2	4.0	* Technetium	Tc	43	(98)
Holmium	Ho	67	164.9	Tellurium	Te	52	127.6
Hydrogen	H	1	1.0	Terbium	Tb	65	158.9
Indium	In	49	114.8	Thallium	Tl	81	204.4
Iodine	I	53	126.9	* Thorium	Th	90	232.0
Iridium	Ir	77	192.2	Thulium	Tm	69	168.9
Iron	Fe	26	55.8	Tin	Sn	50	118.7
Krypton	Kr	36	83.8	Titanium	Ti	22	47.9
Lanthanum	La	57	138.9	Tungsten	W	74	183.9
* Lawrencium	Lr	103	(260)	* Uranium	U	92	238.0
Lead	Pb	82	207.2	Vanadium	V	23	50.9
Lithium	Li	3	6.9	Xenon	Xe	54	131.3
Lutetium	Lu	71	175.0	Ytterbium	Yb	70	173.0
Magnesium	Mg	12	24.3	Yttrium	Y	39	88.9
Manganese	Mn	25	54.9	Zinc	Zr	30	65.4
* Mendelevium	Md	101	(256)	Zirconium	Zr	40	91.2

* Unstable elements.
Value in brackets is the mass number of the isotope with the longest half-life.

Appendix 8 PROPERTIES OF THE NUCLIDES

Z = atomic number = number of protons
A = atomic mass = number of protons plus neutrons
M = exact mass of the nuclide including electrons (in u)

Z	A	M
-1e	0	0.000549
0n	1	1.008665
1p	1	1.007276
1H	1	1.007825
	2	2.014102
	3	3.016050
2He	3	3.016030
	4	4.002603
	6	6.018893
	8	8.034
3Li	6	6.015124
	7	7.016004
	8	8.022487
	9	9.02680
4Be	6	6.01972
	7	7.016929
	9	9.012186
5B	8	8.024609
	10	10.012938
	11	11.009305
6C	9	9.03104
	10	10.01686
	11	11.011432
	12	12.000000
	13	13.003354
	14	14.003242
7N	12	12.01864
	13	13.005738
	14	14.003074
8O	13	13.0248
	16	15.994915
	17	16.999133
	18	17.999161
11Na	22	21.994437
	23	22.989771
	24	23.990964
13Al	27	26.981541
15P	30	29.97832
	31	30.973765
	32	31.973909
16S	35	34.969033
17Cl	36	35.968307
19K	40	39.964000
28Ni	61	60.93106
	64	63.92796
29Cu	64	63.929757
30Zn	64	63.929140
	65	64.92923
36Kr	84	83.911505
	90	89.9197
	91	90.923
	92	91.92182
37Rb	90	89.9148
38Sr	88	87.905628
	90	89.90775
	93	92.9142
	94	93.9154
42Mo	100	99.9076
47Ag	107	106.905091
	108	107.905953
48Cd	113	112.904408
49In	115	114.90387
	116	115.90553
50Sn	116	115.90179
52Te	137	136.910
54Xe	135	134.91350
55Cs	130	129.90676
	135	134.90590
56Ba	136	135.90456
	141	140.91402
	143	142.921
	144	143.923
81Tl	208	207.98201
82Pb	206	205.97447
	208	207.97666
83Bi	212	211.99128
84Po	212	211.988865
	216	216.00192
86Rn	220	220.01139
	222	222.01761
88Ra	224	224.02020
	226	226.02544
	228	228.03110
89Ac	228	228.03104
91Pa	234	234.04342
92U	233	233.03965
	234	234.04098
	235	235.04394
	236	236.04559
	238	238.05082
	239	239.05433
93Np	239	239.05295
94Pu	239	239.05218
	240	240.05384
	241	241.05687
95Am	239	239.05304
	241	241.05685

Index

555 timer IC 537
741 op-amp IC 722, 536
absolute refractive index 402
absolute zero 257
absorbed dose 751
AC induction motor 568
AC reactance 711, 713
AC ripple 520
acceleration,
  centripetal 154
  defined 33
  gravitational 41
  instantaneous 35
accuracy 17
achromatic doublet 426
acoustics 373
action potentials 474
active device 528
ADC – DAC 525, 526
alternating current (AC) 470
alternator 584, 585
ammeters 482
Amonton's law 257
Ampere 469, 564
amplifier
  (AC voltage) 717
  (DC-direct current) 529
  audio 366, 723
  class A 718
  operational (op-amp) 535, 536, 721
amusement park 226
analog signal 525
angle of,
  incidence 388, 401, 384
  reflection 384, 388
  refraction 401
antimatter 669
aphelion 149
apogee 149
apparent depth 408
Archimedes' principle 171
Aristotle 112, 597
astable multivibrator 538
astigmatism 432
astrology 157
astronomical distances 135
atom,
  mass 611, 613
  structure 607, 609
atomic force microscope (AFM) 747
aurorae 556
Avogadro, Amedo 259
Avogadro, hypothesis 259
ballet 186
ballistics 116
balloons 176
barometer
  aneroid 168
  manometer 168
  mercury 168
baseball 194
beat frequency 362
Becquerel, Henri. 603
Becquerel, unit 626
Bell, Alexander Graham 364
bends, diving178
big bang 707
billiards 207
binary signals 527, 727
binding energy, 613
bits and bytes 526
black body radiation 655
black hole 143, 156
Bohr atom 659
Bohr, Neils 609
Boltzmann, Ludwig 259
Boltzmann's constant 259
boson particles 667, 669

Bourdon gauge 167
Boyle, Robert 256
Boyle's law 256
calorimetry 247
camber, road 119
camera systems 434
capacitors and capacitance 509, 720
cars
  accidents 101
  battery 174
  brakes 96
  collisions 190
  road camber 116
  torque 202
  tyres 166, 229
Cartesian diver 175
cathode ray oscilloscope (CRO) 503
Celsius, Anders 243
centre of curvature 388, 419
centre of mass 184
centripetal
  acceleration 154
  force 153
Chadwick, James 608
Charles, Jacques 257
Charles' law 256
chromatic aberration 426
chromosphere 733
circular motion 149
  non-uniform 120
  uniform 117
  vertical 117
coefficient, linear expansion 264
coherency 323
collisions, energy 190, 218, 220
Coloumb, Charles Augustin 451
Coloumb's law 453
compact discs 337, 367, 368
Compton collisions 657
Compton, Arthur Holly 657
computed tomography (CT) 754
concrete 167
conductivity
  heat 272
  thermal 274
conductors and insulators 448
conservation of charge 447
convection 276
conventional current 469
conversions 9
Copernicus 134
corner reflector 387
cosmic microwave radiation 138
cosmology 672
critical angle 409
Crookes, William 598
Crookes'
  radiometer 655
  tube 598
Curie temperature 552
Curie, Marie 603
Curie, Pierre 603
current amplification (gain) 529
curved mirrors 387, 396
cyclotron accelerator 571
Dalton, John 597
dark matter 671
Darlington pair 530
data projector 371
data-logger 48, 502
DC bell, buzzer 561
decibel scale (sound) 364
defibrillation 494
Democritus 596
density 79
derived units 7
deuterons 634
diamagnetic substance 549
dielectric 510
diffraction 321, 352
diffraction, water waves 311
diffuse reflection 385
digital
  counter circuit 726, 727
  signal 525
  television 371
direct current (DC) 469
disintegration 629
displacement 27
Dolby digital sound 369
Doppler effect 374, 376
Doppler ultrasonography 748
dose equivalent 751
double insulated 490
drag coefficient 89
DVD (digital versatile disc) 369
Earth's magnetism 555
echogenicity 748
eddy currents 587
efficiency 212
Einstein, Albert 636, 656
electric

ammeter – voltmeter 482
chair 497
charge 444, 450
circuit 469
current 451
fence 497
field 454
fish and eels 461
motors 566

potential 456,457
shock 493
electrical
generators 583
power 488
power transmission 588
resistance 474
signals 524
electrochemical series 471
electrolyte 467
electromagnetic
circuit breaker (ELCB – RCD) 491
induction 576
relay 514
torque 564
electromagnetism 557
electromagnets 557, 560
electromotive force (emf) 471
electron
drift velocity 468
flow 469
gun (CRO) 462
electronic resistors 506
electron-volt (eV) 459
electrophorus 451
electroscope 450
electrostatic
induction 450
potential energy 457
elevation angle 111
energy
level diagram 659
binding 613
conservation 225
defined 211
EPE 230
GPE 222
internal 241
kinetic 212, 216
potential 212
relativistic 700
rest 700
thermal 241
entropy 251
equations of motion 38
equipotentials 458
errors
absolute 18
random 15
relative 18
systematic 15
evaporation 250
explosions 189, 198
Faraday, Michael 576
Faraday's law 578
feedback resistor 536
ferromagnetic substance 549
fire walking 275
fish 175
floating 173
fluids 164
fluorescence 342
focal length 388, 419
force
centripetal
strong 620
balanced 75
angled 214
planar 217
forensic science 196
Fosbury flop 187
frames of reference 681, 685
Franck-Hertz experiment 661
fraunhofer lines 732
freefall
cats 90
motion 45, 84
raindrops 91
cats 208
Fresnel, Augustine 327
friction 95-97
full scale deflection (FSD) 482
fundamental frequency 354, 356
fuse 490
fusion, latent heat 248
Galileo 112, 435
Galvani, Luigi 466
galvanometers 482, 566
gamma rays 342
gas laws 255, 258
general relativity 701
graph types 65-68

graphs v/t 36
graphs, d/t 31
gravitation, Newton's law
gravitational
  field 155
  force 654
  lensing 704
  red shift 706
gravitron 120
greenhouse effect 733
Grimaldi, Francesco 327
guns, bullets 205
hadron therapy 753
hadrons and leptons 667
half life 626
harmonics 354, 357, 358, 356
hearing range 351
heat flow 273
heat sink 716
home theatre systems 367
Hooke, Robert 231
Hooke's law 231
household electricity 489
Hubble space telescope 151
Hubble's law 137
human eye 429
human nervous system 473
Huygens, Christian 307
Huygens' principle 306
hydraulic press 170
hydrometer 174
hydrostatics 164
ignition coil 586
impulse 188
inclined planes 91-93, 100
inductors and inductance 513
inertia 204
infrasound 351
insulator, heat 273, 280
integrated circuit (IC) 524, 534, 537
interference 321, 352
  sound 315
  water waves 311
internal resistance 477
isotherms 280
isotopes 610
Joliot, Frederick, 604
Joule, James 241
Kelvin, Lord 243
Kennedy assassination 196, 197
Kepler's laws 145
kilowatt hour (kW h) 488
kinetic theory, gases 254
Kirchhoff, Gustav Robert 480
Kirchhoff's laws 480
Larmor frequency 756
laser 383
latent heat
  fusion 248
  vaporisation 249
lateral inversion 386
Lego RCX brick 541
length 5
  contraction 692
  proper 692
lens formula 423
lenses, biconcave and biconvex 418
Lenz's law 580
light
  dependent resistor (LDR) 519, 531
  emitting diode (LED) 518
  pressure 658
  rays 382
  reflection 384
  refraction 400, 407
  waves 320, 338
  gravitational deflection 704
  speed 684
Lippershey, Hans 435
liquid crystal display (LCD) 344
logic circuits 724, 725
loudspeaker systems 366, 564
lubricants 97
luminous object 382
magnetic
  bottle 571
  constant 559
  domains 551
  induction 550
  materials 548, 551
  monopoles 553
  poles 549
  resonance imaging (MRI) 756
  spin 552
magnetosphere 555
magnets (rare earth) 549
mass
  centre of 184
  defined 6
  defect 612, 636
  spectrometer 570
matter waves 665

Maxwell, James Clerk 338
Maxwell's screw rule 558
medical isotopes 748, 750
metallic lattice 467
Michelson-Morley experiment 685
microcontroller 539
micrometer 20
microphones 365, 585
microprocessor (CPU) 524
microscope 438
microwaves 341
Millikan, Robert 455, 600
Millikan's experiment 600
mirage 413
mirror formula 393
mirrors, concave and convex 387
mirrors, plane (flat) 383
momentum
  angular 203-206
  conservation 189
  defined 187
monochromatic 323
Montgolfier brothers 175
motor principle 562
multimeter 484, 501
multiple earth neutral system (MEN) 489
mutual induction 586
myopia and hypermetropia 431
neutrinos 669
neutron
  star 143, 205
  discovery 608
Newton, Isaac 148
Newton's
  1st Law 76
  2nd Law 80
  3rd Law 83
  law of gravitation 148
  rings 336
nodal lines 314, 323, 324, 328
nuclear
  applications 648
  chain reaction 637
  Chernobyl 642
  control rods 639
  coolant 639
  fast breeder 638, 640
  fission 634
  fuel 639
  fusion 634, 643, 732
  medicine 748
  meltdown 637
  moderator 639
  power 637
  reactor 638
  stability 614
  subcritical 639
  transmutations 620
  waste 641
nucleons 614
nucleus 610
nuclide 610
octave 354
Ohm, Georg Simon 475
Ohm's law 475
ohmic conductors 480
Olber's paradox 135
optical
  density 401
  fibres 411
  images, curved mirrors 389
  images, lenses 421
  images, plane mirrors 386
  magnification 391, 423
  power (dioptre) 419
  prisms 411
  wedges 336
order of magnitude 14
overtones 354, 356
ozone layer 741
parallel connection 472
paramagnetic substance 549
particle accelerator 571
Pascal, Blaise 165
Pascal, unit 165
Pascal's principle 169
Pauli exclusion principle 659
peak voltage 504
pendulum 127, 128
pendulum, ballistic 227
perihelion 149
phase change 333
phase, waves 291
photoelectric effect 656
photons 656
photosphere 732
photothermal devices 735
photovoltaic devices 734, 738
PIC microchip 540
piezo electric crystals 747
pitchblende 603
Planck, Max 655

Planck's constant 655
PN junction diode 517
polarisation 343
polychromatic 327
positron 621
positron emission tomography (PET) 758
positronium 145
potential difference 457, 472, 471
potentiometer (rheostat) 477
power 223
power supply 502
precision 17
pressure
  blood 169
  depth 177
  types 165
principal axis 388, 419
principal focus 388, 419
projectiles 109
proton, discovery 608
pulleys 93
quantum
  chromodynamics (QCD) 667
  mechanics 654, 659, 665
  orbitals 659
quarks 455, 669
quicksand 175
radar 341
radiation
  therapy 752
  absorbed dose 645
  background 646, 651
  biological effect 645
  emr 278
  heat 278
  ionizing 605
  magnetic effect 606
  penetrating 605, 616
  quality factor 646
  radiolytic 648
  risks 646
  Sievert 646
radio pharmaceuticals 748, 750
radioactive dating 632
radioactivity 604
  decay 621, 623, 625
  detectors 604, 617
  laws 628
rainbow 413
range 111, 113
Rayleigh criterion 331
RC filters 512
recessional velocity (RV) 663
record player 123
rectification 519
red shift 706
reflection, waves 290, 308
refraction, water 309
relative motion
relative refractive index 405
relativity
  general 701
  simultaneity 687
  special 680
resistor colour code 507
resistors 508
resonance 361
restitution 221
Ride, Sally 161
RLC resonance & circuits 710
robolab 542
robotics 539
rockets 197
rods and cones 430
Roentgen, Willhelm. 602
Rutherford, Ernest 605
sampling rate 526
satellites 150, 152
scalar quantity 27
scanning electron microscope (SEM) 746
scanning tunnelling microscope (STM) 747
scientific notation 10
scintigraphy 748
seismic wave 310
self inductance 513
self-induction 577
semiconduction 515
semiconductor biasing 517
semiconductor doping 516
semiconductors 449
series connection 472
ships 175
shunt resistor 482
SI units 3
significant figures 12
silicon (N and P types) 516
simple cell (battery) 470
single photon emission computes tomography (SPECT) 757
slide projector 439
Snell, Willebrod 402
Snell's law 402

solar
cell 518, 738
collectors 736
corona 733
design (passive and active) 735
radiation 733
winds 556
solenoid 559, 561, 579
sound 350
barrier 377
intensity 363
special relativity 680, 686
specific
gravity 171
heat 245
spectral dispersion 406
spectroscopy 662
specular reflection 385
speed
average 29
instantaneous 35
spherical aberration 395, 425
sphygmomanometer 169
spin dryer 120
sport
squash 128
tennis 128
boats 71
ballet 186
baseball 296
basketball 194
bicycles 81, 208
billiards 207
cars 48, 118, 124, 166, 170, 208
cricket 128
diving 178, 205
jumping 187
parachute 89
planes 81
running 228
shooting 192
shotput 194
skiing 208
soccer 221
tennis 220, 296
walking 228
spring constant 125, 231
square wave clock 537
standard model of particle physics 654, 667, 668
standards 4
stepper motors 568
stroboscope 21
stroboscope, water waves 306
strong nuclear force 654
subwoofer 367
Sun (Sol) 731
superman 233
superposition, waves 294
tank oscillator 715
telescope
Galilean 437
reflecting 437
refracting 436
terrestrial 436
temperature 242
Celsius 242
Fahrenheit 243
Kelvin 243
Tesla, Nicola 554
thermal expansion 263
thermistor 519
thermodynamics, defined 241
thermodynamics, laws 251
thermography 278
thermometers 244
thermos flask 279
Thompson, J.J. 598, 600
thyristor (SCR) 531
ticker timer 46
time 5
time
constant 512, 513
dilated 691
proper 691, 693
variability 688
torque 202
total internal reflection 409
trains 229
transformer 586
transistor
biasing 719
DC switch 530
field effect (FET) 532
power 716
structure 528
transmission electron microscope (TEM) 746
triboelectric series 446
trombe wall 736
tuning circuits 715
tweeter 367
twin paradox 695
ultrasonic 355

ultrasonic waves 747
ultraviolet radiation (UV) 740
uncertainty 16
uncertainty principle 666
universe
baryogenesis 141
universe
  dark matter 142
  inflation 141
  nucleosynthesis 142
valence electrons 467
Van Allen, James 556
Van de Graaff generator 448
vector
  addition 57
  quantity 27, 56
  resolution 62
  subtraction 60
velocity
  angular 122
  average 29
  escape 152
  instantaneous 35
  radial 118
  tangent 118
  relativistic 698
  terminal 88
Vernier caliper 20
vertical motion 43
Villard, Paul. 605
virtual image 386
viscosity 164
visible spectrum 320, 339
Volta, Alessandro 466
voltage 471
  amplification 721
  divider 508
  induced 579
  regulator 520, 715
  root mean square (RMS) 504
voltmeters 482
water, abnormal expansion 268
Watt 488
wave
  characteristics 322, 341, 349
  equation 289, 656
  function generator 722
wave-particle duality 658
waves
  electromagnetic 338
  infrared and ultraviolet 341
  sound 348
  television and radio 339
  characteristics 287
  EMR 284
  graphs 297
  longitudinal 286
  mechanical 284
  ripple tank 305
  seismic 310
  standing 295
  superposition 294
  transverse 286
  water 305
weak nuclear force 654
Weber, Wilhelm 554
weight
  apparent 87
  defined 86
wimps 671
wind instruments 357
work 211, 213
X-rays 342
Young, Thomas 322
Zener diode 518