PEARSON

PHYSICS

WESTERN AUSTRALIA

STUDENT BOOK

12

Pearson Australia
(a division of Pearson Australia Group Pty Ltd)
707 Collins Street, Melbourne, Victoria 3008
PO Box 23360, Melbourne, Victoria 8012

www.pearson.com.au

First published 2018 by Pearson Australia
2021 2020 2019 2018
10 9 8 7 6 5 4 3 2 1

Publisher: Malcolm Parsons
Project Manager: Shelly Wang, Beth Zeme
Production Manager: Anji Bignell, Natalie Lincoln
Development Editor: Vicky Chadfield
Cover designer: Anne Donald
Typesetter: Aptara
Rights & Permissions Editor: Amirah Fatin
Editor: Marta Veroni
Senior Publishing Services Analyst: Rob Curulli
Proofreader: Jane Fitzpatrick
Indexer: Max McMaster
Illustrator/s: DiacriTech
Printed in China

National Library of Australia cataloguing-in-publication data:

Pearson Physics 12 : Western Australia / Greg Moran.

ISBN: 9781488613579 (pbk.)

Target Audience: For secondary school age.

Subjects: Physics—Study and teaching (Secondary)—Western Australia.

Physics—Textbooks.

Pearson Australia Group Pty Ltd ABN 40 004 245 943

Attributions
WACE Physics ATAR Course Year 12 Syllabus © School Curriculum and Standards Authority, Government of Western Australia, 2017

Some of the images used in *Pearson Physics 12 Western Australia Student Book* might have associations with deceased Indigenous Australians. Please be aware that these images might cause sadness or distress in Aboriginal or Torres Strait Islander communities.

PEARSON
PHYSICS
WESTERN AUSTRALIA 12

Writing and Development Team

Greg Moran
Head of Science
Past President STAWA
Coordinating Author

Doug Bail
Education Consultant
Past Head of Science
Author

Dylan Bayley
Electrical Engineer
Answer Checker

Stephen Brown
Teacher
Reviewer

Keith Burrows
Educator
Contributing Author

Robert Chapman
Educator
Contributing Author

Ann Conibear
Teacher
Contributing Author

Subrat Das
Educator
Reviewer

Tracey Fisher
Teacher,
Former Lecturer, Research Scientist
Author and reviewer

Carmel Fry
Head of Snr Curriculum
Contributing Author

Mark Hamilton
Teacher
Contributing Author

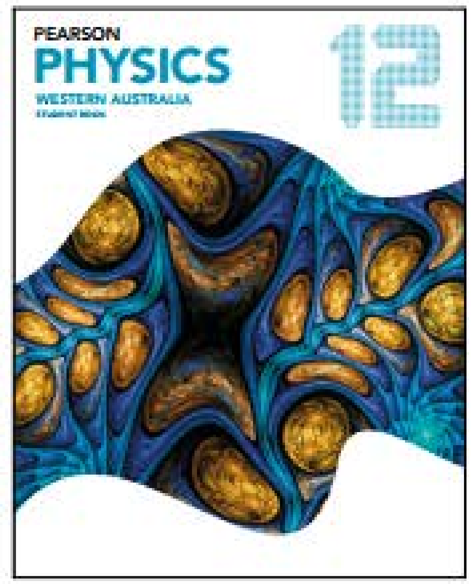
Geoff Lewis
Head of Science
Past President STAWA
Author

Elke McKay
Teacher
Author and Reviewer

David Madden
Educator
Contributing Author

Cameron Parsons
Product Engineer
Answer Checker

Gregory White
Physics Educator
Author, Answer Checker



Contents

UNIT 3 Gravity and electromagnetism

UNIT 4 Revolutions in modern physics

AREA OF STUDY 3

Wave-particle duality and the quantum theory

AREA OF STUDY 4

Special relativity

AREA OF STUDY 5

The Standard Model

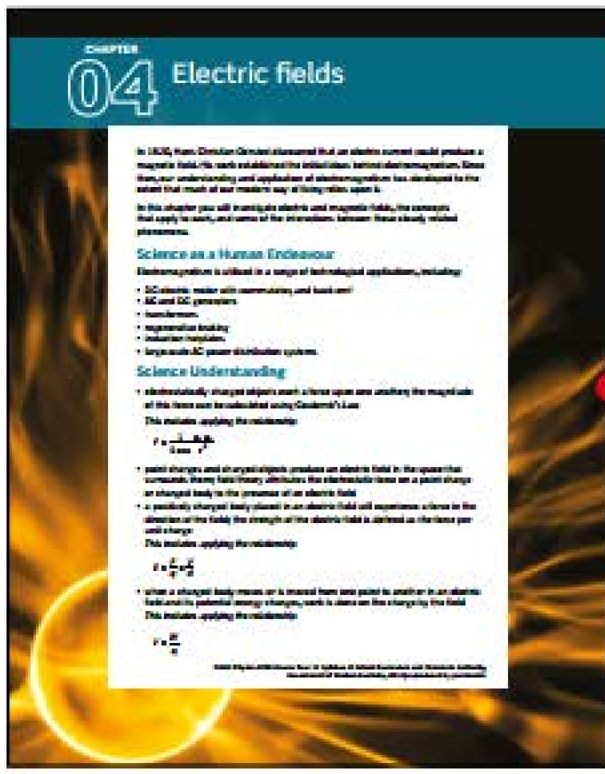
AREA OF STUDY 6

Science inquiry skills

How to use this book

Pearson Physics 12 Western Australia

Pearson Physics 12 Western Australia has been written to the *WACE Physics ATAR Course, Year 12 Syllabus 2017*. Each chapter is clearly divided into manageable sections of work. Best practice literacy and instructional design are combined with high quality, relevant photos and illustrations. Explore how to use this book below.

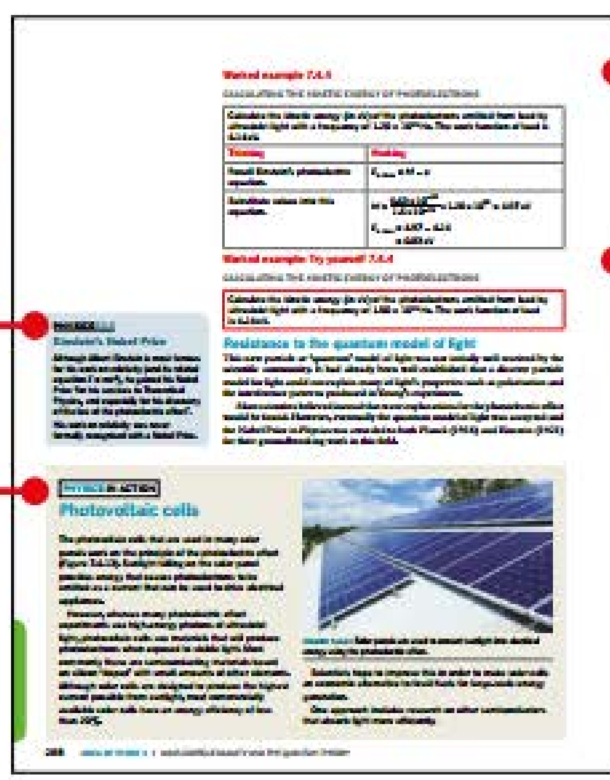


Chapter opening page

The chapter opening page links the syllabus to the chapter content. Science Understanding and Science as a Human Endeavour addressed in the chapter is clearly listed.

REVISION

Revises important concepts from Year 11



PHYSICSFILE

PhysicsFiles include a range of interesting information and real world examples.

PHYSICS IN ACTION

Physics in Action boxes place physics in an applied situation or relevant context and encourage students to think about the development of physics and its use and influence of physics in society.

Worked examples

Worked examples are set out in steps that show both thinking and working. This enhances student understanding by clearly linking underlying logic to the relevant calculations.

Each Worked example is followed by a Try yourself: Worked example. This mirror problem allows students to immediately test their understanding.

Section summary

Each section includes a summary to assist students consolidate key points and concepts.

Section review questions

Each section finishes with questions to test students' understanding and ability to recall the key concepts of the section.

Chapter review

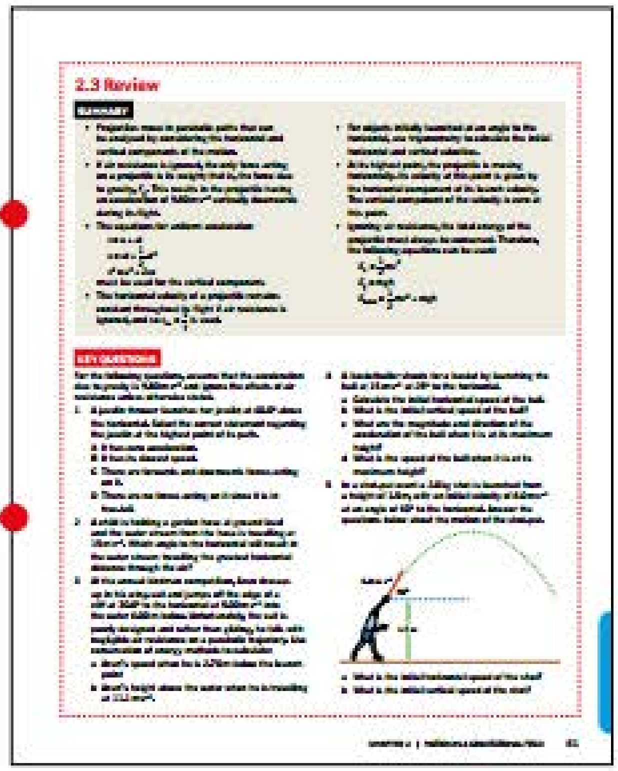
Each chapter finishes with a set of higher order questions to test students' ability to apply the knowledge gained from the chapter.

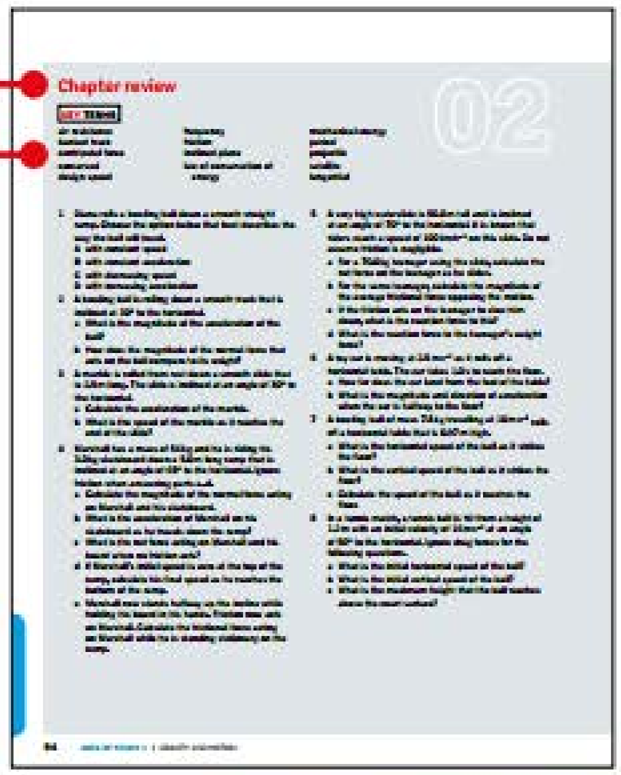
Key terms and glossary

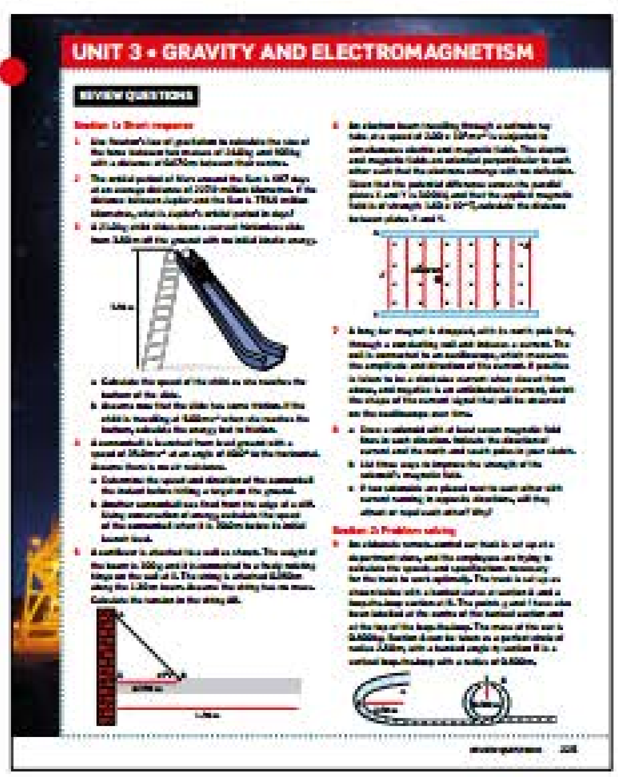
Key terms are shown in bold and listed at the end of each chapter. A comprehensive glossary at the end of the book includes and defines all the key terms.

Unit review

Each unit finishes with a comprehensive set of exam-style questions that assist students to draw together their knowledge and understanding and apply it to this style of question.

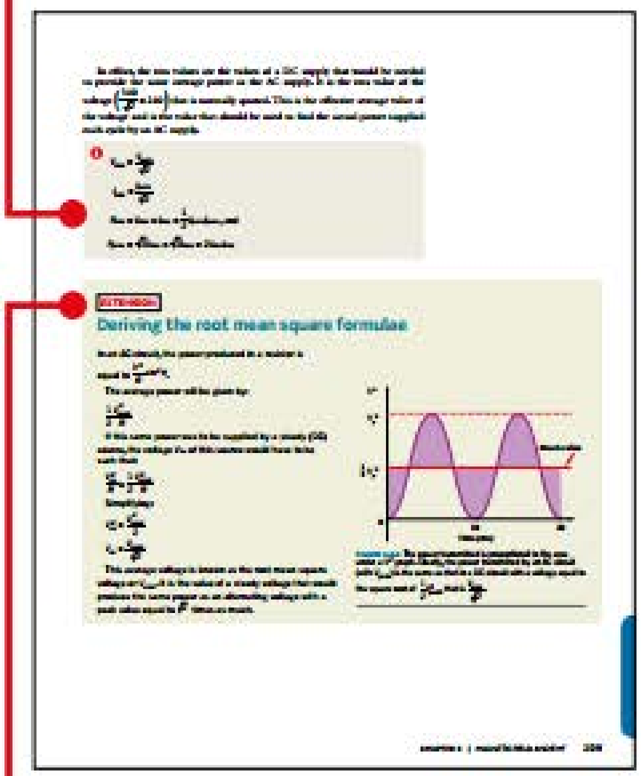






Highlight box

Focuses students' attention on important information such as key definitions, formulae and summary points.



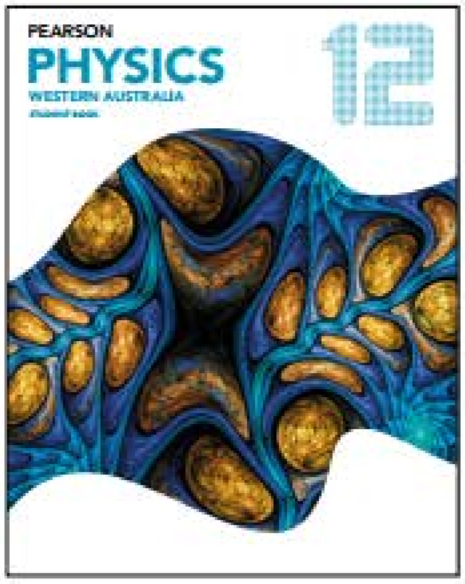
EXTENSION

Extension boxes include material that goes beyond the core content of the syllabus. They are intended for students who wish to expand their depth of understanding in a particular area.

Answers

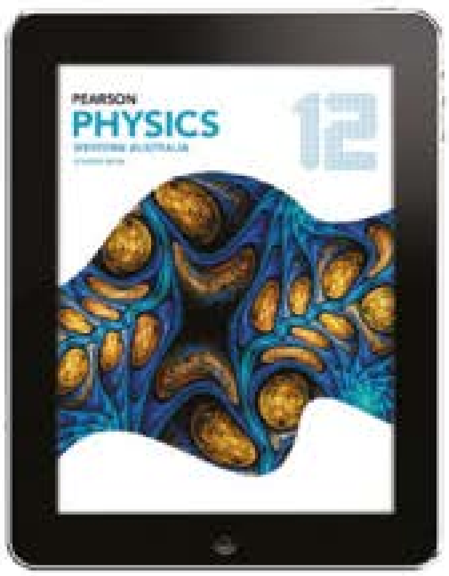
Numerical answers and key short response answers are included at the back of the book. Comprehensive answers and fully worked solutions for all section review questions, Try yourself: Worked examples, chapter review questions and Unit review questions are provided via *Pearson Physics 12 Western Australia Reader+* and *Teacher Resource*

Pearson Physics 12 Western Australia



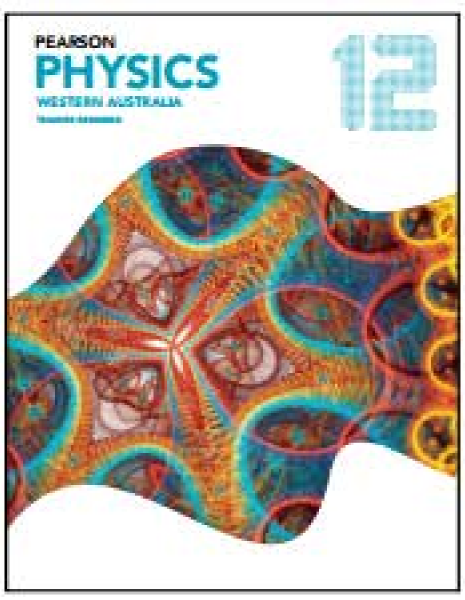
Student Book

Pearson Physics 12 Western Australia has been written to fully align with the *WACE Physics ATAR Course, Year 12 Syllabus 2017*. The series includes the very latest developments and applications of physics and incorporates best practice literacy and instructional design to ensure the content and concepts are fully accessible to all students.



Reader+

Pearson Reader+ lets you use the Student Book online or offline on any device. Pearson Reader+ includes interactive activities and videos to enhance learning and test understanding. The teacher version includes support material to assist planning and implementing the syllabus.



Teacher Resource

The Pearson Physics 12 Western Australia Teacher Resource consists of a print book and online resources, and provides comprehensive teacher support including teaching programs, fully worked solutions to all questions and chapter tests and practice exams.

PearsonDigital

Browse and buy at pearson.com.au
Access the ProductLink at pearsonplaces.com.au

UNIT 3 Gravity and electromagnetism

Field theories have enabled physicists to explain a vast array of natural phenomena and have contributed to the development of technologies that have changed the world, including electrical power generation and distribution systems, artificial satellites and modern communication systems. In this unit, students develop a deeper understanding of motion and its causes by using Newton's Laws of Motion and the gravitational field model to analyse motion on inclined planes, the motion of projectiles, and satellite motion. They investigate electromagnetic interactions and apply this knowledge to understand the operation of direct current motors, direct current (DC) and alternating current (AC) generators, transformers, and AC power distribution systems. Students also investigate the production of electromagnetic waves.

Learning outcomes

By the end of this unit, students:

- understand that motion in gravitational, electric and magnetic fields can be explained using Newton's Laws of Motion
- understand how the electromagnetic wave model explains the production and propagation of electromagnetic waves across the electromagnetic spectrum
- understand transformations and transfer of energy in electromagnetic devices, as well as transformations and transfer of energy associated with motion in gravitational, electric and magnetic fields
- understand how models and theories have developed over time, and the ways in which physical science knowledge and associated technologies interact with social, economic, cultural and ethical considerations
- use science inquiry skills to design, conduct, analyse and evaluate investigations into uniform circular motion, projectile motion, satellite motion and gravitational and electromagnetic phenomena, and to communicate methods and findings
- use algebraic and graphical representations to calculate, analyse and predict measurable quantities related to motion, gravitational effects and electromagnetic phenomena
- evaluate, with reference to evidence, claims about motion, gravity and electromagnetic phenomena and associated technologies, and justify evaluations
- communicate physics understanding using qualitative and quantitative representations in appropriate modes and genres.

CHAPTER

01 The force due to gravity

Gravity is the force that drives the universe. It was gravity that first caused particles to coalesce into atoms, and atoms to congregate into nebulas, planets and stars. An understanding of gravity is fundamental to understanding the universe.

This chapter centres on Newton's law of universal gravitation. This will be used to predict the size of the force experienced by an object at various locations on the Earth and other planets. It will also be used to develop the idea of a gravitational field and to refine the understanding of gravitational potential energy developed in Year 11. Since the field concept is also used to describe other basic forces such as electromagnetism and the strong and weak nuclear forces, this will provide an important foundation for further areas of study later in this unit.

Science as a Human Endeavour

Artificial satellites are used for communication, navigation, remote-sensing and research. Their orbits and uses are classified by altitude (low, medium or high Earth orbits) and by inclination (equatorial, polar and sun-synchronous orbits). Communication via satellite is now used for global positioning systems (GPS), satellite phones and television. Navigation services support management and monitoring of traffic and aircraft movement. Geographic information science uses data from satellites to monitor population movement, biodiversity and ocean currents.

Science Understanding

- all objects with mass attract one another with a gravitational force; the magnitude of gravitational force, F_g, can be calculated using Newton's Law of Universal Gravitation
  This includes applying the relationship

  $$F_g = G\frac{m_1 m_2}{r^2}$$

- gravitational field strength is defined as the net force per unit mass at a particular point in the field
  This includes applying the relationships

  $$g = \frac{F_g}{m} = G\frac{M}{r^2}$$

- the movement of free-falling bodies in Earth's gravitational field is predictable
- objects with mass produce a gravitational field in the space that surrounds them; field theory attributes the gravitational force on an object to the presence of a gravitational field
  This includes applying the relationship

  $$F_{weight} = mg$$

- when a mass moves or is moved from one point to another in a gravitational field and its potential energy changes, work is done on the mass by the field
  This includes applying the relationships

  $$E_p = mg\Delta h,\ W = Fs,\ W = \Delta E,\ E_k = \frac{1}{2}mv^2$$

1.1 Newton's law of universal gravitation

FIGURE 1.1.1 Sir Isaac Newton was one of the most influential physicists who ever lived.

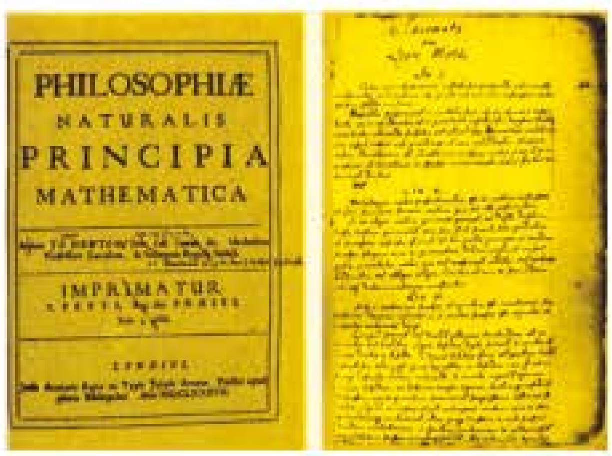


FIGURE 1.1.2 The *Principia* is one of the most influential books in the history of science.

In 1687, Sir Isaac Newton (Figure 1.1.1) published a book that changed the world. Entitled *Philosophiæ Naturalis Prinicipia Mathematica* (Mathematical Principles of Natural Philosophy), Newton's book (Figure 1.1.2) used a new form of mathematics, now known as calculus, and outlined his famous laws of motion.

The *Principia* also introduced Newton's law of universal gravitation. This was particularly significant because, for the first time in history, it was possible to scientifically explain the motion of the planets. This led to a change in humanity's understanding of its place in the universe.

UNIVERSAL GRAVITATION

Newton's law of universal gravitation states that any two bodies in the universe attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

As the radius, r, appears in the denominator of Newton's law of universal gravitation, the relationship indicates an inverse relationship. Since r is also squared, this relationship is known as an **inverse square law**. The implication is that as r increases, F_g will decrease dramatically. This law will reappear again later in the chapter when gravitational fields are examined in detail.

Mathematically, Newton's law of universal gravitation can be expressed as:

$$F_g = G\frac{m_1 m_2}{r^2}$$

where F_g is the **gravitational force (N)**
m_1 is the mass of object 1 (kg)
m_2 is the mass of object 2 (kg)
r is the distance between the centres of m_1 and m_2 (m)
G is the gravitational constant, 6.67×10^{-11} N m^2 kg^{-2}

PHYSICS IN ACTION

Measuring the gravitational constant, G

The **gravitational constant**, G, was first accurately measured by the British scientist Henry Cavendish in 1798, over a century after Newton's death. Cavendish used a **torsion balance** (Figure 1.1.3), a device that can measure very small twisting forces. Cavendish's experiment could measure forces smaller than 1 μN (i.e. 10^{-6} N). He used this balance to measure the force of attraction between lead balls held a small distance apart. Once the size of the force was known for a given combination of masses at a known separation distance, a value for G could be determined.

FIGURE 1.1.3 Henry Cavendish used a torsion balance to measure the small twisting force created by the gravitational attraction of lead balls. A very similar method is still in use in introductory labs in senior secondary schools and universities today.

The law of universal gravitation predicts that *any* two objects that have mass *will attract each other*. However, because the value of G is so small, the **gravitational force** between two everyday objects, such as you and the person seated next to you, is too small to be noticed.

Worked example 1.1.1

GRAVITATIONAL ATTRACTION BETWEEN SMALL OBJECTS

A man with a mass of 90 kg and a woman with a mass of 75 kg have a distance of 80 cm between their centres. Calculate the force of gravitational attraction between them.	
Thinking	**Working**
Recall the formula for Newton's law of universal gravitation.	$F_g = G\frac{m_1 m_2}{r^2}$
Identify the information required, and convert values into appropriate units when necessary.	$G = 6.67 \times 10^{-11}\,\text{N m}^2\,\text{kg}^{-2}$ $m_1 = 90\,\text{kg}$ $m_2 = 75\,\text{kg}$ $r = 80\,\text{cm} = 0.80\,\text{m}$
Substitute the values into the equation.	$F_g = 6.67 \times 10^{-11} \times \frac{90 \times 75}{0.80^2}$
Solve the equation.	$F_g = 7.0 \times 10^{-7}\,\text{N}$

Worked example: Try yourself 1.1.1

GRAVITATIONAL ATTRACTION BETWEEN SMALL OBJECTS

> Two bowling balls are sitting next to each other on a shelf so that the centres of the balls are 60 cm apart. Ball 1 has a mass of 7.0 kg and ball 2 has a mass of 5.5 kg. Calculate the force of gravitational attraction between them.

GRAVITATIONAL ATTRACTION BETWEEN MASSIVE OBJECTS

Gravitational forces between everyday objects are so small (as seen in Worked example 1.1.1) that they are hard to detect without specialised equipment, and can usually be considered as negligible.

For the gravitational force to become significant, at least one of the objects must have a very large mass—for example, our planet Earth (Figure 1.1.4).

FIGURE 1.1.4 Gravitational forces become significant when at least one of the objects has a large mass. Both the Earth and the Moon have significant mass.

Worked example 1.1.2

GRAVITATIONAL ATTRACTION BETWEEN MASSIVE OBJECTS

Calculate the force of gravitational attraction between the Sun and the Earth, given the following data: $m_{Sun} = 2.0 \times 10^{30}\,\text{kg}$ $m_{Earth} = 6.0 \times 10^{24}\,\text{kg}$ $r_{Sun-Earth} = 1.5 \times 10^{11}\,\text{m}$	
Thinking	**Working**
Recall the formula for Newton's law of universal gravitation.	$F_g = G\frac{m_1 m_2}{r^2}$

Identify the information required.	$G = 6.67 \times 10^{-11}\,N\,m^2\,kg^{-2}$ $m_1 = 2.0 \times 10^{30}\,kg$ $m_2 = 6.0 \times 10^{24}\,kg$ $r = 1.5 \times 10^{11}\,m$
Substitute the values into the equation.	$F_g = 6.67 \times 10^{-11} \times \frac{2.0 \times 10^{30} \times 6.0 \times 10^{24}}{(1.5 \times 10^{11})^2}$
Solve the equation.	$F_g = 3.6 \times 10^{22}\,N$

Worked example: Try yourself 1.1.2

GRAVITATIONAL ATTRACTION BETWEEN MASSIVE OBJECTS

Calculate the force of gravitational attraction between the Earth and the Moon, given the following data:
$m_{Earth} = 6.0 \times 10^{24}\,kg$
$m_{Moon} = 7.3 \times 10^{22}\,kg$
$r_{Moon-Earth} = 3.8 \times 10^{8}\,m$

The force in Worked example 1.1.2 is much greater than that in Worked example 1.1.1, illustrating the difference in the gravitational force when at least one of the objects has a much larger mass.

EXTENSION

Understanding the structure of the solar system

In the century before Newton, there had been some controversy about the structure of the solar system. In 1543, the commonly accepted geocentric (i.e. Earth-centred) model of the solar system had been challenged by the Polish astronomer Nicolaus Copernicus. He proposed that the Sun was the centre of the solar system. Unfortunately, some faulty assumptions meant that the predictions of Copernicus' Sun-centred or heliocentric model (shown in Figure 1.1.5) did not match observations any better than the geocentric model.

The Danish astronomer Tycho Brahe had been observing and studying the heavens for many years, accumulating a comprehensive collection of data. Using Brahe's documentation, his assistant, German mathematician Johannes Kepler, refined the Copernican model to reflect actual observations.

Through these calculations, Kepler discovered that the orbits of the planets around the Sun are elliptical and not circular as previously thought (Figure 1.1.6). At the time, this discovery challenged conventional beliefs about the 'perfection' of heavenly bodies, and, as a consequence, Kepler's ideas were not widely accepted. In fact, in some countries his books were banned and publicly burned.

One of Newton's great achievements was that he was able to use his law of universal gravitation to mathematically derive all of Kepler's planetary laws.

This allowed Newton to accurately explain the motion of the planets in terms of gravitational attraction. Within a few years of the publication of Newton's work, the geocentric model of our solar system had largely been abandoned in favour of the heliocentric model.

FIGURE 1.1.5 Nicolaus Copernicus proposed a heliocentric model of the solar system.

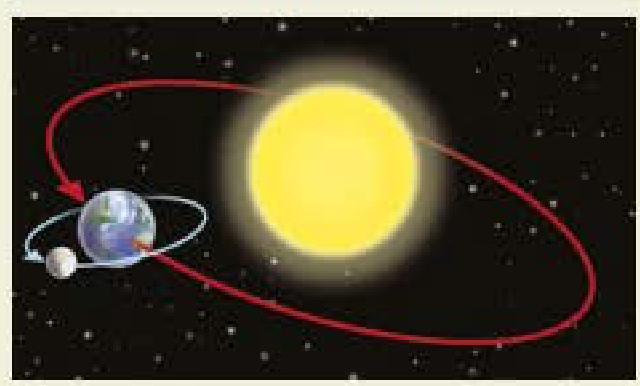


FIGURE 1.1.6 Johannes Kepler discovered that the orbits of planets around the Sun are elliptical.

EFFECT OF GRAVITY

According to Newton's third law of motion, forces occur in action–reaction pairs. An example of such a pair is shown in Figure 1.1.7. The Earth exerts a gravitational force on the Moon and, conversely, the Moon exerts an equal and opposite force on the Earth. Using Newton's second law of motion, you can see that the effect of the gravitational force of the Moon on the Earth will be much smaller than the corresponding effect of the Earth on the Moon. This is because of the Earth's larger mass.

FIGURE 1.1.7 The Earth and Moon exert gravitational forces on each other.

Worked example 1.1.3

ACCELERATION CAUSED BY A GRAVITATIONAL FORCE

The force of gravitational attraction between the Moon and the Earth is approximately 2.0×10^{20} N. Calculate the acceleration of the Earth and the Moon caused by this force. Compare these accelerations by calculating the ratio $\frac{a_{Moon}}{a_{Earth}}$.

Use the following data:

$m_{Earth} = 6.0 \times 10^{24}$ kg

$m_{Moon} = 7.3 \times 10^{22}$ kg

Thinking	Working
Recall the formula for Newton's second law of motion.	$F = ma$
Transpose the equation to make *a* the subject.	$a = \frac{F}{m}$
Substitute values into this equation to find the accelerations of the Moon and the Earth.	$a_{Earth} = \frac{2.0 \times 10^{20}}{6.0 \times 10^{24}} = 3.3 \times 10^{-5}\,\text{m s}^{-2}$ $a_{Moon} = \frac{2.0 \times 10^{20}}{7.3 \times 10^{22}} = 2.7 \times 10^{-3}\,\text{m s}^{-2}$
Compare the two accelerations.	$\frac{a_{Moon}}{a_{Earth}} = \frac{2.7 \times 10^{-3}}{3.3 \times 10^{-5}} = 82$ The acceleration of the Moon is 82 times greater than the acceleration of the Earth.

Worked example: Try yourself 1.1.3

ACCELERATION CAUSED BY A GRAVITATIONAL FORCE

The force of gravitational attraction between the Sun and the Earth is approximately 3.6×10^{22} N. Calculate the acceleration of the Earth and the Sun caused by this force. Compare these accelerations by calculating the ratio $\frac{a_{Earth}}{a_{Sun}}$.

Use the following data:

$m_{Earth} = 6.0 \times 10^{24}$ kg

$m_{Sun} = 2.0 \times 10^{30}$ kg

PHYSICSFILE

Extrasolar planets

In recent years, scientists have been interested in discovering whether other stars have planets like those in our own solar system. One of the ways in which these 'extrasolar planets' (or 'exoplanets') can be detected is from their gravitational effect.

When a large planet (Jupiter-sized or larger) orbits a star, it causes the star to wobble. This causes variations in the star's appearance, which can be detected on Earth. Hundreds of exoplanets have been discovered using this technique, including one orbiting the closest star to our Sun, Proxima Centauri, in the potentially habitable zone where temperatures are similar to those on Earth.

Gravity in the solar system

Although the accelerations caused by gravitational forces in Worked example 1.1.3 are small, over billions of years they have created the motion of the solar system.

In the Earth–Moon system, the acceleration of the Moon is many times greater than that of the Earth, which is why the Moon orbits the Earth. Although the Moon's gravitational force causes a much smaller acceleration of the Earth, it does have other significant effects, such as causing the tides.

Similarly, the Earth and other planets orbit the Sun because their masses are much smaller than the Sun's mass. However, the combined gravitational effect of the planets of the solar system (and Jupiter in particular) causes the Sun to wobble slightly as the planets orbit it.

WEIGHT AND GRAVITATIONAL FORCE

In Year 11 Physics the **weight** of an object was calculated using the formula $F_{weight} = mg$. Weight is another name for the gravitational force acting on an object near the Earth's surface.

Worked example 1.1.4 below shows that the formula $F_{weight} = mg$ and Newton's law of universal gravitation give the same answer for the gravitational force acting on objects on the Earth's surface. It is important to note that the distance used in these calculations is the distance between the centres of the two objects, which is effectively the radius of the Earth.

Worked example 1.1.4

GRAVITATIONAL FORCE AND WEIGHT

Compare the weight of an 80 kg person calculated using $F_{weight} = mg$ with the gravitational force calculated using $F_g = G\frac{m_1m_2}{r^2}$.

Use the following dimensions of the Earth in your calculations:

$g = 9.80\,m\,s^{-2}$
$m_{Earth} = 6.0 \times 10^{24}\,kg$
$r_{Earth} = 6.4 \times 10^6\,m$

Thinking	Working
Apply the weight equation.	$F_{weight} = mg$ $= 80 \times 9.80$ $= 784\,N$ $= 780\,N$ (to two significant figures)
Apply Newton's law of universal gravitation.	$F_g = G\frac{m_1m_2}{r^2}$ $= 6.67 \times 10^{-11} \times \frac{6.0 \times 10^{24} \times 80}{(6.4 \times 10^6)^2}$ $= 780\,N$
Compare the two values.	The equations give the same result to two significant figures.

Worked example: Try yourself 1.1.4

GRAVITATIONAL FORCE AND WEIGHT

Compare the weight of a 1.0 kg mass on the Earth's surface calculated using the formulae $F_{weight} = mg$ and $F_g = G\frac{m_1m_2}{r^2}$. Use the following dimensions of the Earth where necessary:
$g = 9.80\,m\,s^{-2}$
$m_{Earth} = 6.0 \times 10^{24}\,kg$
$r_{Earth} = 6.4 \times 10^6\,m$

Worked example 1.1.4 shows that the constant for the **acceleration due to gravity**, g, can be derived directly from the dimensions of the Earth. An object with mass m sitting on the surface of the Earth is a distance of 6.4×10^6 m from the centre of the Earth.

Given that the Earth has a mass of 6.0×10^{24} kg, then:

$$F_{\text{weight}} = F_g$$

$$\therefore\ mg = G\frac{m_{\text{Earth}}m}{(r_{\text{Earth}})^2}$$

$$= mG\frac{m_{\text{Earth}}}{(r_{\text{Earth}})^2}$$

$$\therefore\ g = G\frac{m_{\text{Earth}}}{(r_{\text{Earth}})^2}$$

$$= 6.67\times10^{-11}\times\frac{6.0\times10^{24}}{(6.4\times10^6)^2}$$

$$= 9.8\,\text{m s}^{-2}$$

EXTENSION

Multi-body systems

So far, only gravitational systems involving two objects have been considered, such as the Moon and the Earth. In reality, all objects experience gravitational force from every other object around them. Usually, most of these forces are negligible and only the gravitational effect of the largest object nearby (i.e. the Earth) needs to be considered.

When there is more than one significant gravitational force acting on a body, the gravitational forces must be added together as vectors to determine the net gravitational force (Figure 1.1.8).

The direction and relative magnitude of the net gravitational force in a multi-body system depends entirely on the masses and the positions of the attracting objects (i.e. m_1, m_2 and m_3 in Figure 1.1.8).

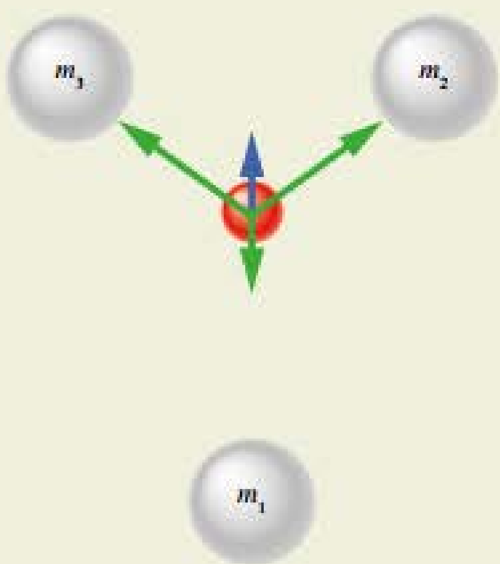


FIGURE 1.1.8 For the three masses $m_1 = m_2 = m_3$, the gravitational forces acting on the central red ball are shown by the green arrows. The vector sum of the green arrows is shown by the blue arrow. This will be the direction of the net (or resultant) gravitational force on the red ball due to the other three masses.

The rate of acceleration of objects near the surface of the Earth is a result of the Earth's mass and radius. A planet with a different mass and/or different radius will have a different value for g. Likewise, if an object is above the Earth's surface, the value of r will be greater and the value of g will be smaller (due to the inverse square law). This explains why the strength of the Earth's gravity reduces as you travel away from the Earth.

APPARENT WEIGHT

Scientists use the term 'weight' simply to mean 'the force due to gravity'. It is also correct to interpret weight as the contact force (or **normal reaction force**) between an object and the Earth's surface. In most situations these two definitions are effectively the same. However, there are some cases, for example when a person is accelerating up or down in an elevator, where they give different results. In these situations, the normal force (F_N) is referred to as the **apparent weight** since you do not feel the force you apply to the floor, you will only experience with your senses the forces that are applied on *you*. What you feel is the normal force acting up on you from the floor. Normally, when you stand on a surface that is either stationary or in constant vertical motion, your apparent weight is constant and equal to your weight force (Figure 1.1.9).

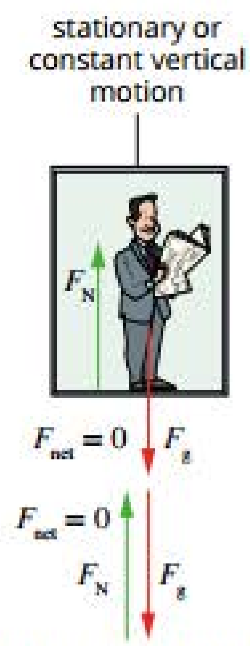


FIGURE 1.1.9 In this case, the forces that act on the person, F_N and F_g, are equal in size. The person will 'feel' his or her normal apparent weight.

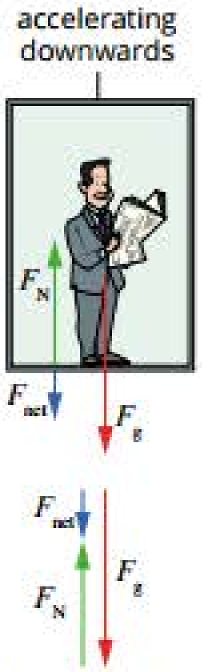


FIGURE 1.1.10 In this case, the forces that act on the person in the lift cause him to feel lighter than his normal apparent weight. When accelerating downwards, $F_N < F_g$.

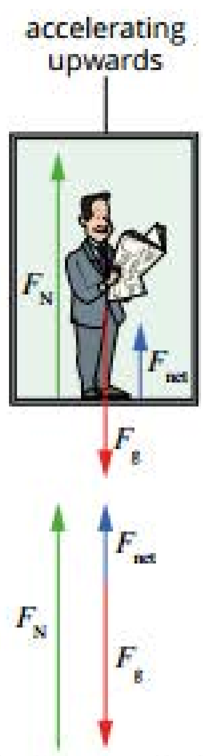


FIGURE 1.1.11 In this case, the forces that act on the person in the lift cause him to feel heavier than his normal apparent weight. When accelerating upwards, $F_N > F_g$.

The apparent weight that you experience changes when the surface you are standing on is accelerating upwards or downwards. If the floor is accelerating downwards at a rate less than 9.80 m s^{-2}, your feet will be pressing less firmly on the surface than when the floor was not accelerating. Therefore, the normal force is also less and so your apparent weight appears to be less. That is, you would feel lighter than usual (Figure 1.1.10).

The opposite happens when the floor is accelerating upwards. In this case, the floor is pushing up against your feet with a greater force than the normal reaction force due to your weight alone. The upwards push of the floor must provide the force to accelerate you upwards. This accelerating force adds to the normal force to make it appear that your apparent weight is greater than it would be if you weren't accelerating. That is, you would feel heavier than usual (Figure 1.1.11).

The normal reaction force (felt as apparent weight) and the force due to gravity (weight force) add as vectors to give the net force that causes the acceleration.

$$F_{net} = F_N + F_{weight}$$

where F_N is the apparent weight force that acts upwards on your feet

F_{weight} is the weight force due to gravity (which never changes)

F_{net} is the net force causing the acceleration

Worked example 1.1.5

CALCULATING APPARENT WEIGHT

A 79.0 kg student rides a lift up to the top floor of an office block. During the journey, the lift accelerates upwards at 1.26 m s^{-2} before travelling at a constant velocity of 3.78 m s^{-1} and then finally decelerating at 1.89 m s^{-2}.

a Calculate the apparent weight of the student in the first part of the journey while accelerating upwards at 1.26 m s^{-2}.

Thinking	Working
Ensure that the variables are in their standard units.	$m = 79.0$ kg $a = 1.26$ m s^{-2} up $g = 9.80$ m s^{-2} down
Apply the sign and direction convention for motion in one dimension. Up is positive and down is negative.	$m = 79.0$ kg $a = +1.26$ m s^{-2} $g = -9.80$ m s^{-2}
Apply the equation for apparent weight (the normal force).	$F_{net} = F_N + F_g$ $F_N = F_{net} - F_g$ $= ma - mg$ $= (79.0 \times 1.26) - (79.0 \times -9.80)$ $= 99.54 + 774.2$ $= 874$ N

b Calculate the apparent weight of the student in the second part of the journey while travelling at a constant speed of 3.78 m s^{-1}.

Thinking	Working
Ensure that the variables are in their standard units.	$m = 79.0$ kg $a = 0$ m s^{-2} $g = 9.80$ m s^{-2} down

Apply the sign and direction convention for motion in one dimension. Up is positive and down is negative.	$m = 79.0\,\text{kg}$ $a = 0\,\text{m}\,\text{s}^{-2}$ $g = -9.80\,\text{m}\,\text{s}^{-2}$
Apply the equation for apparent weight (the normal force).	$F_{net} = F_N + F_{weight}$ $F_N = F_{net} - F_{weight}$ $= ma - mg$ $= (79.0 \times 0) - (79.0 \times -9.80)$ $= 0 + 774.2$ $= 774\,\text{N}$

c Calculate the apparent weight of the student in the last part of the journey while travelling upwards and decelerating at $1.89\,\text{m}\,\text{s}^{-2}$.

Thinking	Working
Ensure that the variables are in their standard units. Also consider that deceleration is a negative acceleration.	$m = 79.0\,\text{kg}$ $a = -1.89\,\text{m}\,\text{s}^{-2}$ up $g = 9.80\,\text{m}\,\text{s}^{-2}$ down
Apply the sign and direction convention for motion in one dimension. Up is positive and down is negative.	$m = 79.0\,\text{kg}$ $a = -1.89\,\text{m}\,\text{s}^{-2}$ $g = -9.80\,\text{m}\,\text{s}^{-2}$
Apply the equation for apparent weight (the normal force).	$F_{net} = F_N + F_{weight}$ $F_N = F_{net} - F_{weight}$ $= ma - mg$ $= (79.0 \times -1.89) - (79.0 \times -9.80)$ $= -149.3 + 774.2$ $= 625\,\text{N}$

Worked example: Try yourself 1.1.5

CALCULATING APPARENT WEIGHT

A 79.0 kg student rides a lift down from the top floor of an office block to the ground. During the journey the lift accelerates downwards at $2.35\,\text{m}\,\text{s}^{-2}$, before travelling at a constant velocity of $4.08\,\text{m}\,\text{s}^{-1}$ and then finally decelerating at $4.70\,\text{m}\,\text{s}^{-2}$.

a Calculate the apparent weight of the student in the first part of the journey while accelerating downwards at $2.35\,\text{m}\,\text{s}^{-2}$.

b Calculate the apparent weight of the student in the second part of the journey while travelling at a constant speed of $4.08\,\text{m}\,\text{s}^{-1}$.

c Calculate the apparent weight of the student in the last part of the journey while travelling downwards and decelerating at $4.70\,\text{m}\,\text{s}^{-2}$.

From Worked example 1.1.5, you can see that:

- when accelerating upwards the student will feel heavier than normal ($F_N > mg$) (Note: this is the same as decelerating while travelling downwards)
- when accelerating downwards, the student will feel lighter than normal ($F_N < mg$) (Note: this is the same as decelerating while travelling upwards)
- when travelling upwards or downwards at a constant velocity, the student will feel their normal weight, just as they would if the lift was stationary ($F_N = mg$).

Apparent weightlessness

Defining apparent weight makes it possible to identify the situations in which you will experience **apparent weightlessness**. Your apparent weight is a contact reaction force that acts upwards on you from a surface because gravity is pulling you down on that surface. So if you are not standing on a surface, you will experience zero apparent weight; that is, apparent weightlessness. This means that you will experience apparent weightlessness the moment you step off the top platform of a diving pool or as you skydive from a plane, although the rushing air will hardly let you experience the sensation of floating as you skydive.

Felix Baumgartner experienced apparent weightlessness as he fell from his balloon 39 km above the Earth (Figure 1.1.12). This vertical height is equivalent to the distance from Perth to the west end of Rottnest.

FIGURE 1.1.12 Felix Baumgartner experienced apparent weightlessness on his return to Earth from 39 000 m.

Astronauts also experience apparent weightlessness in the International Space Station, which orbits about 370 km above the surface of the Earth (about the horizontal distance from Perth to Albany).

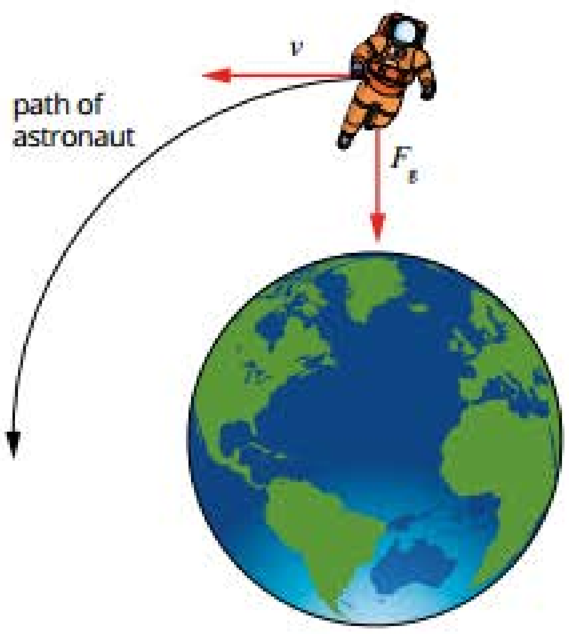


FIGURE 1.1.13 Astronauts are in free-fall while orbiting the Earth.

Whenever you are in **free-fall**, you experience apparent weightlessness. It follows then that whenever you experience apparent weightlessness, you must be in free-fall. When astronauts experience apparent weightlessness, they are not floating in space as they orbit the Earth. They are actually in free-fall. Astronauts and their spacecraft are both falling, but not directly towards the Earth like Baumgartner. The astronauts are actually moving horizontally, as shown in Figure 1.1.13. Baumgartner stayed approximately above the same place on the Earth from where he departed. Astronauts, on the other hand, are moving at a velocity relative to the Earth, so they are moving across the sky at the same time as they are falling. The combined effect is that they fall in a curved path that almost mirrors the curve of the Earth. So they fall, but continually miss the Earth as the surface of the Earth curves away from their path.

Importantly there is a significant difference between apparent weightlessness and true weightlessness. True weightlessness only occurs when the gravitational field strength is zero and hence $F_{weight} = 0$. This only occurs in deep space, far enough away from any planets so that their gravitational effect is zero. Apparent weightlessness, however, can occur when still under the influence of a gravitational field.

EXTENSION

Falling at constant speed

Galileo was able to show, more than 400 years ago, that the mass of a body does not affect the rate at which it falls towards the ground. However, our common experience is that not all objects behave in this way. A light object, such as a feather or a balloon, does not accelerate at 9.80 m s^{-2} as it falls. It drifts slowly to the ground, far slower than other dropped objects. Parachutists and skydivers also eventually fall with a constant speed. However, they can change their falling speed by changing their body profile, as pictured in Figure 1.1.14. If they assume a tuck position, they will fall faster, and if they spread out their arms and legs, they will fall slower. This enables them to form spectacular patterns as they fall.

FIGURE 1.1.14 Skydivers performing intricate manoeuvres in free-fall.

Skydivers, base-jumpers and air-surfers are able to use the force of air resistance to their advantage. As a skydiver first steps out of the plane, the forces acting on them are drag (air resistance), F_{ar}, and weight due to gravity, F_{weight}. Since their speed is low, the drag force is small, as shown in Figure 1.1.15(a). There is a large net force (F_{net}) downwards, so they will experience a large downwards acceleration of just less than 9.80 m s^{-2}, causing them to speed up. This causes the drag force to increase because they are colliding harder with the air molecules. In fact, the drag force increases in proportion to the square of the speed, $F_{ar} \propto v^2$. This results in a smaller net force downwards as shown in Figure 1.1.15(b). Their downwards acceleration is therefore reduced. (It is important to remember that they are still speeding up, but at a reduced rate.)

As their speed continues to increase, so too does the magnitude of the drag force. Eventually, the drag force becomes as large as the weight force due to gravity, as shown in Figure 1.1.15(c). When this happens, the net force is zero and the skydiver will fall with a constant velocity. As the velocity is now constant, the drag force will also remain constant and the motion of the skydiver will not change, as shown in Figure 1.1.15(d). This velocity is commonly known as the *terminal velocity*.

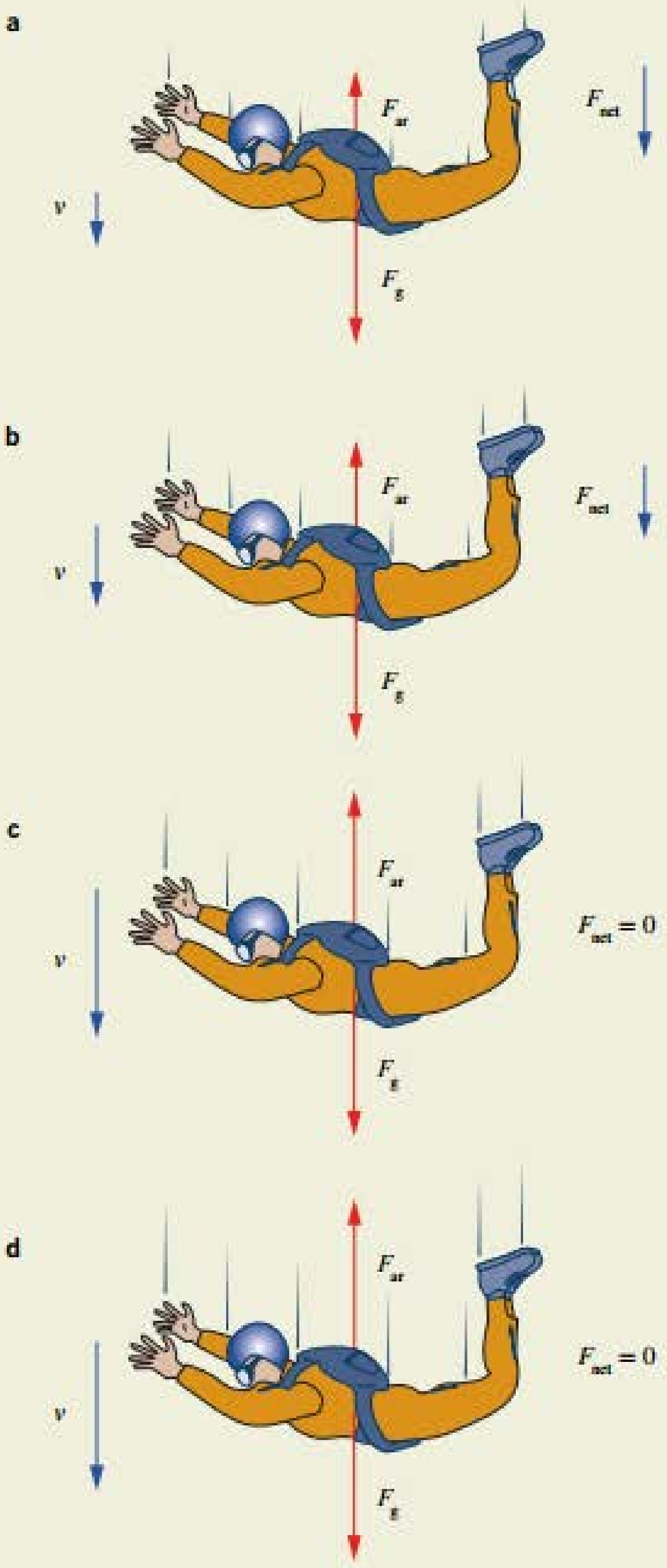


FIGURE 1.1.15 The forces involved in skydiving.

PHYSICS IN ACTION

Satellites

Natural satellites have existed throughout the universe for billions of years. The planets and asteroids of the solar system are natural satellites of the Sun.

Earth has one natural satellite: the Moon. The largest planets—Jupiter and Saturn—each have more than sixty natural satellites in orbit around them. Most of the stars in the Milky Way galaxy have planets and more of these exoplanets are being discovered each year.

Since the space age began in 1957 with the launch of Sputnik, about 6000 **artificial satellites** have been launched into orbit around the Earth. Today there are about 4000 artificial satellites still in orbit, although only about 1200 of these are operational.

Satellites in orbit around the Earth are classified as being in low, medium or high orbit.

- Low orbit: 180 km to 2000 km **altitude**. Most satellites orbit in this range (an example is shown in Figure 1.1.16). These include the Hubble Space Telescope, which is used by astronomers to view objects right at the edge of the universe.
- Medium orbit: 2000 km to 36 000 km altitude. The most common satellites in this region are the global positioning system (GPS) satellites used to run navigation systems.
- High orbit: 36 000 km altitude or greater. Australia uses the Optus satellites for communications, and deep-space weather pictures come from the Japanese Himawari-8 satellite. The satellites that sit at an altitude of 36 000 km and orbit with a period of 24 hours are known as **geostationary satellites** (or geosynchronous satellites). Most communications satellites are geostationary.

Earth satellites can have different orbital paths depending on their function:

- equatorial orbits in which the satellite always travels above the equator
- polar or near-polar orbits in which the satellite travels over or close to the North and South poles as it orbits
- inclined orbits, which lie between equatorial and polar orbits.

Satellites are used for a multitude of different purposes, with 60% used for communications. Many low-orbit American NOAA satellites have an inclination of 99° and an orbit that allows them to pass over each part of the Earth at the same time each day. These satellites are also known as Sun-synchronous satellites.

Artificial and natural satellites are not propelled by rockets or engines. They orbit in free-fall, and the only force acting on them is the gravitational attraction between the satellite and the body about which it orbits. This means that the satellites have a **centripetal acceleration** that is equal to the gravitational field strength at their location (Figure 1.1.18). Centripetal acceleration is covered in more detail in Chapter 2 'Motion in a gravitational field'.

Artificial satellites are often equipped with tanks of propellant that is squirted in the appropriate direction when the orbit of the satellite needs to be adjusted.

Artificial satellites of particular value to Australia

Geostationary meteorological satellites Himawari-8 and 9

The Himawari-8 satellite was launched from the Tanegashima Space Centre, Japan, on 7 October 2014, and orbits at approximately 35 786 km directly over the equator in a geostationary orbit of 24 hours. At its closest point to the Earth (perigee), its altitude is 35 784 km. At its furthest point from the Earth (apogee), it is at 35 789 km. Himawari-8 orbits at a longitude of around 140.7°E, so it is just to the north of Cape York. At this position it is ideally located for use by Australian forecasters.

FIGURE 1.1.16 A low-orbit satellite called the Soil Moisture and Ocean Salinity (SMOS) probe was launched in August 2014. Its role is to measure water movements and salinity levels on Earth as a way of monitoring climate change. It was launched from northern Russia by the European Space Agency (ESA).

Scans from Himawari-8 are made every 10 minutes and transmitted in near real-time to a satellite dish on the roof of the head office at the Bureau of Meteorology in Perth. Himawari-8 is the first geostationary weather satellite to take true-colour pictures at a much greater resolution than previous satellites. 16 different types of image show the temperature variations in the atmosphere and are invaluable in weather forecasting. Himawari-8 is box-like and measures about 2.6 m along each side. It had a mass of 3500 kg when it was launched, and is powered by solar panels that, when deployed, take its overall length to approximately 8 m. Himawari-9 was launched on 2 November 2016 and is positioned very near to Himawari-8. Its role is to remain on standby until 2022, when Himawari-8 will switch from observing to standing-by while Himawari-9 takes over.

Hubble Space Telescope (HST)

This cooperative venture between NASA and the European Space Agency (ESA) was launched by the crew of the space shuttle Discovery on 25 April 1990. The HST is a permanent unoccupied space-based observatory with a 2.4 m-diameter reflecting telescope, spectrographs and a faint-object camera. It orbits above the Earth's atmosphere, producing images of distant stars and galaxies far clearer than those from ground-based observatories (Figure 1.1.17). The HST is in a low-Earth orbit inclined at 28° to the equator. Its expected life span was originally around 15 years, but service and repair missions have extended its life and it is still in use today.

FIGURE 1.1.17 In August 2014, astronomers used the Hubble Space Telescope to detect the blue companion star of a white dwarf in a distant galaxy. The white dwarf slowly siphoned fuel from its companion, eventually igniting a runaway nuclear reaction in the compact star, which produced a supernova blast.

National Oceanic and Atmospheric Administration Satellite (NOAA-19)

Many of the US-owned and operated NOAA satellites are located in low-altitude near-polar orbits. This means that they pass close to the poles of the Earth as they orbit. NOAA-19 was launched in February 2009 and orbits at an inclination of 99° to the equator. Its low altitude means that it captures high-resolution pictures of small bands of the Earth. The data is used in local weather forecasting as well as to provide enormous amounts of information for monitoring global warming and climate change.

Table 1.1.1 provides data for the three satellites discussed in this section.

Seeing the International Space Station (ISS) and other satellites

It is easy to see low-orbit satellites if you are away from city lights. The best time to look is just after sunset. If you can, go outside and look for any slow-moving objects passing across the star background.

There are also many websites that will allow you to track and predict the real-time paths of satellites. You can use the NASA 'Spot the Station' website to see when the ISS is passing over your part of the planet. The ISS is so bright that it is easy to see from most locations.

FIGURE 1.1.18 The only force acting on these artificial and natural satellites is the gravitational attraction of the Earth. Both orbit with a centripetal acceleration equal to the gravitational field strength at their locations.

TABLE 1.1.1 A comparison of the three satellites discussed in this section

Satellite	Orbit	Inclination	Perigee (km)	Apogee (km)	Period
Himawari-8	equatorial	0°	35 784	35 789	1 day
Hubble	inclined	28°	591	599	96.6 min
NOAA-19	near polar	99°	846	866	102 min

1.1 Review

SUMMARY

- All objects with mass attract one another with a gravitational force.
- The gravitational force acts equally on each of the masses.
- The magnitude of the gravitational force is given by Newton's law of universal gravitation:

  $$F_g = G\frac{m_1 m_2}{r^2}$$

- Gravitational forces are usually negligible unless one of the objects is massive, e.g. a planet.
- The weight of an object on the Earth's surface is due to the gravitational attraction of the Earth, i.e. weight = F_{weight}.
- The acceleration due to gravity of an object near the Earth's surface can be calculated using the dimensions of the Earth:

  $$g = G\frac{m_{Earth}}{(r_{Earth})^2} = 9.80\,\text{m s}^{-2}$$

- Objects can have an apparent weight that is greater or less than their normal weight. This occurs when they are accelerating vertically.
- Artificial satellites are used for communication, navigation, remote-sensing and research. Their orbits and uses are classified by altitude (low-, medium- or high-Earth orbits) and by inclination (equatorial, polar and Sun-synchronous orbits).

KEY QUESTIONS

1 Newton's law of universal gravitation links the relationships (or proportionalities) between several key factors influencing the force due to gravity between two objects. What are the individual proportionalities?

2 What does the symbol *r* represent in Newton's law of universal gravitation?

3 Calculate the force of gravitational attraction between the Sun and Mars, given the following data:

$m_{Sun} = 2.0 \times 10^{30}$ kg

$m_{Mars} = 6.4 \times 10^{23}$ kg

$r_{Sun-Mars} = 2.2 \times 10^{11}$ m

4 The force of gravitational attraction between the Sun and Mars is 1.8×10^{21} N. Calculate the acceleration of Mars given that $m_{Mars} = 6.4 \times 10^{23}$ kg.

5 On 14 April 2014, Mars came within 93 million km of Earth. Its gravitational effect on the Earth was the strongest it had been for over 6 years. Use the following data to answer the questions below.

$m_{Sun} = 2.0 \times 10^{30}$ kg

$m_{Earth} = 6.0 \times 10^{24}$ kg

$m_{Mars} = 6.4 \times 10^{23}$ kg

a Calculate the gravitational force between the Earth and Mars on 14 April 2014.

b Calculate the force of the Sun on the Earth if the distance between them was 153 million km.

c Compare your answers to parts (a) and (b) above by expressing the Mars–Earth force as a percentage of the Sun–Earth force.

6 The acceleration of the Moon caused by the gravitational force of the Earth is much larger than the acceleration of the Earth due to the gravitational force of the Moon. What is the reason for this?

7 Calculate the acceleration of an object dropped near the surface of Mercury if this planet has a mass of 3.3×10^{23} kg and a radius of 2500 km. Assume that the gravitational acceleration on Mercury can be calculated similarly to that on Earth.

8 Calculate the weight of a 65 kg cosmonaut standing on the surface of Mars, given that the planet has a mass of 6.4×10^{23} kg and a radius of 3.4×10^{6} m.

9 In your own words, explain the difference between the terms 'weight' and 'apparent weight', giving an example of a situation where the magnitudes of these two forces would be different.

10 Calculate the apparent weight of a 50 kg person in an elevator under the following circumstances.

a accelerating upwards at $1.2\,\text{m s}^{-2}$

b moving upwards at a constant speed of $5\,\text{m s}^{-1}$

11 Calculate the apparent weight of a 45.0 kg child standing in a lift that is decelerating at $3.15\,\text{m s}^{-2}$ while travelling upwards.

12 Which statement best describes the motion of astronauts when orbiting the Earth?

A They float in a zero-gravity environment.

B They float in a reduced gravity environment.

C They fall down very slowly due to the very small gravity.

D They fall in a reduced gravity environment.

13 Select the statement below that correctly describes how a satellite in a stable circular orbit 200 km above the Earth will move.

A It will have an acceleration of $9.80\,m\,s^{-2}$.

B It will have constant velocity.

C It will have zero acceleration.

D It will have acceleration of less than $9.80\,m\,s^{-2}$.

14 What can be said about an object if that object is orbiting the Earth in space and appears to be weightless?

A It is in free-fall.

B It is in zero gravity.

C It has no mass.

D It is floating.

15 A geostationary satellite orbits above Singapore, which is on the equator. Which of the following statements about the satellite is correct?

A It is in a low orbit.

B It is in a high orbit.

C It passes over the North Pole.

D It is not moving.

1.2 Gravitational fields

Newton's law of universal gravitation describes the force acting between two mutually attracting bodies. In reality, complex systems like the solar system involve a number of objects (i.e. the Sun and planets shown in Figure 1.2.1) that are all exerting attractive forces on each other at the same time.

In the 18th century, to simplify the process of calculating the effect of simultaneous gravitational forces, scientists developed a mental construct known as the gravitational field. In the following centuries, the idea of a **field** was also applied to other forces and has become a very important concept in physics.

FIGURE 1.2.1 The solar system is a complex gravitational system.

PHYSICS IN ACTION

Discovery of Neptune

The planet Neptune was discovered through its gravitational effect on other planets. Two astronomers, Urbain Le Verrier of France and John Couch Adams of England, each independently identified that the observed orbit of Uranus varied significantly from predictions made based on the gravitational effects of the Sun and other known planets. Both suggested that this was due to the influence of a distant, undiscovered planet.

Le Verrier sent a prediction of the location of the new planet to Gottfried Galle at the Berlin Observatory and, on 23 September 1846, Neptune was discovered within 1° of Le Verrier's prediction (Figure 1.2.2).

FIGURE 1.2.2 This star chart published in 1846 shows the location of Neptune in the constellation Aquarius when it was discovered on 23 September, and its location one week later.

GRAVITATIONAL FIELDS

A **gravitational field** is a region in which a gravitational force is exerted on all matter within that region. Every physical object has an accompanying gravitational field. For example, the space around your body contains a gravitational field because any other object that comes into this region will experience a (small) force of gravitational attraction to your body.

The gravitational field around a large object such as a planet is much more significant than that around a small object. The Earth's gravitational field exerts a significant influence on objects on its surface and even up to thousands of kilometres into space.

REPRESENTING GRAVITATIONAL FIELDS

Over time, scientists have developed a commonly understood method of representing fields using a series of arrows known as field lines (Figure 1.2.3). For gravitational fields, these are constructed as follows:

- The direction of the arrowhead indicates the direction of the gravitational force.
- The space between the field lines indicates the relative magnitude of the field.
  - Closely spaced field lines indicate a strong field.
  - Widely spaced field lines indicate a weaker field.
  - Parallel field lines indicate constant or **uniform** field strength.

An infinite number of field lines could be drawn, so only a few are chosen to represent the rest. The size of the gravitational force acting on a mass in the region of a gravitational field is determined by the strength of the field, and the force acts in the direction of the field.

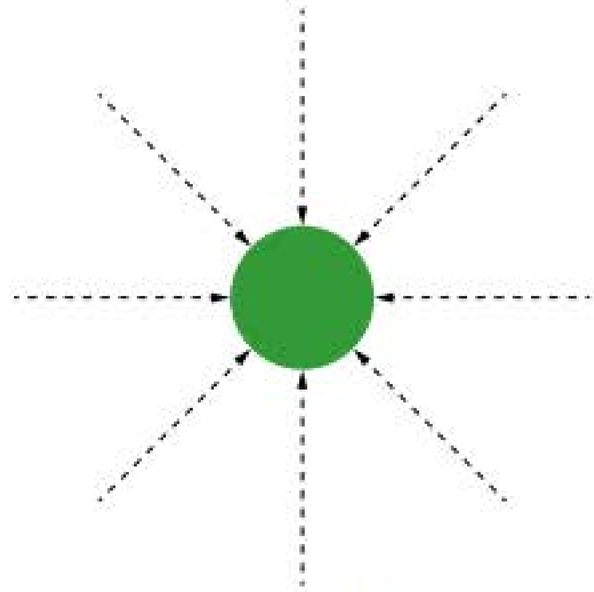
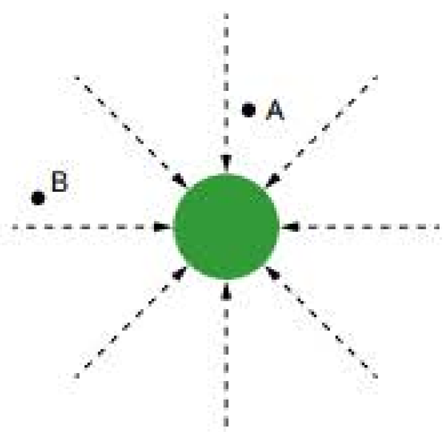


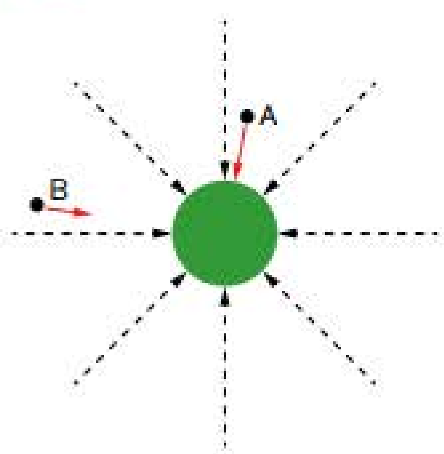
FIGURE 1.2.3 The arrows in this gravitational field diagram indicate that objects will be attracted towards the mass in the centre; the spacing of the lines shows that force will be strongest at the surface of the central mass and weaker further away from it.

Worked example 1.2.1

INTERPRETING GRAVITATIONAL FIELD DIAGRAMS

The diagram below shows the gravitational field of a moon.



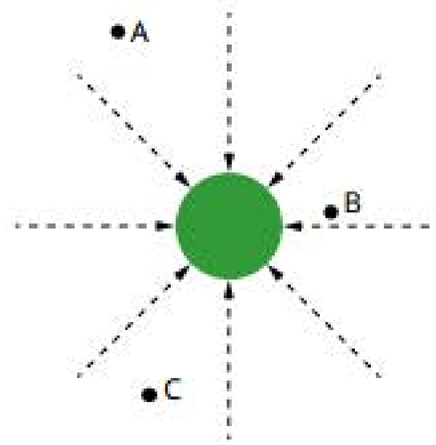
<table>
<tr><td colspan="2">a Use arrows to indicate the magnitude and direction of the gravitational force acting at points A and B.</td></tr>
<tr><td>Thinking</td><td>Working</td></tr>
<tr><td>The direction of the field arrows indicates the direction of the gravitational force, which is inwards towards the centre of the moon.</td><td></td></tr>
</table>

b Describe the relative strength of the gravitational field at each point.	
Thinking	**Working**
The closer the field lines, the stronger the force. The field lines are closer together at point A than they are at point B, as point A is closer to the moon.	The field is stronger at point A than at point B.

Worked example: Try yourself 1.2.1

INTERPRETING GRAVITATIONAL FIELD DIAGRAMS

The diagram below shows the gravitational field of a planet.



a Use arrows to indicate the magnitude and direction of the gravitational force acting at points A, B and C.

b Describe the relative strength of the gravitational field at each point.

GRAVITATIONAL FIELD STRENGTH

In theory, gravitational fields extend infinitely out into space. However, since the magnitude of the gravitational force decreases with the square of the distance, eventually these fields become so weak as to become negligible.

In Section 1.1, it was shown that it is possible to calculate the acceleration due to gravity of objects near the Earth's surface using the dimensions of the Earth:

$$g = G\frac{m_{\text{Earth}}}{(r_{\text{Earth}})^2} = 9.80\,\text{m}\,\text{s}^{-2}$$

The constant g can also be used as a measure of the strength of the gravitational field. When understood in this way, the constant is written with the equivalent units of $\text{N}\,\text{kg}^{-1}$ rather than $\text{m}\,\text{s}^{-2}$. This means $g_{\text{Earth}} = 9.80\,\text{N}\,\text{kg}^{-1}$.

These units indicate that objects on the surface of the Earth experience 9.80 N of gravitational force for every kilogram of their mass.

Accordingly, the familiar equation $F_{\text{weight}} = mg$ can be transposed so that the **gravitational field strength**, g, can be calculated.

$$g = \frac{F_{\text{weight}}}{m}$$

where g is gravitational field strength ($\text{N}\,\text{kg}^{-1}$)

F_{weight} is the force due to gravity (N)

m is the mass of an object in the field (kg)

PHYSICSFILE

$\text{N}\,\text{kg}^{-1} = \text{m}\,\text{s}^{-2}$

It can be shown that $\text{N}\,\text{kg}^{-1}$ and $\text{m}\,\text{s}^{-1}$ are equivalent units.

From Newton's second law, $F = ma$, you will remember that:

$$1\,\text{N} = 1\,\text{kg}\,\text{m}\,\text{s}^{-2}$$

$$\therefore\ 1\,\text{N}\,\text{kg}^{-1} = 1\,\text{kg}\,\text{m}\,\text{s}^{-2} \times \text{kg}^{-1} = 1\,\text{m}\,\text{s}^{-2}$$

Worked example 1.2.2

CALCULATING GRAVITATIONAL FIELD STRENGTH

When a student hangs a 1.0 kg mass from a spring balance, the balance measures a downwards force of 9.80 N. According to this experiment, what is the gravitational field strength of the Earth in this location?	
Thinking	**Working**
Recall the equation for gravitational field strength.	$g = \frac{F_{weight}}{m}$
Substitute in the appropriate values.	$g = \frac{9.8}{1.0}$
Solve the equation.	$g = 9.8\,\mathrm{N\,kg^{-1}}$

Worked example: Try yourself 1.2.2

CALCULATING GRAVITATIONAL FIELD STRENGTH

A student uses a spring balance to measure the weight of a piece of wood as 2.5 N.

If the piece of wood is thought to have a mass of 260 g, calculate the gravitational field strength indicated by this experiment.

Newton's law of universal gravitation can be written as $F_g = G\frac{Mm}{r^2}$, to indicate the difference in mass between the large central body and the smaller object within the gravitational field. This formula can be used to develop the formula for gravitational field strength:

$$g = \frac{F_{weight}}{m} = \frac{G\frac{Mm}{r^2}}{m}$$

This simplifies to:

$$g = G\frac{M}{r^2}$$

where g is the gravitational field strength ($\mathrm{N\,kg^{-1}}$)

G is the gravitational constant, $6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$

M is the mass of the planet or moon (the central body; kg)

r is the radius of the planet or moon (m)

REVISION

Inverse square law

The concept of a field is a very powerful tool for understanding forces that act at a distance. It has also been applied to forces such as the electrostatic force between charged objects and the force between two magnets.

The study of gravitational fields introduces the concept of the inverse square law. From the point source of a field, whether it be gravitational, electric or magnetic, the field will spread out radially in three dimensions. When the distance from the source is doubled, the field will be spread over four times the original area.

In Figure 1.2.4, going from r to $2r$ to $3r$, the area shown increases from one square to four squares (2^2) to nine squares (3^2). Using the inverse part of the inverse square law, at a distance $2r$ the strength of the field will be reduced to a quarter of that at r, as is the force that the field would exert. At $3r$ from the source, the field will be reduced to one-ninth of that at the source, and so on.

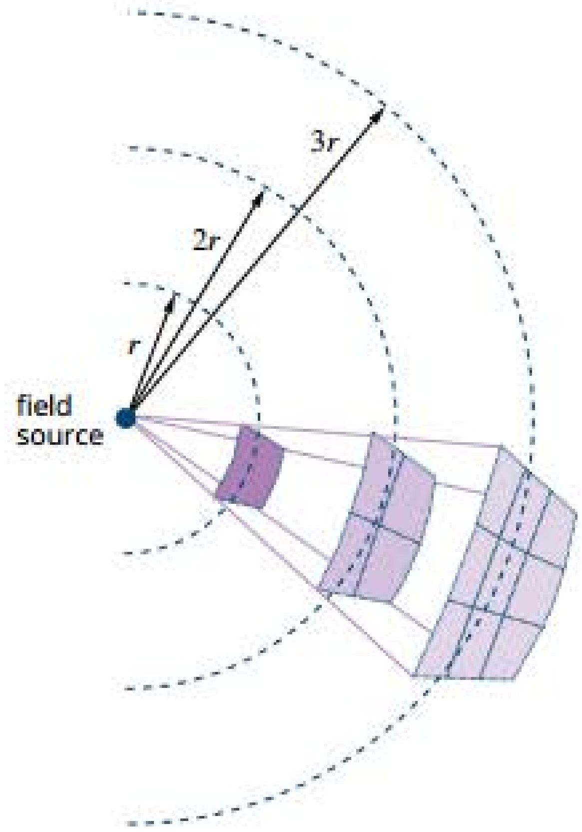


FIGURE 1.2.4 As the distance from the source of a field increases, the field is spread over an area that increases with the square of the distance from the source, resulting in the strength of the field decreasing by the same ratio.

In terms of the gravitational field, the strength of the force varies inversely with the square of the distance between the objects:

$$F \propto \frac{M}{r^2}$$

where F is the force

r is the distance from the source of the gravitational field.

This is referred to as the inverse square law.

One key difference between the gravitational force and other inverse square forces is that the gravitational force is always attractive, whereas like charges or magnetic poles repel one another.

Inverse square laws are an important concept in physics, not only in the study of fields but also for other phenomena where energy is moving away from its source in three dimensions, such as in sound and other waves.

Variations in gravitational field strength of the Earth

The gravitational field strength of the Earth, g, is usually assigned a value of $9.81\,\mathrm{N\,kg^{-1}}$ (and generally rounded further to $9.80\,\mathrm{N\,kg^{-1}}$ for Physics exams). However, the field strength experienced by objects on the surface of the Earth can vary between $9.76\,\mathrm{N\,kg^{-1}}$ and $9.83\,\mathrm{N\,kg^{-1}}$, depending on location.

Geological formations can also create differences in gravitational field strength, depending on their composition. Geologists use a sensitive instrument known as a **gravimeter** (Figure 1.2.5) that detects small local variations in gravitational field strength to indicate underground features. Rocks with above-average density, such as those containing mineral ores, create slightly stronger gravitational fields, whereas less-dense sedimentary rocks produce weaker fields.


FIGURE 1.2.5 A gravimeter can be used to measure the strength of the local gravitational field.

If the surface of the Earth is considered a flat surface as it appears in everyday life, then the gravitational field lines are approximately parallel, indicating a uniform field (Figure 1.2.6).

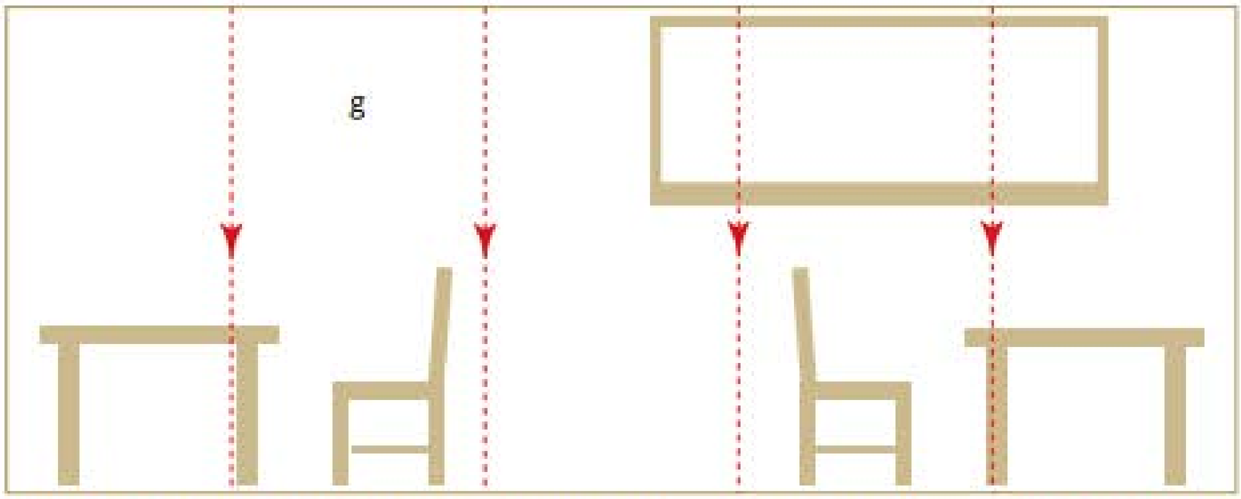


FIGURE 1.2.6 The uniform gravitational field, g, is represented by evenly spaced parallel lines in the direction of the force.

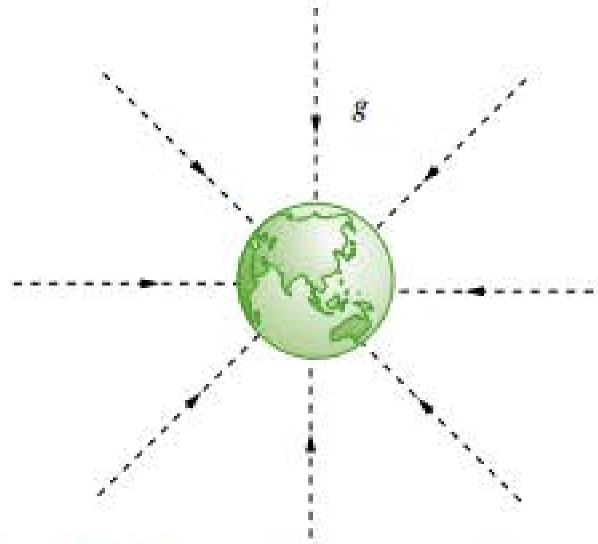


FIGURE 1.2.7 The Earth's gravitational field becomes progressively weaker out into space.

However, when the Earth is viewed from a distance as a sphere, it becomes clear that the Earth's gravitational field is not uniform at all (Figure 1.2.7). The increasing distance between the field lines indicates that the field becomes progressively weaker out into space.

This is because gravitational field strength, like gravitational force, is governed by the inverse square law:

$$g = G\frac{M_{\text{Earth}}}{(r_{\text{Earth}})^2}$$

The gravitational field strength at different altitudes can be calculated by adding the altitude to the radius of the Earth to calculate the distance of the object from Earth's centre (Figures 1.2.8 and 1.2.9).

$$g = \frac{GM_{\text{Earth}}}{(r_{\text{Earth}} + \text{alititude})^2}$$

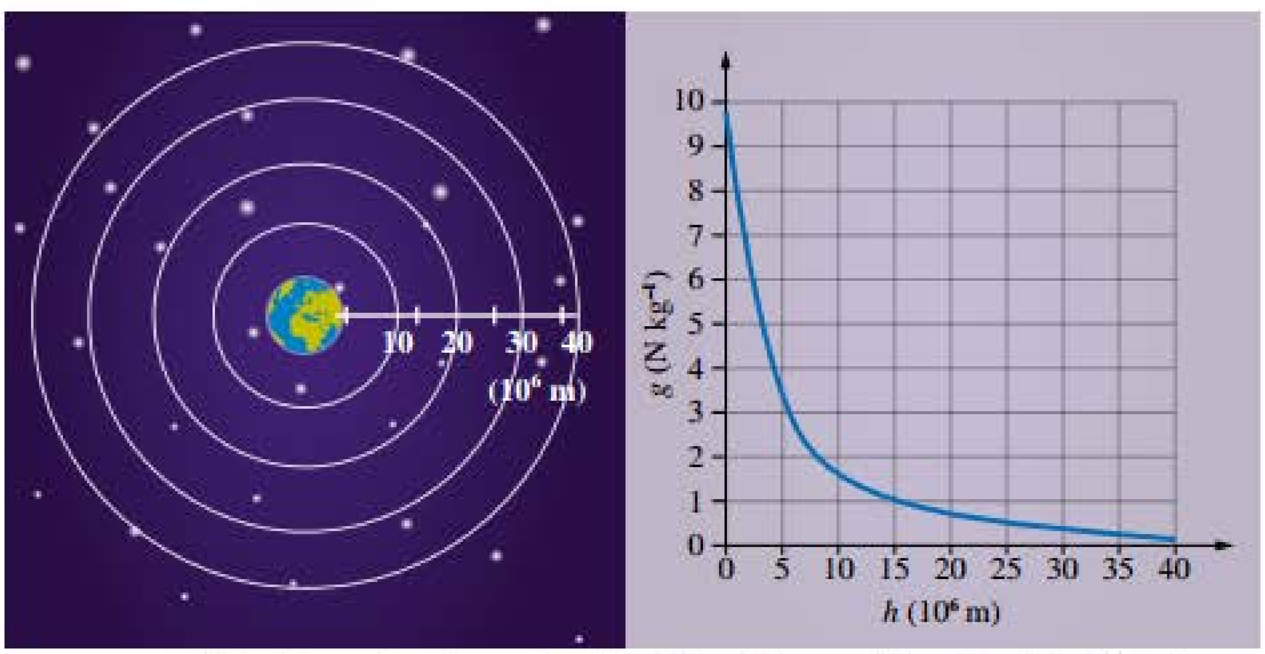


FIGURE 1.2.8 As the distance from the surface of the Earth is increased from 0 to 40×10^6 m, the value for g decreases rapidly from 9.80 N kg^{-1}, according to the inverse square law. The blue line on the graph gives the value of g at various altitudes (h).

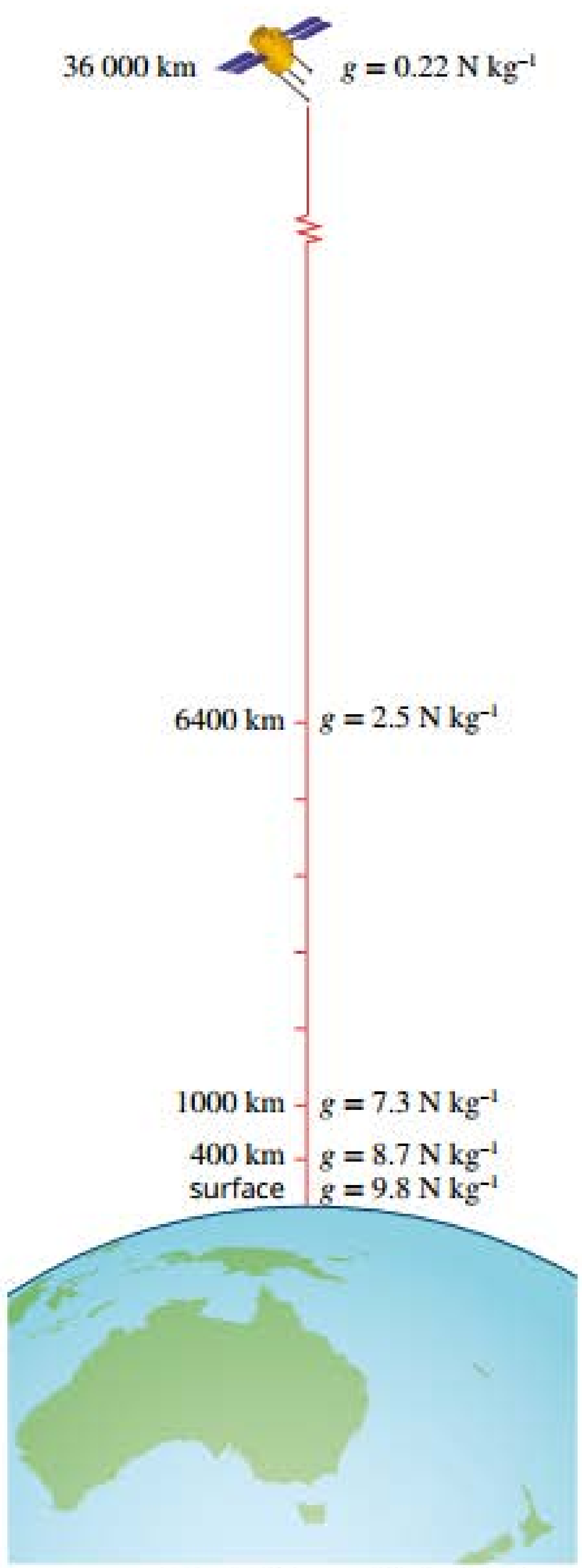


FIGURE 1.2.9 The Earth's gravitational field strength is weaker at higher altitudes.

PHYSICSFILE

Variations in gravitational field strength

The Earth's gravitational field strength is not the same at every point on the Earth's surface. As the Earth is not a perfect sphere, objects near the equator are slightly further from the centre of the Earth than objects at the poles. This means that the Earth's gravitational field is slightly stronger at the poles than at the equator.

The shape of the Earth is known as an oblate spheroid (Figure 1.2.10). Mathematically, this is the shape that's made when an ellipse is rotated around its minor axis. The diameter of the Earth between the North and South poles is approximately 40 km shorter than its diameter at the equator.

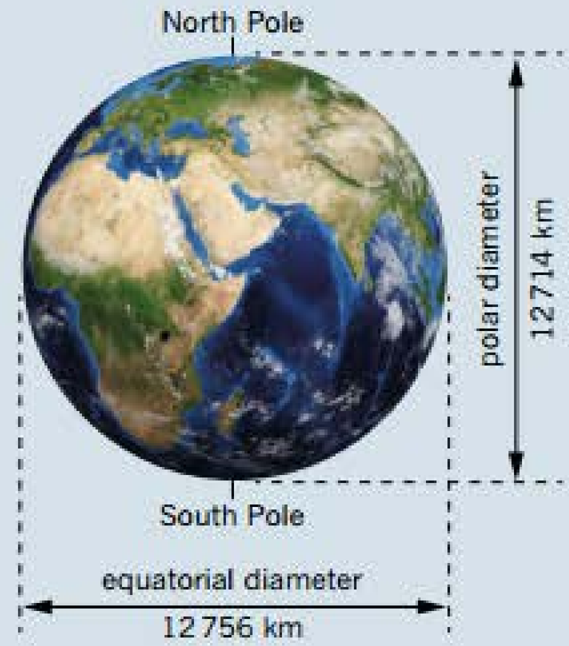


FIGURE 1.2.10 The Earth is a flattened sphere, which means its gravitational field is slightly stronger at the poles.

Worked example 1.2.3

CALCULATING GRAVITATIONAL FIELD STRENGTH AT DIFFERENT ALTITUDES

Calculate the strength of the Earth's gravitational field at the top of Mt Everest using the following data:

$r_{Earth} = 6.38 \times 10^6$ m

$m_{Earth} = 5.97 \times 10^{24}$ kg

height of Mt Everest = 8850 m

Thinking	Working
Recall the formula for gravitational field strength.	$g = G\frac{M}{r^2}$
Add the height of Mt Everest to the radius of the Earth.	$r = 6.38 \times 10^6 + 8850$ m $= 6.389 \times 10^6$ m
Substitute the values into the formula.	$g = G\frac{M}{r^2}$ $= 6.67 \times 10^{-11} \times \frac{5.97 \times 10^{24}}{(6.389 \times 10^6)^2}$ $= 9.76\ \text{N kg}^{-1}$

Worked example: Try yourself 1.2.3

CALCULATING GRAVITATIONAL FIELD STRENGTH AT DIFFERENT ALTITUDES

Commercial airlines typically fly at an altitude of 11 000 m. Calculate the gravitational field strength of the Earth at this height using the following data:

$r_{Earth} = 6.38 \times 10^6$ m

$m_{Earth} = 5.97 \times 10^{24}$ kg

Gravitational field strengths of other planets

The gravitational field strength on the surface of the Moon is much less than on Earth, at approximately 1.6 N kg^{-1}. This is because the Moon's mass is smaller than that of the Earth.

The formula $g = G\frac{M}{r^2}$ can be used to calculate the gravitational field strength on the surface of any astronomical object, such as Mars (Figure 1.2.11).

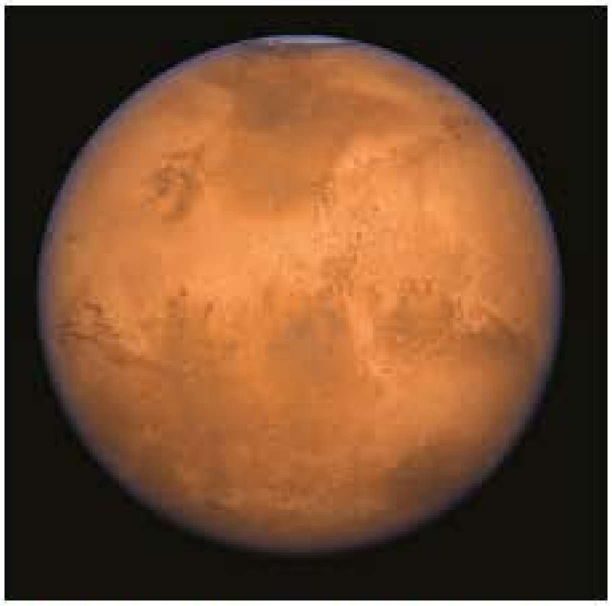


FIGURE 1.2.11 The gravitational field strength on the surface of Mars (shown here) is different from the gravitational field strength on the surface of the Earth, which, in turn, is different from that on the Moon.

Worked example 1.2.4

GRAVITATIONAL FIELD STRENGTH ON ANOTHER PLANET OR MOON

Calculate the strength of the gravitational field on the surface of the Moon given that the Moon's mass is 7.35×10^{22} kg and its radius is 1740 km.
Give your answer correct to two significant figures.

Thinking	Working
Recall the formula for gravitational field strength.	$g = G\frac{M}{r^2}$
Convert the Moon's radius to m.	$r = 1740$ km $= 1740 \times 1000$ m $= 1.74 \times 10^6$ m
Substitute values into the formula.	$g = G\frac{M}{r^2}$ $= 6.67 \times 10^{-11} \times \frac{7.35 \times 10^{22}}{(1.74 \times 10^6)^2}$ $= 1.6$ N kg^{-1}

Worked example: Try yourself 1.2.4

GRAVITATIONAL FIELD STRENGTH ON ANOTHER PLANET OR MOON

Calculate the strength of the gravitational field on the surface of Mars.
$m_{Mars} = 6.42 \times 10^{23}$ kg
$r_{Mars} = 3390$ km
Give your answer correct to two significant figures.

PHYSICSFILE

Moon walking

Walking is a process of repeatedly stopping yourself from falling over. When astronauts first tried to walk on the Moon, they found that they fell too slowly to walk easily. Instead, they invented a kind of shuffling jump that was a much quicker way of moving around in the Moon's weak gravitational field (Figure 1.2.12). This type of 'moon walk' should not be confused with the famous dance move of the same name!

FIGURE 1.2.12 Astronauts had to invent a new way of walking to deal with the Moon's weak gravitational field.

1.2 Review

SUMMARY

- A gravitational field is a region in which a gravitational force is exerted on all matter within that region.
- A gravitational field can be represented by a gravitational field diagram:
  - The arrowheads indicate the direction of the gravitational force.
  - The spacing of the field lines indicates the relative strength of the field. The closer the line spacing, the stronger the field.
- The strength of a gravitational field can be calculated using the following formulae:

$$g = \frac{F_{weight}}{m} \text{ or } g = G\frac{M}{r^2}$$

- The gravitational field strength on the Earth's surface is approximately 9.80 N kg^{-1}. This varies from location to location and with altitude.
- The gravitational field strength on the surface of any other planet depends on the mass and radius of the planet.

KEY QUESTIONS

1 What is the most appropriate unit for measuring gravitational field strength in the context of gravitational fields?

2 Students use a spring balance to measure the weight of a 150 g set of slotted masses to be 1.4 N. According to this measurement, what is the gravitational field strength in their classroom?

3 A gravitational field, g, is measured as 5.5 N kg^{-1} at a distance of 40 000 km from the centre of a planet. The distance from the centre of the planet is then increased to 120 000 km. What would be the ratio of the magnitude of the gravitational field at this new distance to the magnitude of original measurement?

4 Different types of satellite have different types of orbit, as shown in the table below. Calculate the strength of the Earth's gravitational field in each orbit.

r_{Earth} = 6380 km

m_{Earth} = 5.97 × 10^{24} kg

	Type of orbit	Altitude (km)
a	low-Earth orbit	2000
b	medium-Earth orbit	10 000
c	semi-synchronous orbit	20 200
d	geosynchronous orbit	35 786

5 On 12 November 2014, the Rosetta spacecraft landed a probe on the comet 67P/Churyumov–Gerasimenko. Assuming this comet is a roughly spherical object with a mass of 1 × 10^{13} kg and a diameter of 1.8 km, calculate the gravitational field strength on its surface.

6 The masses and radii of three planets are given in the following table. Calculate the gravitational field strength, g, at the surface of each planet.

Planet	Mass (kg)	Radius (m)
Mercury	3.30 × 10^{23}	2.44 × 10^{6}
Saturn	5.69 × 10^{26}	6.03 × 10^{7}
Jupiter	1.90 × 10^{27}	7.15 × 10^{7}

7 There are bodies outside our solar system, such as neutron stars, that produce very large gravitational fields. A typical neutron star can have a mass of 3.0 × 10^{30} kg and a radius of just 10 km. Calculate the gravitational field strength at the surface of such a star.

8 A newly discovered solid planet located in a distant solar system is found to be distinctly non-spherical in shape. Its polar radius is 5000 km, and its equatorial radius is 6000 km.
The gravitational field strength at the poles is 8.0 N kg^{-1}. How would the gravitational field strength at the poles compare with the strength at the equator?

9 There is a point between the Earth and the Moon where the total gravitational field is zero. The significance of this is that returning lunar missions are able to return to Earth under the influence of the Earth's gravitational field once they pass this point. Given that the mass of Earth is 6.0 × 10^{24} kg, the mass of the Moon is 7.3 × 10^{22} kg and the radius of the Moon's orbit is 3.8 × 10^{8} m, calculate the distance of this point from the centre of the Earth.

10 An astronaut travels away from Earth to a region in space where the gravitational force due to Earth is only 1.0% of that at Earth's surface. What distance, in Earth radii, is the astronaut from the centre of the Earth?

1.3 Work in a gravitational field

The concept of gravitational potential energy should be familiar to you from Unit 2 Physics. However, the nature of a gravitational field means that a more sophisticated understanding of gravitational potential energy is needed when considering the motion of objects such as rockets and satellites (Figure 1.3.1).

FIGURE 1.3.1 Satellites in orbit have gravitational potential energy.

REVISITING WORK AND CONSERVATION OF ENERGY

The **gravitational potential energy** of an object, E_g, is directly proportional to the mass of the object, m, its height above the surface of the planet, Δh, and the strength of the gravitational field, g:

$$E_g = mg\Delta h$$

where E_g is the gravitational potential energy of an object (J)

m is the mass of the object (kg)

g is the gravitational field strength (N kg^{-1}; 9.80 N kg^{-1} near the surface of the Earth)

Δh is the height of the object above a reference point (m)

The formula for gravitational potential energy is developed from the work–energy theorem, which assumes that work done against the force of gravity is converted into potential energy:

$$\Delta E = W = Fs$$

where ΔE is the change in gravitational potential energy (J)

W is the work done (J)

F is the force of gravity (N)

s is the distance moved in the gravitational field (m)

Worked example 1.3.1

WORK DONE FOR A CHANGE IN GRAVITATIONAL POTENTIAL ENERGY

A mountaineer climbs from a height of 700 m above sea level to the top of Mount Everest, which is 8848 m above sea level.

The total mass of the mountaineer (with equipment) is 100 kg. Assuming that the gravitational field strength of the Earth (g) is 9.80 N kg^{-1}, calculate the amount of work done (in MJ) by the mountaineer in climbing to the summit of the mountain.

Thinking	Working
Calculate the change in height.	$\Delta h = 8848 - 700$ $= 8148\text{ m}$
Substitute appropriate values into $E_g = mg\Delta h$. Remember to give your answer in MJ to two significant figures.	$E_g = mg\Delta h$ $= 100 \times 9.80 \times 8148$ $= 7\,985\,040\text{ J}$ $= 8.0\text{ MJ}$
The work done by the mountaineer is equal to the change in gravitational potential energy.	$W = \Delta E = 8.0\text{ MJ}$

Worked example: Try yourself 1.3.1

WORK DONE FOR A CHANGE IN GRAVITATIONAL POTENTIAL ENERGY

Calculate the work done (in MJ) to lift a weather satellite of 3.2 tonnes from the Earth's surface to the limit of the atmosphere, which ends at the Karman line (exactly 100 km up from the surface of the Earth). Assume $g = 9.80$ N kg^{-1}.

INTERPLAY BETWEEN GRAVITATIONAL, KINETIC AND MECHANICAL ENERGY

Gravitational potential energy calculations are important because, when combined with the concepts of kinetic energy and conservation of mechanical energy, they allow the speed of a falling object to be determined.

Accordingly, we can define **kinetic energy** by the following equation:

> $E_k = \frac{1}{2}mv^2$
>
> where E_k is the kinetic energy of an object (J)
>
> m is the mass of the object (kg)
>
> v is the speed of the object ($m\,s^{-1}$)

Worked example 1.3.2

SPEED OF A FALLING OBJECT

In a unique demonstration of Galileo's famous experiment, Apollo 15 astronaut Dave Scott simultaneously dropped a hammer and a feather while standing on the surface of the Moon (Figure 1.3.2).

FIGURE 1.3.2 An artist's impression of Astronaut Dave Scott dropping a hammer and a feather on the Moon.

If the gravitational field strength on the Moon is $1.6\,N\,kg^{-1}$, the hammer had a mass of 450 g and it was dropped from a height of 1.4 m, calculate the speed of the hammer as it hit the Moon's surface.

Thinking	Working
Calculate the gravitational potential energy of the hammer on the Moon. Change the units of measurement when necessary.	$E_g = mg\Delta h$ $= 0.45 \times 1.6 \times 1.4$ $= 1.0\,J$
Assume that when the hammer hit the surface of the Moon, all of its gravitational potential energy had been converted into kinetic energy.	$E_k = E_g = 1.0\,J$
Use the definition of kinetic energy to calculate the speed of the hammer as it hit the ground.	$E_k = \frac{1}{2}mv^2$ $1.0 = \frac{1}{2} \times 0.45 \times v^2$ $\frac{1.0 \times 2}{0.45} = v^2$ $v = 2.1\,m\,s^{-1}$

Worked example: Try yourself 1.3.2

SPEED OF A FALLING OBJECT

Calculate how fast a 450 g hammer would be going as it hit the ground if it was dropped from a height of 1.4 m on Earth, where $g = 9.80\,N\,kg^{-1}$.

WORK IN A NON-CONSTANT GRAVITATIONAL FIELD

The formula $E_g = mg\Delta h$ is developed assuming that work is done against a constant force of gravity: $\Delta E = W = Fs$. Although this assumption holds true for objects close to the surface of a planet, it is not adequate for objects such as satellites that move to altitudes at which the gravitational field of the planet becomes significantly diminished.

Newton's law of universal gravitation indicates that the strength of the Earth's gravitational field changes with altitude. The field is stronger close to the ground and weaker at high altitudes (Figure 1.3.3).

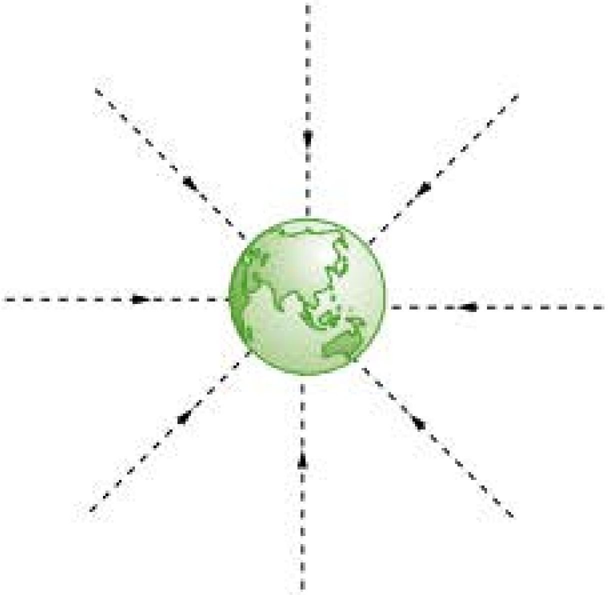


FIGURE 1.3.3 The Earth's gravitational field extends out into space, and the field is strongest close to the Earth where the field lines are closest together.

Consider the example of a 10 kg meteor falling towards the Earth from deep space, as shown in Figure 1.3.4. Closer to the Earth, the meteor moves through regions of increasing gravitational field strength. So the gravitational force, F_g, on the meteor increases as it approaches Earth. As the force is not constant, this means that the work done on the meteor (which corresponds to its change in gravitational potential energy) cannot be found by simply multiplying the gravitational force by the distance travelled.

FIGURE 1.3.4 As the meteor approaches Earth, it moves through an increasingly stronger gravitational field and so is acted upon by a greater gravitational force.

Using the force–distance graph

When a free-falling body, like the meteor in Figure 1.3.4, is acted upon by a varying gravitational force, the energy changes of the body can be analysed either by calculus or, more simply, by using a gravitational force–distance graph. As with other force–distance graphs, the area under the graph is equal to the work done (i.e. the energy change of the body). The area under the graph has the unit newton metres (N m), which is equivalent to the joule (J).

The area under a gravitational force–distance graph gives the change in energy that an object will experience as it moves through the gravitational field.

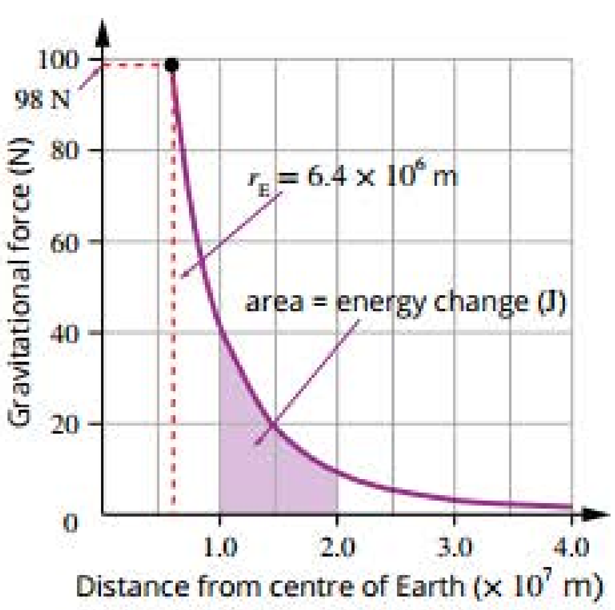


FIGURE 1.3.5 The gravitational force acting on a 10 kg meteor at different distances from the Earth. The shaded region represents the work done by the gravitational field as the body moves between 2.0×10^7 m and 1.0×10^7 m from the centre of the Earth.

The shaded area in Figure 1.3.5 represents the decrease in gravitational potential energy of the 10 kg meteor as it falls from a distance of 2.0×10^7 m to 1.0×10^7 m from the centre of the Earth. This area also represents the amount of kinetic energy that the meteor gains as it approaches Earth.

Note that the energy change of the meteor will be the same regardless of whether the meteor falls directly towards the planet (Figure 1.3.6(a)) or follows a more indirect path (Figure 1.3.6(b)).

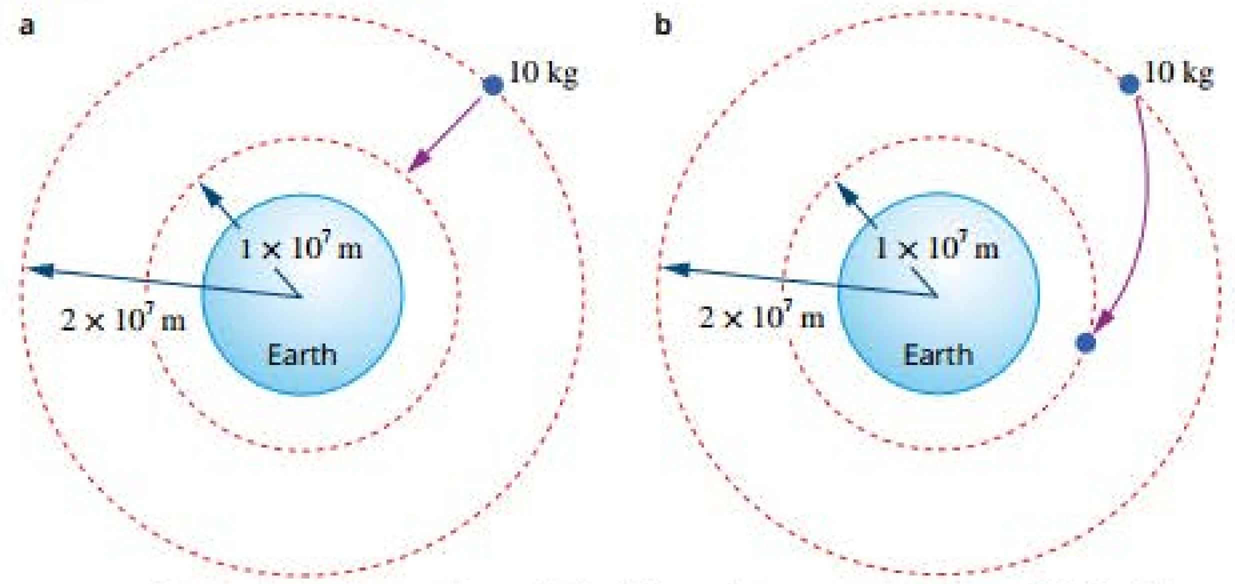


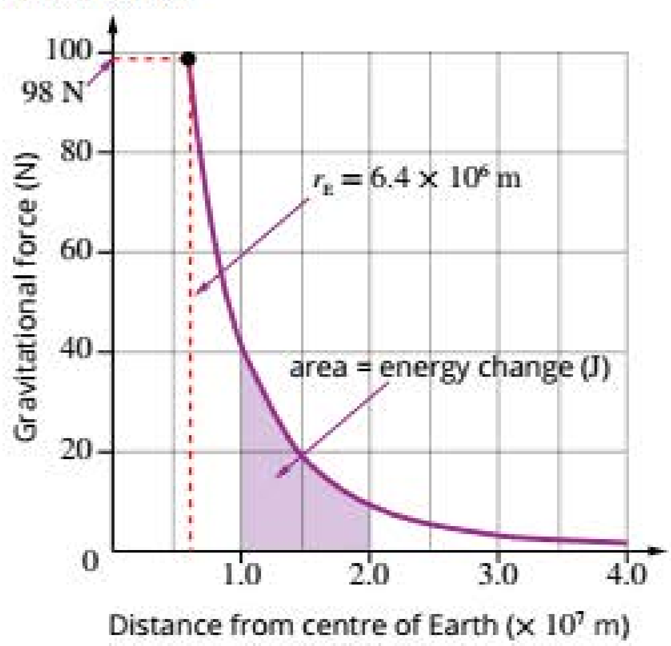
FIGURE 1.3.6 The shaded region on the gravitational force–distance graph in Figure 1.3.5 could represent the change in energy in the free-fall situations in either (a) or (b).

Worked example 1.3.3 shows how a force–distance graph can be used to determine the change in gravitational potential energy of a meteor.

Worked example 1.3.3

CHANGE IN GRAVITATIONAL POTENTIAL ENERGY USING A FORCE–DISTANCE GRAPH

A 10 kg meteor falls from a distance of 2.0×10^7 m to 1.0×10^7 m from the centre of the Earth. Use the graph below to determine the change in gravitational potential energy of the meteor.

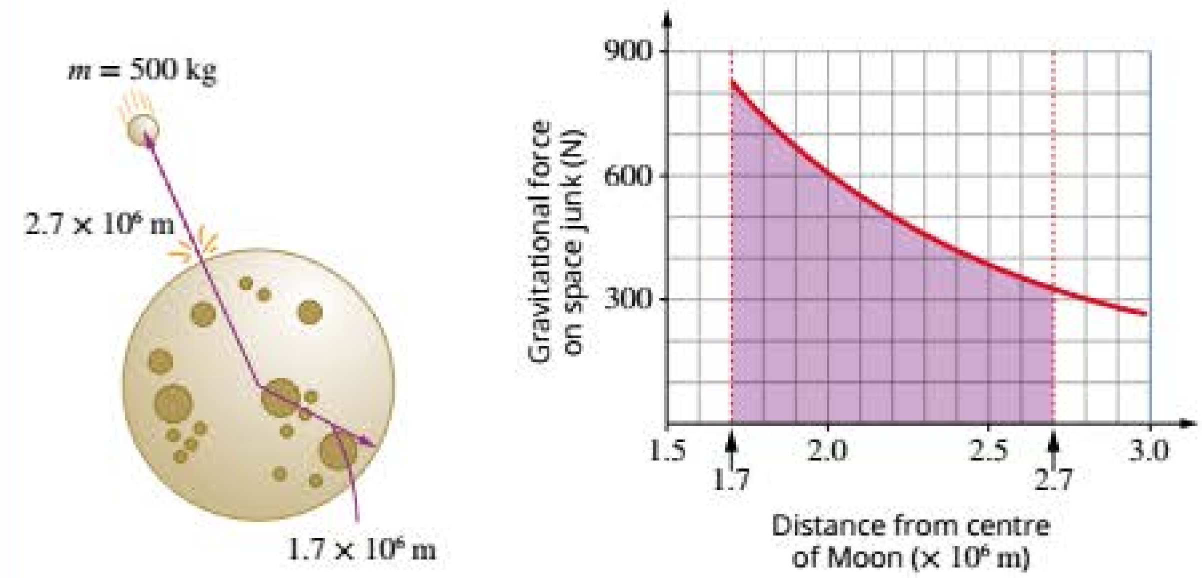


Thinking	Working
Count the number of shaded squares. (In this example, count the partially shaded squares as half squares.)	Number of shaded squares = 2
Calculate the area (energy value) of each square.	$E_{square} = 0.5 \times 10^7 \times 20$ $= 1 \times 10^8$ J
To calculate the energy change, multiply the number of shaded squares by the energy value of each square.	$\Delta E_g = 2 \times (1 \times 10^8)$ $= 2 \times 10^8$ J

Worked example: Try yourself 1.3.3

CHANGE IN GRAVITATIONAL POTENTIAL ENERGY USING A FORCE–DISTANCE GRAPH

A 500 kg lump of space junk is plummeting towards the Moon. The Moon has a radius of 1.7×10^6 m. Using the force–distance graph, determine the decrease in gravitational potential energy of the junk as it falls to the Moon's surface.



PHYSICSFILE

Using a force–distance graph in a constant field

A force–distance graph can also be used to calculate the change in gravitational potential energy of an object falling in a uniform gravitational field. Consider the graph in Figure 1.3.7 of a 10 kg rock that falls from a height of 40 m to 10 m.

The area under the graph in Figure 1.3.7 is 98 N × 30 m = 294 J, which is exactly what would have been calculated using the formula $E_g = mg\Delta h$. Therefore, it is more convenient to use the formula in uniform field situations.

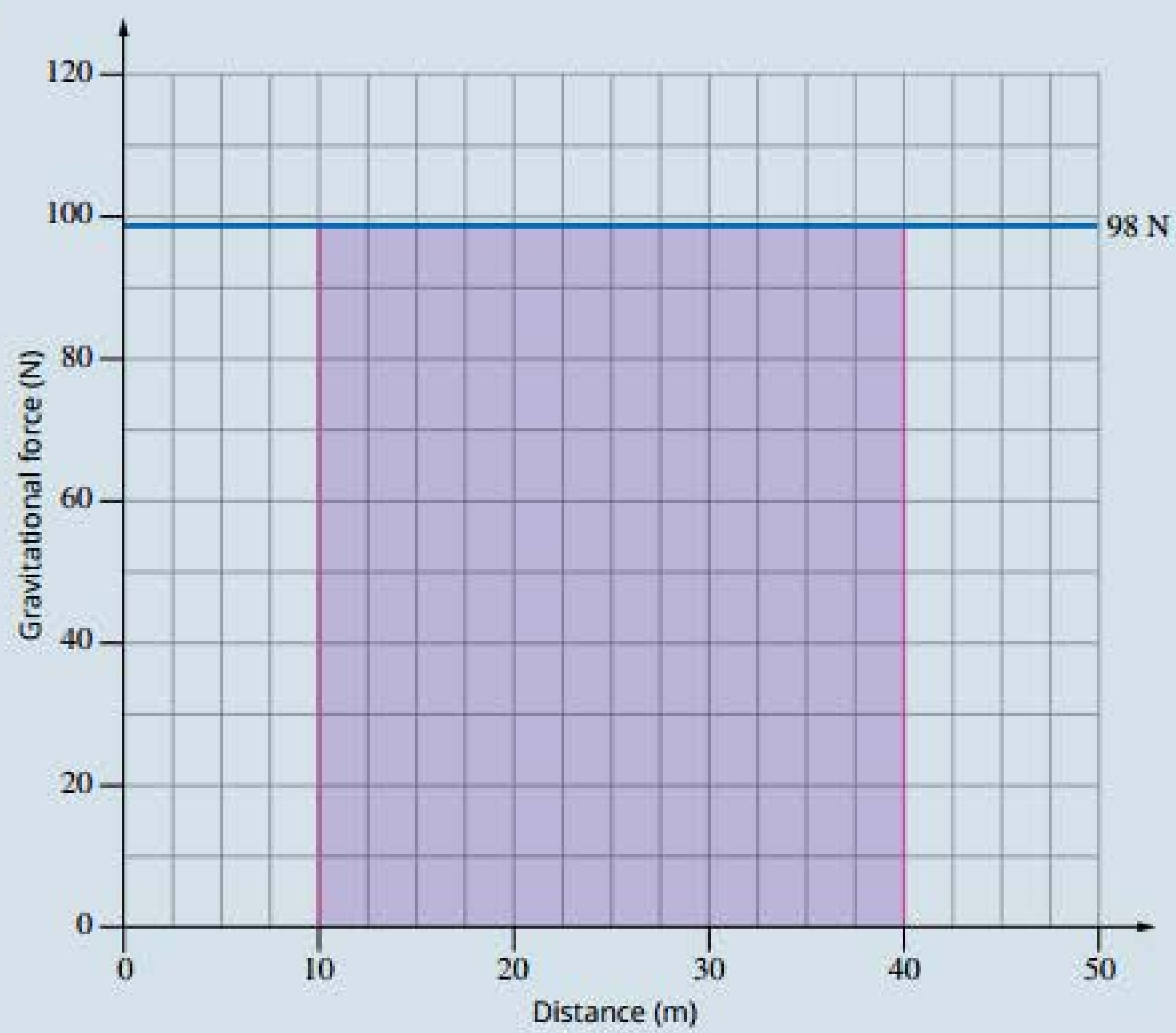


FIGURE 1.3.7 In a uniform field, the gravitational force–distance graph is a horizontal line.

USING THE GRAVITATIONAL FIELD STRENGTH–DISTANCE GRAPH

The change in gravitational potential energy of an object can also be calculated using a graph of the gravitational field strength of an object, as shown in Figure 1.3.8.

The area under a gravitational field strength–distance graph gives a quantity that has units of $N\,kg^{-1} \times m$, which is equivalent to $J\,kg^{-1}$, so the area indicates the change in energy for each kilogram of the object's mass. To find the work done or energy change (J), the area ($J\,kg^{-1}$) must be multiplied by the mass (kg) of the object.

The area under a gravitational field strength–distance graph gives the energy change per kilogram of mass. To find the change in energy, the area must be multiplied by the mass, in kg, of the object.

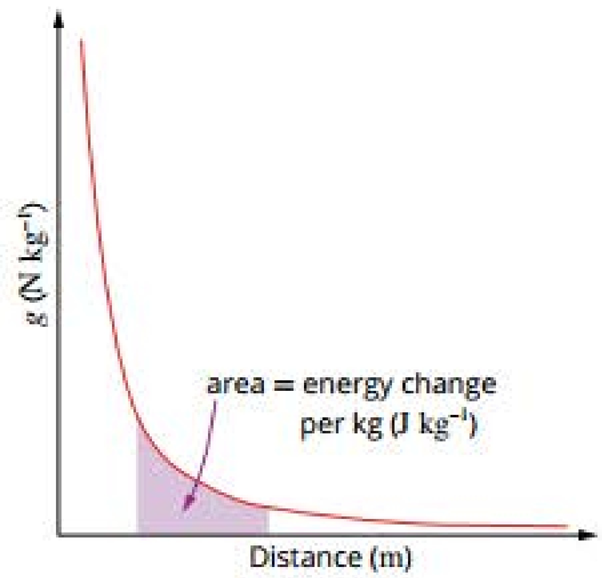
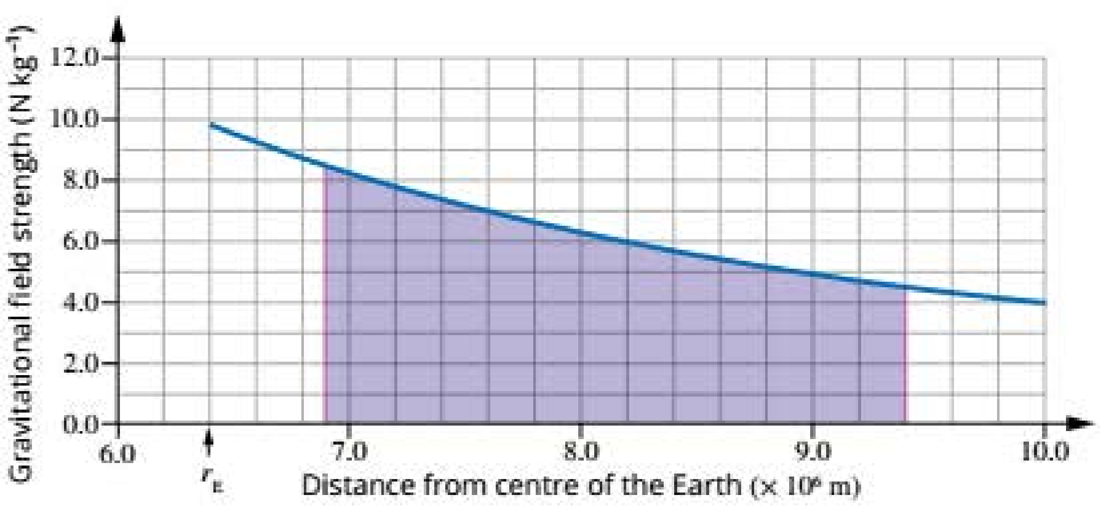


FIGURE 1.3.8 A graph of gravitational field strength–distance can also be used to analyse the energy changes of a body moving through a gravitational field.

Worked example 1.3.4

CHANGE IN GRAVITATIONAL POTENTIAL ENERGY USING A GRAVITATIONAL FIELD STRENGTH–DISTANCE GRAPH

A wayward satellite of mass 1500 kg has developed a highly elliptical orbit around the Earth. At its closest approach (perigee), it is just 500 km above the Earth's surface. At its furthest point (apogee) it is 3000 km from the Earth's surface. Using the graph of the gravitational field strength of the Earth shown below, determine the approximate change in the gravitational potential energy of the satellite as it orbits. (Note: the radius of the Earth is 6400 km.)

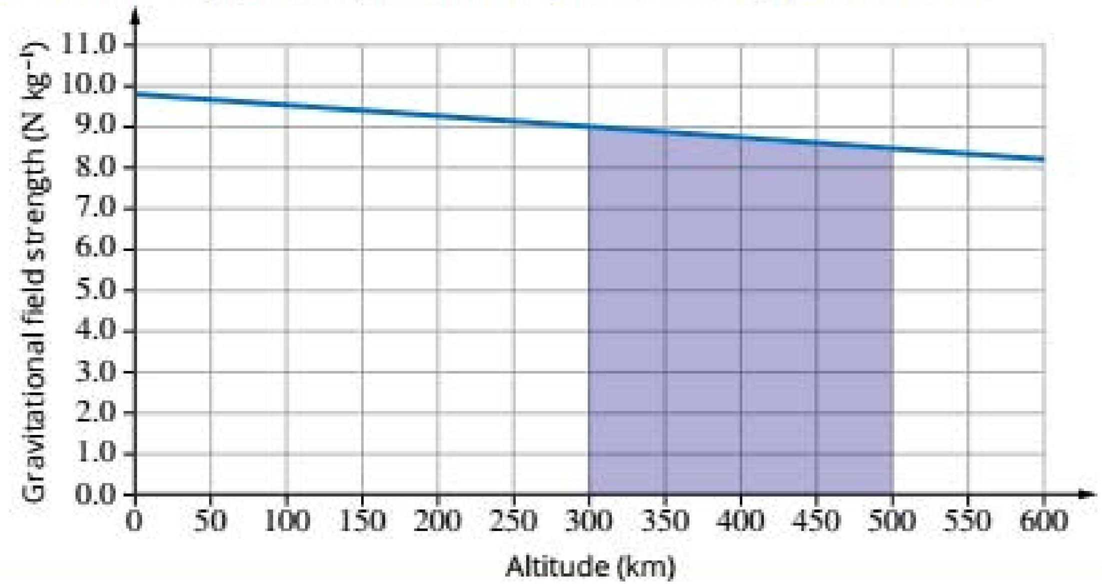


Thinking	Working
Count the number of shaded squares. Only count squares that are at least 50% shaded.	Number of shaded squares = 82
Calculate the energy value of each square.	$E_{square} = 0.2 \times 10^6\,m \times 1\,N\,kg^{-1}$ $= 2 \times 10^5\,J\,kg^{-1}$
To calculate the energy change, multiply the number of shaded squares by the energy value of each square and the mass of the satellite.	$\Delta E_g = 82 \times (2 \times 10^5) \times 1500$ $= 2.5 \times 10^{10}\,J$

Worked example: Try yourself 1.3.4

CHANGE IN GRAVITATIONAL POTENTIAL ENERGY USING A GRAVITATIONAL FIELD STRENGTH–DISTANCE GRAPH

A 3000 kg Soyuz rocket moves from an orbital height of 300 km above the Earth's surface to dock with the International Space Station at a height of 500 km. Use the graph of the gravitational field strength of the Earth below to determine the approximate change in the gravitational potential energy of the rocket.



1.3 Review

SUMMARY

- The principles of work and conservation of energy are useful for understanding gravitational potential energy. This includes the following formulae:

  $W = Fs$

  $W = \Delta E$

  $E_k = \frac{1}{2}mv^2$

- The gravitational potential energy formula $E_g = mg\Delta h$ assumes that the Earth's gravitational field is constant. This is approximately true for objects that are within a few kilometres of the Earth's surface.
- The strength of the Earth's gravitational field decreases as altitude increases.
- The area under a gravitational force–distance graph gives the change in kinetic energy or change in gravitational potential energy of a body, and indicates the work done by the gravitational field.
- The area under a gravitational field strength–distance graph gives the change in energy per kilogram ($J\,kg^{-1}$) of the object. To calculate the energy change, multiply the area by the mass (kg).

KEY QUESTIONS

1 Which one of the following statements is correct? A satellite in a stable circular orbit around the Earth will have:

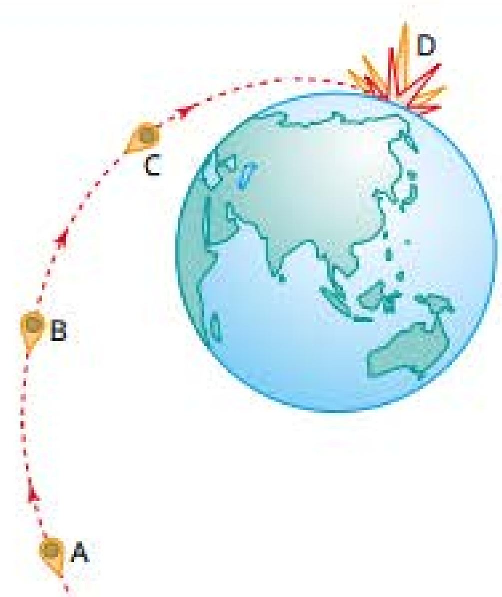
A varying potential energy as it orbits

B varying kinetic energy as it orbits

C constant kinetic energy and constant potential energy

The following information applies to questions 2–4.

Ignore air resistance when answering these questions. The path of a meteor plunging towards the Earth is as shown.



2 How does the gravitational field strength of the Earth change from point A to point D?

3 How will the acceleration of the meteor change as it travels along the path shown?

4 Which one or more of the following statements is correct?

A The kinetic energy of the meteor increases as it travels from A to D.

B The gravitational potential energy of the meteor decreases as it travels from A to D.

C The total energy of the meteor remains constant.

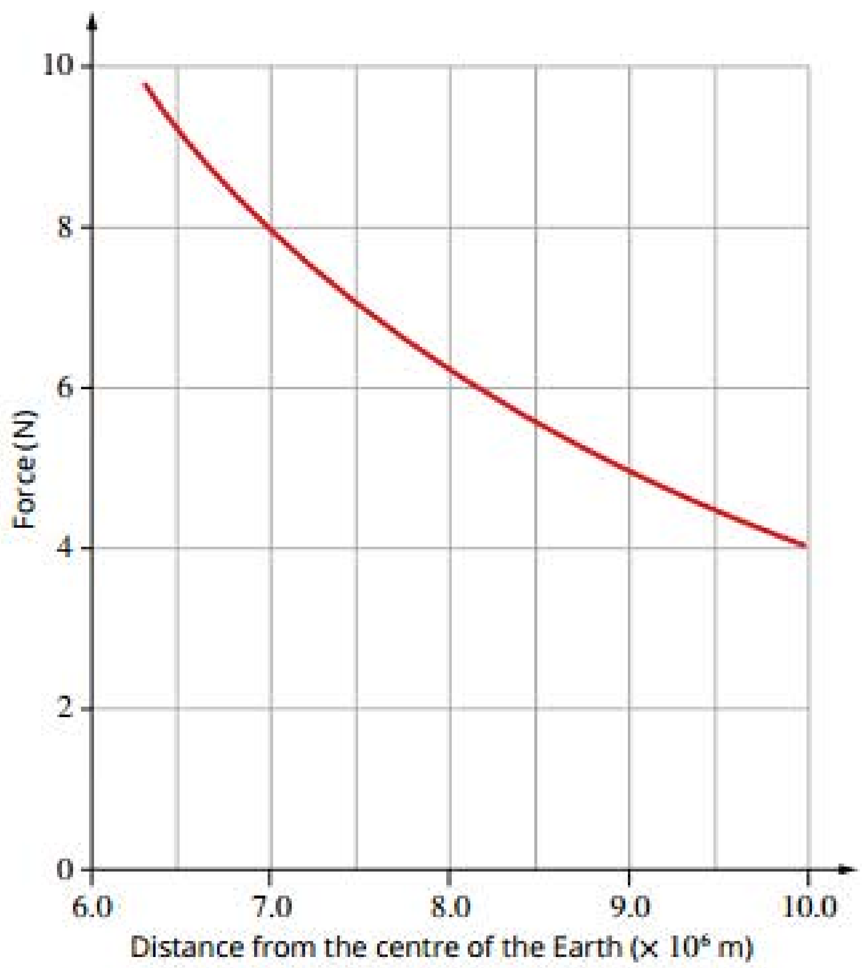
D The total energy of the meteor increases.

5 The Saturn V rocket that took the first astronauts to the Moon had a mass of 3000 tonnes. Its stage I rockets fired for 6 minutes and took the rocket to an altitude of 67 km. How much work did the stage I rockets do in this time?

6 The Valles Marineris on Mars is one of the most spectacular land formations in the solar system: a gigantic canyon 4000 km long, 200 km wide and 7 km deep. If a Martian explorer were to drop a 400 g rock from the edge of the canyon to its floor 7000 m below, how fast would the rock be going when it hit the bottom? The gravitational field strength on Mars is weaker than on Earth: 6.1 N kg^{-1}.

The following information applies to questions 7–9.

The graph shows the force on a mass of 1.0 kg as a function of its distance from the centre of the Earth.



7 **a** Use the graph to determine the gravitational force between the Earth and a 1.0 kg mass 100 km above the Earth's surface.

b Use the graph to determine the height above the Earth's surface at which a 1.0 kg mass would experience a gravitational force of 5.0 N.

8 A meteor of mass 1.0 kg is speeding towards the Earth. When this meteor is at a distance of 9.5×10^6 m from the centre of the planet, its speed is 4.0 km s^{-1}.

a Determine the kinetic energy of the meteor when it is 9.5×10^6 m from the centre of the Earth.

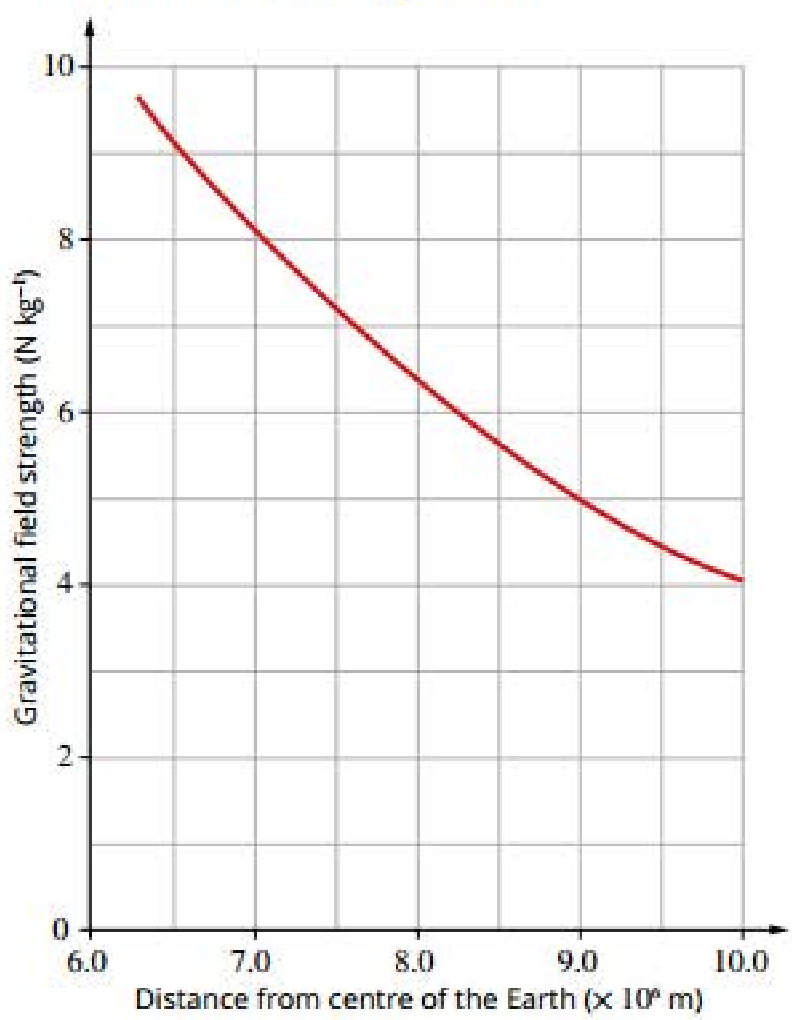
b Calculate the increase in kinetic energy of the meteor as it moves from a distance of 9.5×10^6 m from the centre of the Earth to a point that is 6.5×10^6 m from the centre.

c Ignoring air resistance, what is the kinetic energy of the meteor when it is 6.5×10^6 m from the centre of the Earth?

d How fast is the meteor travelling when it is 6.5×10^6 m from the centre of the Earth?

9 A communications satellite of mass 240 kg is launched from a space shuttle that is in orbit 600 km above the Earth's surface. The satellite travels directly away from the Earth and reaches a maximum distance of 8000 km from the centre of the Earth before stopping due to the influence of the Earth's gravitational field. Use the graph to estimate the kinetic energy of this satellite as it was launched.

10 A 20 tonne remote-sensing satellite is in a circular orbit around the Earth at an altitude of 600 km. The satellite is moved to a new stable orbit with an altitude of 2600 km. Use the following graph to estimate the increase in the gravitational potential energy of the satellite as it moved from its lower orbit to its higher orbit.



Chapter review

01

KEY TERMS

acceleration due to gravity
altitude
apparent weight
apparent weightlessness
artificial satellites
centripetal acceleration
field
free-fall
geostationary satellite
gravimeter
gravitational constant
gravitational field
gravitational field strength
gravitational force
gravitational potential energy
inverse square law
kinetic energy
natural satellite
Newton's law of universal gravitation
normal reaction force
torsion balance
uniform
weight

1 Use Newton's law of universal gravitation to calculate the gravitational force acting on a person with a mass of 75 kg. Use the following data:
$m_{Earth} = 6.0 \times 10^{24}$ kg
$r_{Earth} = 6400$ km

2 The gravitational force of attraction between Saturn and Dione, a moon of Saturn, is equal to 2.79×10^{20} N. Calculate the orbital radius of Dione. Use the following data:
mass of Dione = 1.05×10^{21} kg
mass of Saturn = 5.69×10^{26} kg

3 Of all the planets in the solar system, Jupiter exerts the largest force on the Sun: 4.2×10^{23} N. Calculate the acceleration of the Sun due to this force, using the following data: $m_{Sun} = 2.0 \times 10^{30}$ kg.

4 The planet Jupiter and the Sun exert gravitational forces on each other.
a Compare, qualitatively, the force exerted on Jupiter by the Sun to the force exerted on the Sun by Jupiter.
b Compare, qualitatively, the acceleration of Jupiter caused by the Sun to the acceleration of the Sun caused by Jupiter.

5 Calculate the acceleration due to gravity on the surface of Mars if it has a mass of 6.4×10^{23} kg and a radius of 3400 km.

6 Calculate the apparent weight of a 50 kg person in an elevator under the following circumstances.
a accelerating downwards at 0.6 m s^{-2}
b moving downwards at a constant speed of 2 m s^{-1}

7 A comet of mass 1000 kg is plummeting towards Jupiter. Jupiter has a mass of 1.90×10^{27} kg and a planetary radius of 7.15×10^{7} m. If the comet is about to crash into Jupiter, calculate the:
a magnitude of the gravitational force that Jupiter exerts on the comet
b magnitude of the gravitational force that the comet exerts on Jupiter
c acceleration of the comet towards Jupiter
d acceleration of Jupiter towards the comet.

8 A person standing on the surface of the Earth experiences a gravitational force of 900 N. What gravitational force will this person experience at a height of two Earth radii above the Earth's surface?
A 900 N
B 450 N
C zero
D 100 N

9 During a space mission, an astronaut of mass 80 kg initially accelerates at 30 m s^{-2} upwards, then travels in a stable circular orbit at an altitude where the gravitational field strength is 8.2 N kg^{-1}.
a What is the apparent weight of the astronaut during lift-off?
A zero
B 780 N
C 2400 N
D 3200 N
b During the lift-off phase, the astronaut will feel:
A lighter than usual
B heavier than usual
C the same as usual
c The weight of the astronaut during the lift-off phase is:
A lower than usual
B greater than usual
C the same as usual
d During the orbit phase, the apparent weight of the astronaut is:
A zero
B 780 N
C 2400 N
D 660 N
e During the orbit phase, the weight of the astronaut is:
A zero
B 780 N
C 2400 N
D 660 N

10 Describe the main rules to follow when drawing gravitational field lines.

11 A set of bathroom scales is calibrated so that when the person standing on it has a weight of 600 N, the scales read 61.5 kg. What gravitational field strength has been assumed in this setting?

12 The Earth is a slightly flattened sphere. Its radius at the poles is 6357 km compared to 6378 km at the equator. The Earth's mass is 5.97×10^{24} kg.

a Calculate the Earth's gravitational field strength at the equator.

b Using the information in part (a), calculate how much stronger the gravitational field would be at the North Pole compared with at the equator. Give your answer as a percentage of the strength at the equator.

13 Neptune has a planetary radius of 2.48×10^7 m and a mass of 1.02×10^{26} kg.

a Calculate the gravitational field strength on the surface of Neptune.

b A 250 kg lump of ice is falling directly towards Neptune. What is its acceleration as it nears the surface of Neptune? Ignore any drag effects.

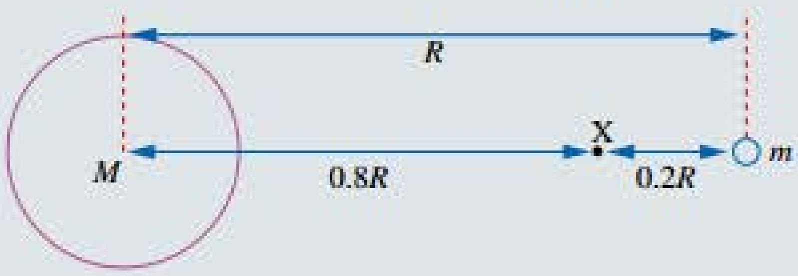
A $9.80\ \text{m s}^{-2}$

B zero

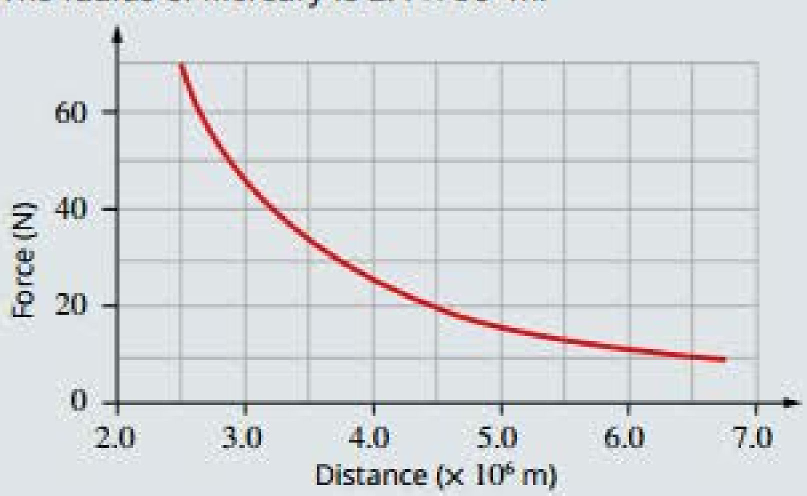
C $11\ \text{m s}^{-2}$

D $1.6\ \text{m s}^{-2}$

14 Two stars of masses *M* and *m* are in orbit around each other. As shown in the following diagram, they are a distance *R* apart. A spacecraft located at point X experiences zero net gravitational force from these stars. Calculate the value of the ratio $\frac{M}{m}$.



15 A 20 kg rock is speeding towards Mercury. The following graph shows the force on the rock as a function of its distance from the centre of the planet. The radius of Mercury is 2.4×10^6 m.



When the rock is 3.0×10^6 m from the centre of the planet, its speed is estimated at $1.0\ \text{km s}^{-1}$. Using the graph, estimate the:

a increase in kinetic energy of the rock as it moves to a point that is just 2.5×10^6 m from the centre of Mercury

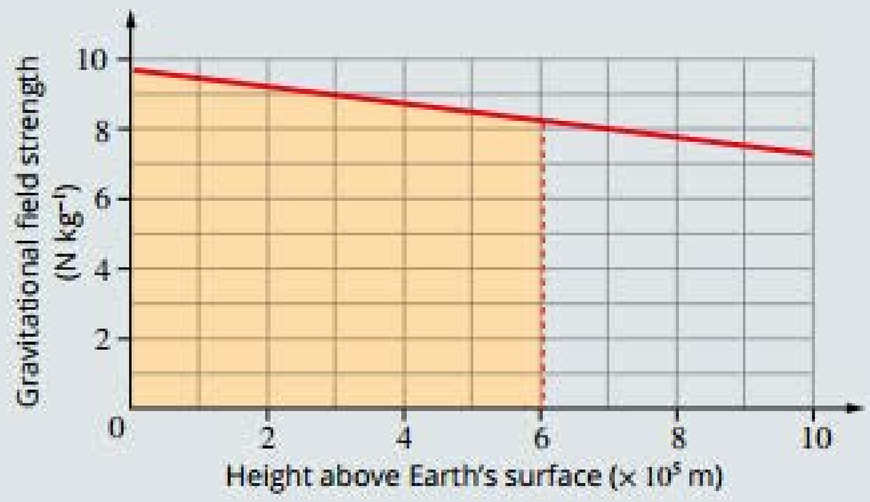
b kinetic energy of the rock at this closer point

c speed of the rock at this point

d gravitational field strength at 2.5×10^6 m from the centre of Mercury.

The following information relates to questions 16–20.

The diagram shows the gravitational field strength and distance near the Earth. A wayward satellite of mass 1000 kg is drifting towards the Earth.



16 What is the gravitational field strength at an altitude of 300 km?

17 Which of the following units is associated with the area under this graph?

A J

B m s^{-2}

C J s

D J kg^{-1}

18 Which one of the following quantities is represented by the shaded area on the graph? (Ignore air resistance.)

A the kinetic energy per kilogram of the satellite at an altitude of 600 km

B the loss in gravitational potential energy of the satellite

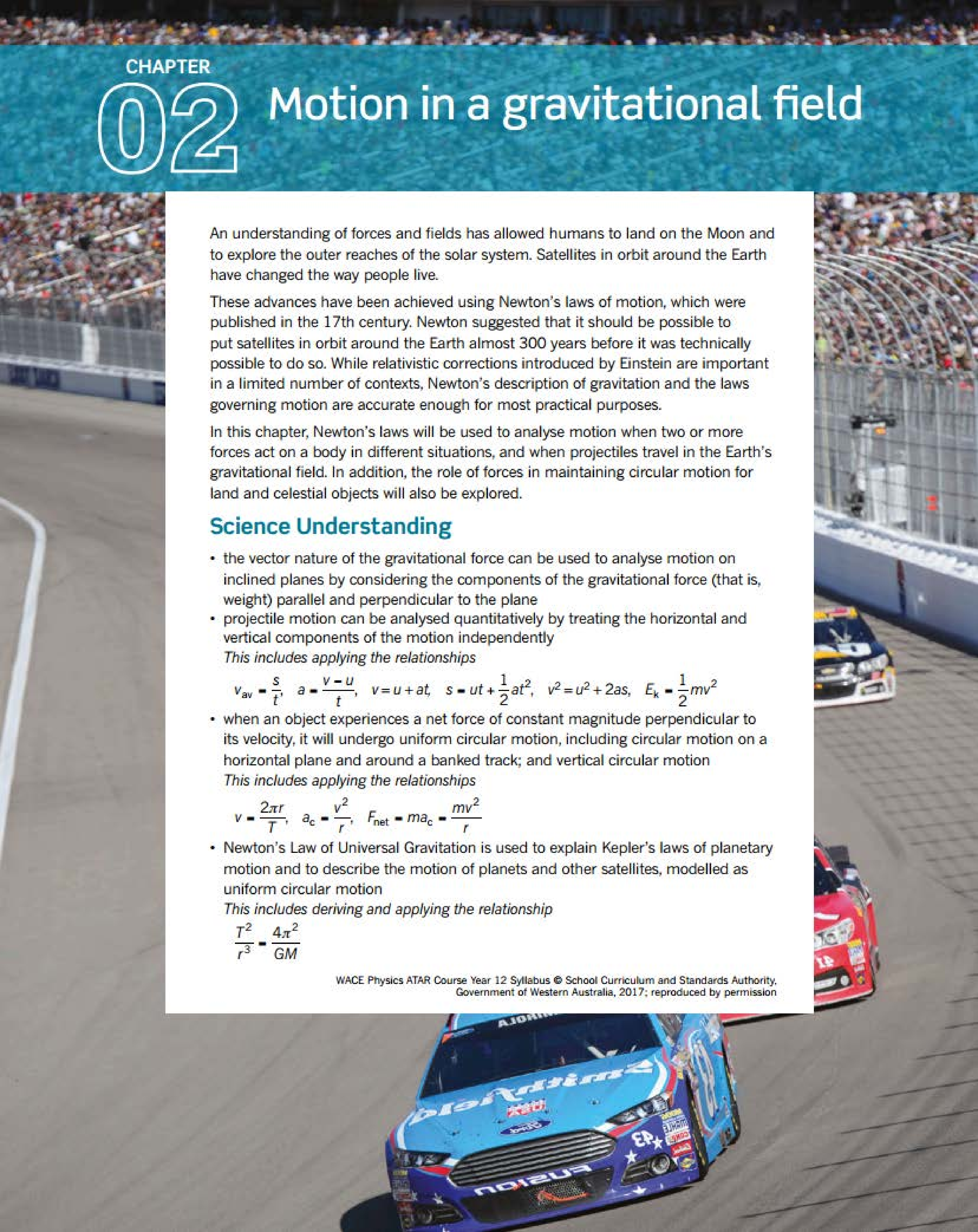
C the loss in gravitational potential energy per kilogram of the satellite as it falls to the Earth's surface

D the increase in gravitational potential energy of the satellite as it falls to the Earth's surface

19 How much kinetic energy does the satellite gain as it travels from an altitude of 600 km to an altitude of 200 km?

20 In reality, would the satellite gain the amount of kinetic energy that you have calculated in Question 19? Explain your answer.


ALMIROLA
Smithfield
43
FUSION
SMITH
41
Motorcraft
BLANEY
CAMRY



CHAPTER 02

Motion in a gravitational field

An understanding of forces and fields has allowed humans to land on the Moon and to explore the outer reaches of the solar system. Satellites in orbit around the Earth have changed the way people live.

These advances have been achieved using Newton's laws of motion, which were published in the 17th century. Newton suggested that it should be possible to put satellites in orbit around the Earth almost 300 years before it was technically possible to do so. While relativistic corrections introduced by Einstein are important in a limited number of contexts, Newton's description of gravitation and the laws governing motion are accurate enough for most practical purposes.

In this chapter, Newton's laws will be used to analyse motion when two or more forces act on a body in different situations, and when projectiles travel in the Earth's gravitational field. In addition, the role of forces in maintaining circular motion for land and celestial objects will also be explored.

Science Understanding

- the vector nature of the gravitational force can be used to analyse motion on inclined planes by considering the components of the gravitational force (that is, weight) parallel and perpendicular to the plane
- projectile motion can be analysed quantitatively by treating the horizontal and vertical components of the motion independently
  This includes applying the relationships
  $$v_{av} = \frac{s}{t}, \quad a = \frac{v-u}{t}, \quad v = u + at, \quad s = ut + \frac{1}{2}at^2, \quad v^2 = u^2 + 2as, \quad E_k = \frac{1}{2}mv^2$$
- when an object experiences a net force of constant magnitude perpendicular to its velocity, it will undergo uniform circular motion, including circular motion on a horizontal plane and around a banked track; and vertical circular motion
  This includes applying the relationships
  $$v = \frac{2\pi r}{T}, \quad a_c = \frac{v^2}{r}, \quad F_{net} = ma_c = \frac{mv^2}{r}$$
- Newton's Law of Universal Gravitation is used to explain Kepler's laws of planetary motion and to describe the motion of planets and other satellites, modelled as uniform circular motion
  This includes deriving and applying the relationship
  $$\frac{T^2}{r^3} = \frac{4\pi^2}{GM}$$

2.1 Inclined planes

If you have ever ridden a bike down a steep hill, you will know that on your way down you will accelerate and be travelling quite fast by the end. An external observer would see you travelling in both the relative horizontal and vertical directions. This motion down the street is an example of motion on an inclined plane, and this chapter aims to quantify your intuition in calculating how Newton's laws affect a body on an inclined plane. The steepest residential street in the world, Baldwin Street, is located in Dunedin, New Zealand (Figure 2.1.1).

FIGURE 2.1.1 The steepest residential street in the world. Its upper level is surfaced in concrete—bitumen would flow down the slope in warmer temperatures.

In Year 11 you looked at Newton's laws of motion in one dimension and how an object will behave with different forces applied. These laws can also be applied in two dimensions to analyse the problem of inclined planes.

REVISION

Newton's third law

Newton's third law states that when one body exerts a force on another body (an action force), the second body exerts an equal force in the opposite direction on the first body (the reaction force):

$F_{\text{on A by B}} = -F_{\text{on B by A}}$

To simplify the notation, this text will use the convention to mean $F_{AB} = F_{\text{on A by B}}$.

Hence the first subscript always shows the body experiencing the force.

It is important to note that action–reaction pairs can never be added together, because they act on different bodies (Figure 2.1.2). The forces in an action–reaction pair:

- are the same magnitude (size)
- act in opposite directions
- are exerted on two different objects.

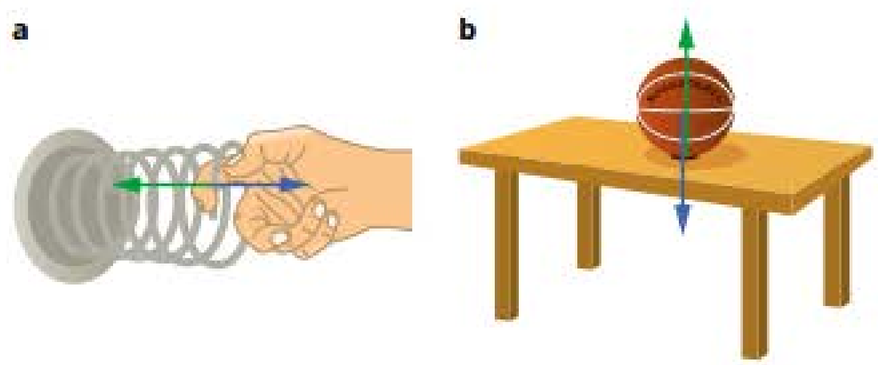


FIGURE 2.1.2 (a) An action–reaction pair. The hand pulls on the spring and the spring pulls back on the hand with an equal and opposite force. (b) Not an action–reaction pair. This is because the force due to gravity and the normal reaction force both act on the same object, the basketball.

THE NORMAL FORCE

For inclined planes, one reaction force in particular deserves an in-depth analysis. When an object exerts a force on a surface, the surface exerts a reaction force on the object that is normal (at right angles) to the surface.

For example, the block in Figure 2.1.3(a) exerts a force on the surface because it is attracted towards the centre of the Earth by gravity. The surface exerts a normal reaction force on the block. The weight F_g is thus balanced by F_N, as shown in the figure. There is no net force on the block, and so Newton's first law applies and the object remains stationary.

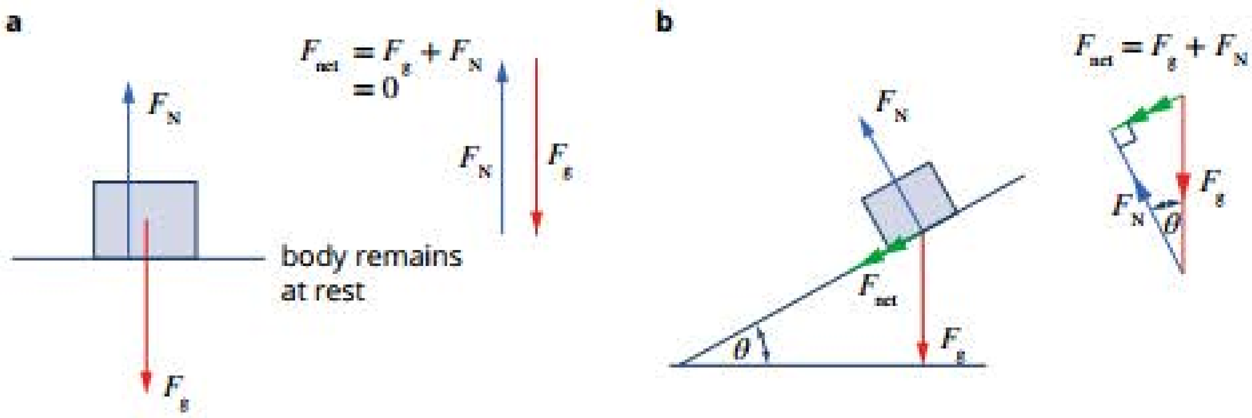


FIGURE 2.1.3 (a) Block on level surface: the net force is zero as F_N and F_g cancel. (b) Block on incline: $F_N = F_g \cos\theta$, and the net force is given by $F_{net} = F_g + F_N$ added as vectors.

On an **inclined plane**, F_N is at an angle to F_g. There is a net force down the slope and the block accelerates as predicted by Newton's second law.

Another way of viewing the forces along the inclined plane is to resolve the weight vector into two components: one perpendicular (at right angles) to the slope, and one parallel to the slope, as shown in Figure 2.1.4. The component perpendicular to the surface is balanced by the normal force F_N. This perpendicular component has a magnitude of $F_g \cos\theta$. Given this, it can be seen that on an inclined plane, the normal force will be less than the weight of the body, and will decrease as the angle of inclination increases. As indicated in Figure 2.1.3(b), the component of the weight of the object directed down the slope is the force responsible for the acceleration down the slope.

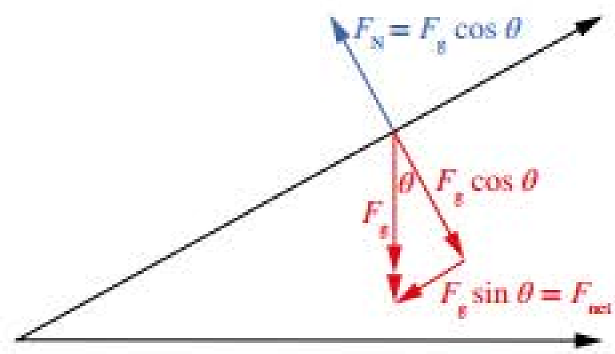
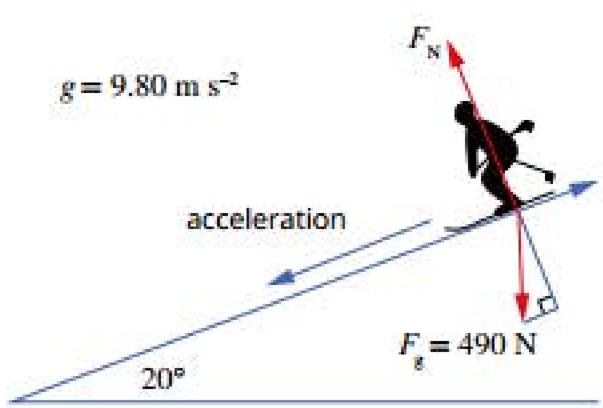


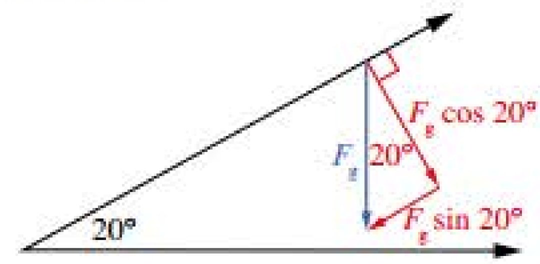
FIGURE 2.1.4 Block on an incline: the weight force can be resolved into a force perpendicular to the surface and a force parallel to the surface.

Worked example 2.1.1

INCLINED PLANES

A skier of mass 50 kg is skiing down an icy slope that is inclined at 20° to the horizontal. Assume that friction is negligible and that the acceleration due to gravity is $9.80\,m\,s^{-2}$.



a Determine the components of the weight of the skier perpendicular to the slope and parallel to the slope.	
Thinking	**Working**
Draw a sketch including the values provided.	
Resolve the weight into a component perpendicular to the slope.	The perpendicular component $F_{\perp} = F_g \cos 20°$ $= 490 \cos 20°$ $= 460\,\text{N}$
Resolve the weight into a component parallel to the slope. This is the net force.	The parallel component $F_{net} = F_g \sin 20°$ $= 490 \sin 20°$ $= 168\,\text{N}$
b Determine the normal force that acts on the skier.	
Thinking	**Working**
The normal force is equal in magnitude to the perpendicular component of the weight force.	$F_N = 460\,\text{N}$
c Calculate the acceleration of the skier down the slope.	
Thinking	**Working**
Apply Newton's second law. The net force along the plane is the component of the weight parallel to the slope.	$a = \frac{F_{net}}{m}$ $= \frac{168}{50}$ $= 3.36\,\text{m}\,\text{s}^{-2}$ down the slope

Worked example: Try yourself 2.1.1

INCLINED PLANES

A heavier skier of mass 85 kg travels down the same icy slope inclined at 20° to the horizontal. Assume that friction is negligible and that the acceleration due to gravity is $9.80\,\text{m}\,\text{s}^{-2}$.

a Determine the components of the weight of the skier perpendicular to the slope and parallel to the slope.

b Determine the normal force that acts on the skier.

c Calculate the acceleration of the skier down the slope.

Aside from rounding differences, the acceleration calculated in the Worked example and Try yourself questions were equal. That is because acceleration is independent of the mass of the object:

$$a = \frac{F_{net}}{m} = \frac{mg\sin\theta}{m} = g\sin\theta$$

That is, acceleration down a frictionless inclined plane is affected only by the gravitational field strength and the angle of the plane.

THE FRICTIONAL FORCE

In real life, the inclined planes you are familiar with are not smooth, and so you cannot neglect the force of friction. The resistance that an object encounters when moving over or through another medium is termed **friction**. This force always acts in the direction opposite to the motion of the object (either up or down the inclined plane). For example, Figure 2.1.5(a) shows the force diagram of a car travelling down a hill at constant velocity with friction acting on it. At constant velocity, Newton's second law states that all forces on the object must be balanced. The forces acting parallel to the incline (the frictional force F_f and the parallel component of the weight $F_g \sin\theta$) will therefore be equal in magnitude; that is, $F_f = F_g \sin\theta$.

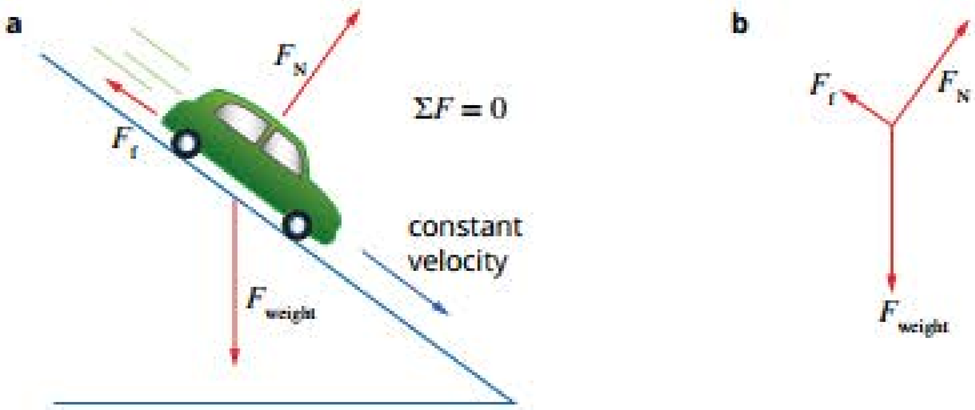


FIGURE 2.1.5 (a) A car rolls with constant velocity down an incline. With friction present, all forces acting on the car are balanced and thus it is either stationary or rolls with constant velocity. (b) Vector diagram shows the forces acting on the car.

STRATEGIES FOR SOLVING INCLINED PLANE PROBLEMS

Where forces on a body are given, Newton's laws can be applied.

Two questions can be asked:

1 Is the object described as stationary or travelling at constant velocity? In this case, $F_{net} = 0$.

2 Is the object accelerating? In this case, $F_{net} = ma$.

For coplanar forces that are not aligned, resolve forces into components. Once these forces have been resolved into two perpendicular directions—the components parallel to, and the components perpendicular to the inclined plane—the net force in each direction can be determined.

Typically, the components perpendicular to the incline will consist of the normal force, F_N, and the component of the weight force perpendicular to the incline, $F_g \cos\theta$.

The components parallel to the incline will consist of the frictional force (if present), F_f, and the component of the weight force parallel to the incline, $F_g \sin\theta$.

Newton's second law can be used to find the acceleration of an object. The other equations of motion may then be used to find quantities such as displacement and final velocity.

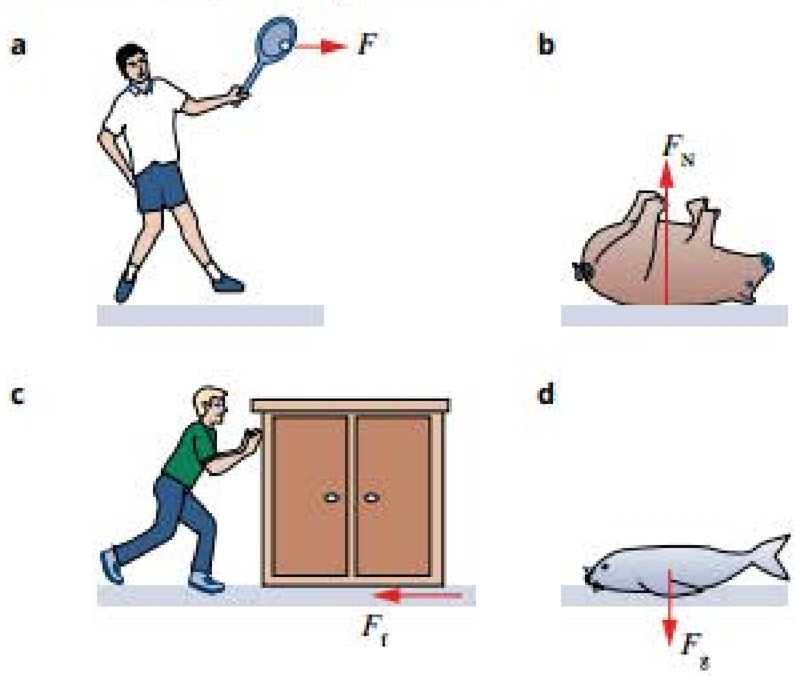
2.1 Review

SUMMARY

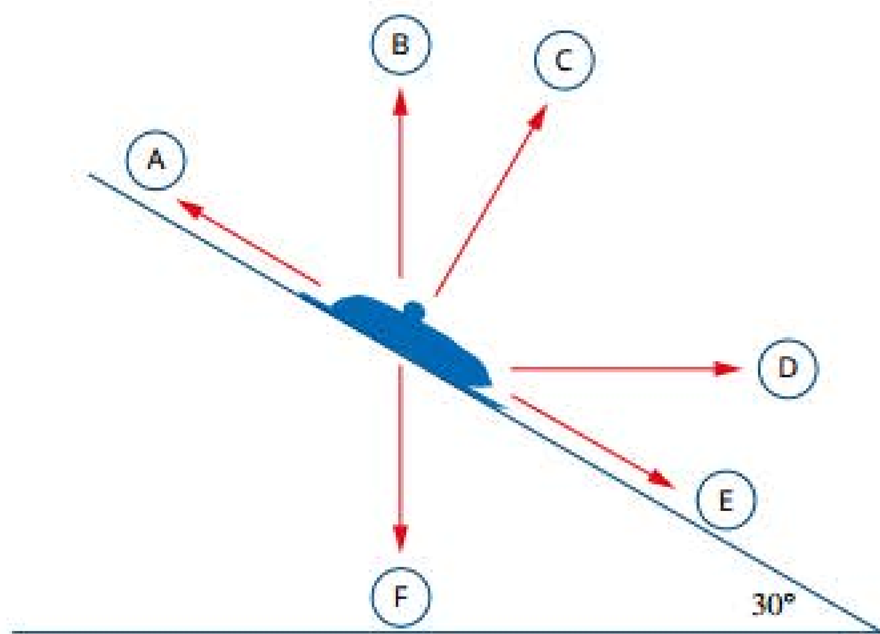
- Newton's third law states that when one body exerts a force on another body (an action force), the second body exerts an equal force in the opposite direction on the first (the reaction force): $F_{AB} = -F_{BA}$. The forces in an action–reaction pair are the same magnitude, act in opposite directions and are exerted on two different objects.
- A normal reaction force, F_N, acts between an object and a surface, at right angles to the surface.
  - On a horizontal surface, $F_N = F_g$.
  - On an inclined surface, F_N is equal and opposite to the component of the weight force acting perpendicular to the plane: $F_N = F_g \cos\theta$.
- The net force acting on an object on a plane inclined at an angle θ is $F_{net} = F_g \sin\theta$ when friction is negligible. It is the net force that causes acceleration down the incline.
- If the force of friction is present, its direction will be opposite to that of the object's motion.

KEY QUESTIONS

1 Complete each of these force diagrams, showing the reaction pair to the action force that is shown. For each force that you draw, state what the force is acting on and what is providing the force.



2 Kirsty is riding in a bobsled that is sliding down a snow-covered hill with a slope of 30° to the horizontal. The total mass of the sled and Kirsty is 100 kg. Initially the brakes are on and the sled moves down the hill with a constant velocity.



a Which one of the arrows (A–F) best represents the direction of the frictional force acting on the sled?

b Which one of the arrows (A–F) best represents the direction of the normal force acting on the sled?

c Calculate the net frictional force acting on the sled.

d Kirsty then releases the brakes and the sled accelerates. What is the magnitude of her initial acceleration?

e In a second run, Kirsty rides the bobsled down the same slope but with the brakes off, so friction can be ignored. The bobsled now has an extra passenger, so that its total mass is now 140 kg. How will this affect the acceleration of the bobsled?

3 Which one or more of the following statements describe(s) the forces acting on an object on a stationary level surface?

A The normal force is always perpendicular to the surface.

B The normal force is always equal in magnitude to the weight.

C The normal force and the weight are an action–reaction pair.

D The normal force and the weight both act on the object, and add to equal zero.

4 Which of the following statements describes the forces acting on an object on a plane inclined at an angle θ?

A The normal force is always perpendicular to the surface.

B The normal force is equal in magnitude to the weight.

C The normal force and the weight cancel out.

D In the absence of friction, a component of the normal force causes the object to accelerate down the slope.

5 A skier of mass 100 kg is skiing down an icy slope that is inclined at 30° to the horizontal. Assume that friction and drag are negligible.

a Determine the component of the weight of the skier perpendicular to the incline.

b Determine the component of the weight of the skier parallel to the incline.

c Determine the acceleration of the skier down the slope.

6 A smooth steel ramp 2.00 m long is inclined at 25° to the horizontal. A steel ball of mass 0.100 kg is rolled down the ramp followed by a steel ball of mass 0.200 kg.

a Determine the normal force acting on each ball.

b Determine the acceleration of each ball down the ramp.

The inclination of the ramp is now increased to 70° and the two balls are separately rolled down it once more.

c Determine again the normal force acting on the two balls.

d Determine the acceleration of the two balls down the ramp.

e Compare your answers for (a) and (b) for the two balls. Then compare your answers for (a) and (b) with those of (c) and (d). What can you say about the effects of increasing mass and angle on the normal force and the acceleration of an object on an inclined plane?

2.2 Projectiles launched horizontally

A **projectile** is any object that is thrown or projected into the air and is moving freely; that is, it has no power source (such as a rocket engine or propeller) driving it. A netball as it is passed, a cricket ball that is hit for six, and an aerial skier flying through the air are all examples of projectiles. People have long argued about the path that projectiles follow; some thought they were circular or had straight sections. It is now known that if projectiles are not launched vertically and if air resistance is ignored, they move in smooth parabolic paths, like that shown in Figure 2.2.1. This section considers projectiles that are launched horizontally and uses Newton's laws to solve problems involving this type of motion.

FIGURE 2.2.1 A multi-flash photograph of a golf ball that has been bounced on a hard surface. The ball moves in a parabolic path.

PROJECTILE MOTION

It is a very common misconception that when a projectile, such as a netball, travels forwards through the air, it has a forwards force acting on it. This is incorrect. There will have been some forwards force acting as the projectile is launched, but once the projectile is released, this forwards force is no longer acting. The continued motion of the object is explained by Newton's first law.

In fact, if air resistance is ignored, the only force acting on a projectile during its flight is its weight, which is the force due to gravity, F_g. This force is constant and always directed vertically downwards. This causes the projectile to continually deviate from a straight-line path to follow a parabolic path, as seen in Figure 2.2.2. This motion is known as free-fall.

Projectile motion is quite complex compared to straight-line motion. It must be analysed by considering the different components—horizontal and vertical—of the actual motion. The vertical and horizontal components of the motion are independent of each other and must be treated separately.

Given that the only force acting on a projectile is the force due to gravity, F_g, it follows that the projectile must have a vertical acceleration of $9.80\,m\,s^{-2}$ downwards throughout its motion.

In the vertical direction, a projectile accelerates due to the force of gravity, that is, at a rate of $9.80\,m\,s^{-2}$ downwards.

In the horizontal direction, a projectile has a uniform velocity since there are no forces acting in this direction (if air resistance is ignored). So, the horizontal acceleration is zero.

FIGURE 2.2.2 The centre of mass of the combined motorcycle and rider travels in a parabolic path as bike and rider fly through the air.

PROJECTILES LAUNCHED HORIZONTALLY

Projectiles can be launched at any angle. The launch velocity needs to be resolved into vertical and horizontal components using trigonometry in order to complete most problems. For projectiles launched horizontally, calculating the vector components of the launch velocity is easy to do. That's because the initial vertical velocity is zero (but increases during the flight). The horizontal velocity is constant and is equal to the launch velocity. This can be verified using trigonometric ratios and a launch angle of 0°.

Tips for solving projectile motion problems

1 Construct a diagram showing the projectile's motion to set the problem out clearly. Write out the information supplied for the horizontal and vertical components separately.

2 In the horizontal direction, the velocity, v, of the projectile is constant, so the only formula needed is $v_{av} = \frac{s}{t}$.

3 In the vertical direction, the projectile is moving with a constant acceleration ($9.80\,m\,s^{-2}$ down), and so the equations of motion for uniform acceleration must be used. These include:

$$v = u + at$$

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

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

PHYSICSFILE

Cartoon physics

It is easy to get the wrong idea about projectile motion when you watch cartoon characters running or driving off cliffs. In many cartoons, the character leaves the cliff and travels horizontally outwards, stopping in mid-air (Figure 2.2.3). Once they realise where they are, they immediately fall vertically downwards. This is not what happens in reality. The character should start falling in a smooth parabolic arc as soon as they leave the cliff top.

FIGURE 2.2.3 Many misconceptions can arise from what is shown in cartoons. In real life, this car would start falling as soon as it leaves the cliff top and it would travel in a parabolic arc.

4 In the vertical direction it is important to clearly specify whether up or down is the positive or negative direction. Either choice will work just as effectively. The same convention needs to be used consistently throughout each problem.

5 If a projectile is launched horizontally, its horizontal velocity throughout the flight is the same as its initial velocity.

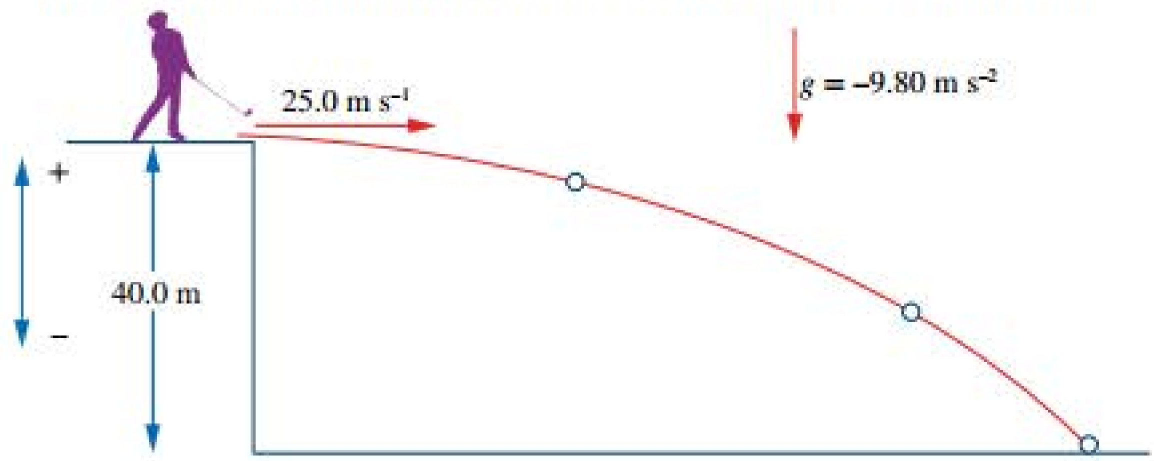
6 Pythagoras' theorem can be used to determine the actual speed of the projectile at any point.

7 If the velocity of the projectile is required, it is necessary to provide a direction with respect to the horizontal plane as well as the speed of the projectile.

Worked example 2.2.1

PROJECTILE LAUNCHED HORIZONTALLY

A golf ball of mass 150 g is hit horizontally from the top of a 40.0 m high cliff with a speed of $25.0\,\text{m s}^{-1}$. In your working, use $g = 9.80\,\text{m s}^{-2}$ and ignore air resistance.



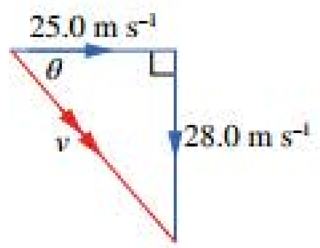
a Calculate the time that the ball takes to land.

Thinking	Working
Let the downwards direction be negative. Write out the information relevant to the vertical component of the motion. Note that the instant the ball is hit, it is travelling only horizontally, so its initial vertical velocity is zero. Note also that the ball ends up travelling 40 m below its original point so the vertical displacement is negative.	Down is negative. Vertically: $s = -40.0\,\text{m}$ $u = 0\,\text{m s}^{-1}$ $v = ?$ $a = -9.80\,\text{m s}^{-2}$ $t = ?$
In the vertical direction, the ball has constant acceleration, so use equations for uniform acceleration. Select the equation that best fits the information you have.	$s = ut + \frac{1}{2}at^2$
Substitute values, rearrange and solve for *t*.	$-40.0 = 0 + \frac{1}{2}(-9.80)t^2$ $t = \sqrt{\frac{-40.0}{-4.90}}$ $t = 2.86\,\text{s}$

b Calculate the distance that the ball travels from the base of the cliff, i.e. the range of the ball.

Thinking	Working
Write out the information relevant to the horizontal component of the motion. As the ball is hit horizontally, the initial speed gives the horizontal component of the velocity throughout the flight.	Horizontally: $v = 25.0\,\text{m s}^{-1}$ $t = 2.86\,\text{s}$ from part (a) $s = ?$
Select the equation that best fits the information you have.	$v_{av} = \frac{s}{t}$
Substitute values, rearrange and solve for *s*.	$25.0 = \frac{s}{2.86}$ $s = 25.0 \times 2.86$ $s = 71.5\,\text{m}$ Note that the mass of the ball does not affect its motion, as with all objects in projectile motion or in free-fall.

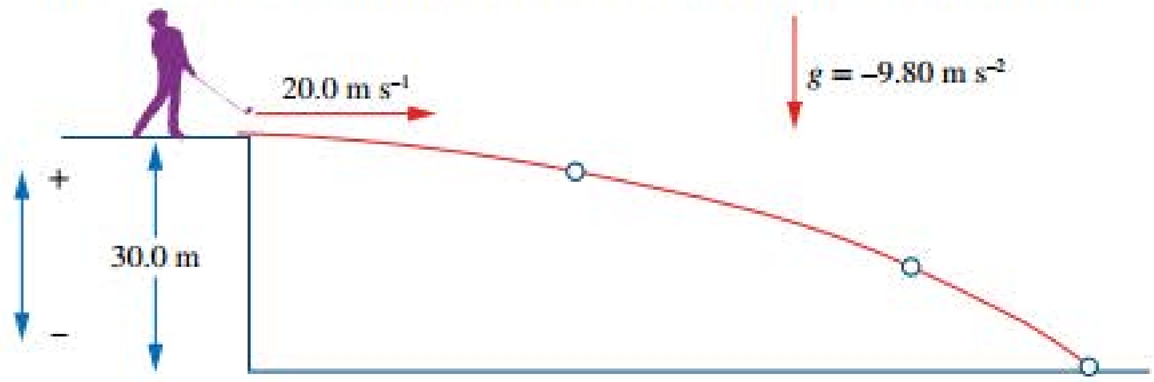
c Calculate the velocity of the ball as it lands.

Thinking	Working
Find the horizontal and vertical components of the ball's speed as it lands. Write out the information relevant to both the vertical and horizontal components.	Horizontally: $u = v = 25.0\,\text{m s}^{-1}$ Vertically, with down as positive: $s = -40.0\,\text{m}$ $u = 0$ $v = ?$ $a = -9.80\,\text{m s}^{-2}$ $t = 2.86\,\text{s}$
To find the final vertical speed, use the equation for uniform acceleration that best fits the information you have.	$v = u + at$
Substitute values, rearrange and solve for the variable you are looking for, in this case *v*.	Vertically: $v = 0 + (-9.80) \times 2.86$ $v = -28.0\,\text{m s}^{-1}$
Add the components as vectors.	
Use Pythagoras' theorem to work out the resultant speed of the ball, *v*.	$v = \sqrt{v_h^2 + v_v^2}$ $= \sqrt{(25.0)^2 + (-28.0)^2}$ $= \sqrt{1409}$ $v = 37.5\,\text{m s}^{-1}$
Use trigonometry to find the angle, *θ*.	$\theta = \tan^{-1} \frac{28.0}{25.0}$ $= 48.2°$
Indicate the velocity with a magnitude and direction relative to the horizontal.	The final velocity of the ball is $37.5\,\text{m s}^{-1}$ at 48.2° below the horizontal.

Worked example: Try yourself 2.2.1

PROJECTILE LAUNCHED HORIZONTALLY

A golf ball of mass 100 g is hit horizontally from the top of a 30.0 m high cliff with a speed of $20.0\,m\,s^{-1}$. In your working, use $g = 9.80\,m\,s^{-2}$ and ignore air resistance.



a Calculate the time that the ball takes to land.
b Calculate the distance that the ball travels from the base of the cliff, i.e. the range of the ball.
c Calculate the velocity of the ball as it lands.

PHYSICSFILE

Aerodynamic design

In the track and field event of javelin, the aerodynamic shape of the javelin was too successful. The javelin was originally designed to reduce the drag force so that the athletes could throw further. This was not a problem until the 1980s, when athletes began to throw so far that runners competing in track events were endangered. The design of the javelin was changed and it was made more snub-nosed to increase drag and reduce distances thrown (this can be seen in Figure 2.2.4). In 1983, the world record was 104.8 m. In 1986 with the modified design, the world record dropped to 85.7 m.

FIGURE 2.2.4 Australia's Kathryn Mitchell in action during the women's javelin final at the London 2012 Olympic Games.

THE EFFECTS OF AIR RESISTANCE

The interaction between a projectile and the air can have a significant effect on the motion of the projectile, particularly if the projectile has a large surface area and a relatively low mass. If you try to throw an inflated party balloon, it will not travel very far compared to throwing a marble at the same speed.

The size of the **air resistance** or drag force that acts on an object as it moves depends on factors such as:

- the speed of the object, v. The faster an object moves, the greater the drag force becomes.
- the cross-sectional area of the object in its direction of motion. Greater area means greater drag.
- the aerodynamic shape of the object. A more streamlined shape experiences less drag.
- the density of the air. Higher air density means greater drag.

When a pilot drops a supply parcel from a plane, the drag force from the air acts in the opposite direction to the parcel's velocity. If the parcel was dropped on the Moon, where there is no air and hence no air resistance, this would not be a factor and the parcel would continue its horizontal motion and would remain directly below the plane as it fell.

Figure 2.2.5 shows a supply parcel being dropped from a plane moving at a constant velocity. If air resistance is ignored, the parcel falls in the parabolic arc shown by the darker blue curved line in Figure 2.2.5(a). It would continue moving horizontally at the same rate as the plane; that is, as the parcel falls it would stay directly beneath the plane until it hits the ground. The effect of air resistance is shown by the light-blue curved path in Figure 2.2.5(a). Air resistance (or drag) is a retarding force and it acts in a direction that is opposite to the motion of the projectile. Air resistance makes the parcel fall more slowly, over a shorter path. If air resistance is taken into account, there are now two forces acting, as shown in Figure 2.2.5(b): weight, F_g, and air resistance, F_a. Therefore, the resultant force, F_{net}, that acts on the projectile is not vertically down and nor is its acceleration.

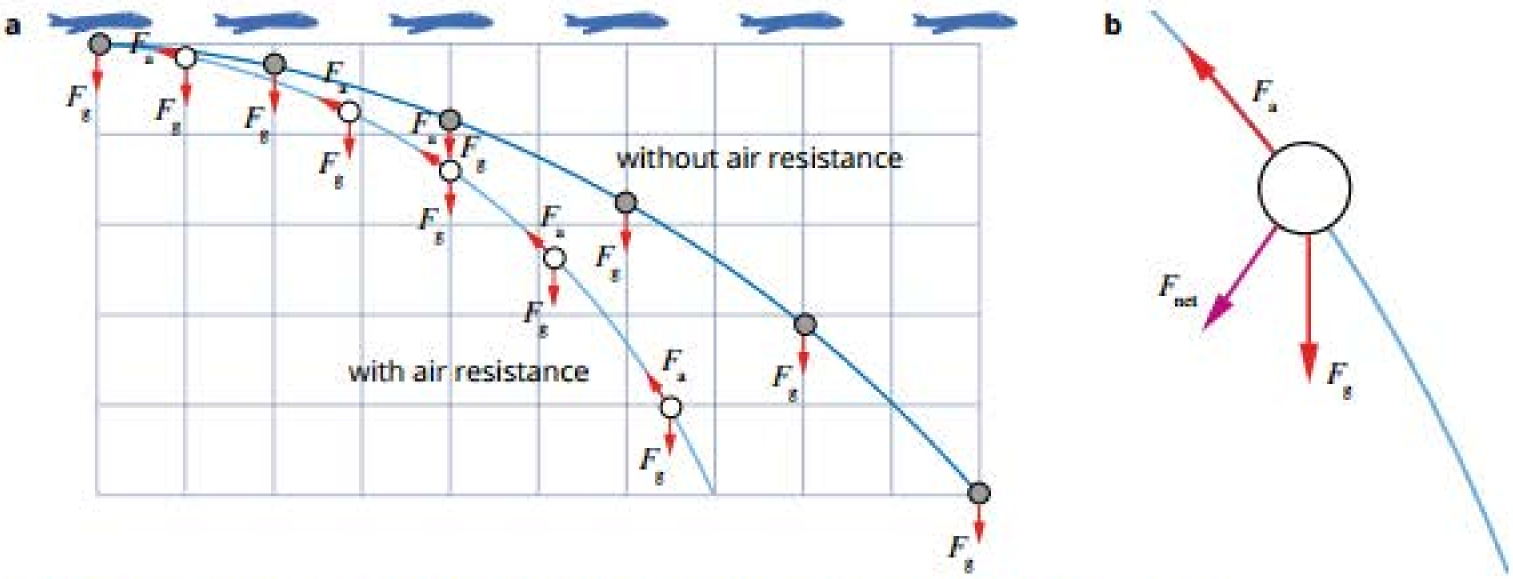


FIGURE 2.2.5 (a) The paths of a supply parcel dropped from a plane, with and without air resistance. (b) When air resistance is acting, the net force on the parcel is not vertically down.

SOLVING PROJECTILE PROBLEMS USING ENERGY

Using conservation of energy for projectiles

In addition to calculating the projectile motion of an object using equations of linear motion and acceleration, motion can be modelled by using the principle of conservation of energy. When a projectile is launched, it is given initial kinetic energy $E_k = \frac{1}{2}mv^2$. As the object falls down to the ground, the vertical component of its velocity increases because of the acceleration due to gravity of $9.80\,\text{m}\,\text{s}^{-2}$ downwards, and therefore its kinetic energy increases.

According to the **law of conservation of energy** (and neglecting air resistance), this increase in kinetic energy is due to work done by the gravitational field, and as a result the gravitational potential energy $E_g = mgh$ decreases. Since the total **mechanical energy** is **conserved,** $\frac{1}{2}mv_i^2 + mgh_i = \frac{1}{2}mv_f^2 + mgh_f$. From this relationship, you can calculate the final speed of the object given the initial speed, the initial height, and the final height of the object above the ground. Note that the direction of the final speed cannot be determined using this strategy.

Tips for solving projectile motion problems considering energy

1 Calculate the initial kinetic energy of the projectile using $E_{k\,i} = \frac{1}{2}mv_i^2$.

2 Calculate the initial gravitational potential energy of the projectile using $E_{p\,i} = mgh_i$.

3 Add the initial gravitational potential energy to the initial kinetic energy. This determines E_t, the total mechanical energy at the beginning of the projectile motion.

4 Calculate the final gravitational potential energy, using the final height of the object and $E_{p\,f} = mgh_f$.

5 Finally, calculate the final speed of the projectile, using the initial total mechanical energy equal to the sum of the final kinetic and potential energy: $E_t = E_{k\,f} + E_{p\,f}$.

Worked example 2.2.2

PROJECTILE LAUNCHED HORIZONTALLY—PROBLEM SOLVING WITH WORK AND ENERGY

A ball of mass 500 g is tossed horizontally at $5.00\,\text{m s}^{-1}$ from a height of 1.50 m. Using $g = 9.80\,\text{m s}^{-2}$ and ignoring air resistance, calculate the speed of the ball as it lands.

Thinking	Working
Start by calculating the initial kinetic energy of the object, using $E_{\text{k initial}} = \frac{1}{2}mv_i^2$.	$E_{\text{k i}} = \frac{1}{2}mv_i^2$ $= \frac{1}{2} \times 0.500 \times 5.00^2$ $= 6.25\,\text{J}$
Calculate the initial potential energy of the object, using $E_{\text{p i}} = mgh_i$.	$E_{\text{p i}} = mgh_i$ $= 0.500 \times 9.80 \times 1.50$ $= 7.35\,\text{J}$
Calculate the total mechanical energy.	$E_t = E_{\text{k i}} + E_{\text{p i}}$ $= 6.25 + 7.35$ $= 13.6\,\text{J}$
Finally, use $E_{\text{k f}} = \frac{1}{2}mv_f^2$ to work out the speed.	$E_t = E_{\text{k f}} + E_{\text{p f}}$ $13.6 = \frac{1}{2}mv^2 + mgh$ $= \frac{1}{2} \times 0.500 \times v^2 + mgh_f$ $= \frac{1}{2} \times 0.500 \times v^2 + 0$ $v^2 = 54.4$ $v = 7.38\,\text{m s}^{-1}$

Worked example: Try yourself 2.2.2

PROJECTILE LAUNCHED HORIZONTALLY—PROBLEM SOLVING WITH WORK AND ENERGY

A toy plane of mass 300 g is tossed horizontally at $6.00\,\text{m s}^{-1}$ from a height of 3.00 m. Using $g = 9.80\,\text{m s}^{-2}$ and ignoring air resistance, calculate the speed of the plane as it lands.

2.2 Review

SUMMARY

- If air resistance is ignored, the only force acting on a projectile is its weight, i.e. the force of gravity, F_g. This results in the projectile having an acceleration of $9.80\,m\,s^{-2}$ vertically downwards during its flight.
- Projectiles move in parabolic paths that can be analysed by considering the horizontal and vertical components of the motion.
- The following equations of motion for uniform acceleration must be used for the vertical component of the motion:

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

- The horizontal velocity of a projectile remains constant throughout its flight if air resistance is ignored. Therefore, the following equation for average velocity can be used for this component of the motion:

$$v_{av} = \frac{s}{t}$$

- Ignoring air resistance, the total energy of the projectile must always be conserved. Therefore, the following equations can be used:

$$E_k = \frac{1}{2}mv^2$$
$$E_g = mgh$$
$$E_{total} = \frac{1}{2}mv^2 + mgh$$

KEY QUESTIONS

For the following questions, assume that the acceleration due to gravity is $9.80\,m\,s^{-2}$ and ignore the effects of air resistance unless otherwise stated.

1 A boy throws a stone horizontally at $5\,m\,s^{-1}$ into a pond. Which one or more of the following statements best describe(s) the stone as it falls towards the pond?

A The only force acting on it is gravity.
B It travels in a circular path.
C There is a driving force acting on it.
D Its speed increases.

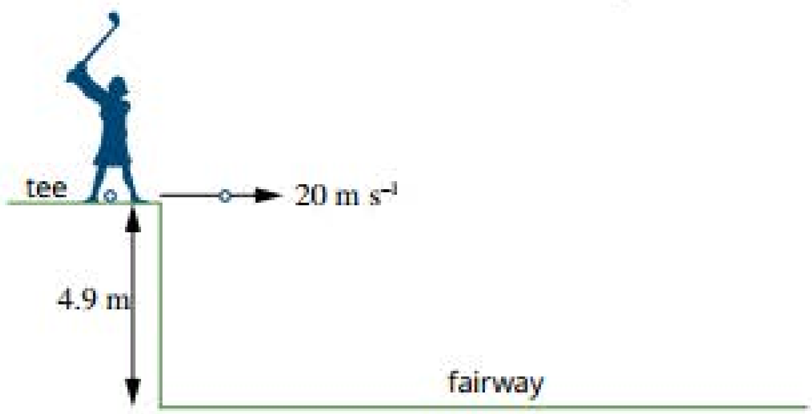
2 A marble travelling at $2.0\,m\,s^{-1}$ rolls off a horizontal bench and takes 0.75 s to reach the floor.

a How far does the marble travel horizontally before landing?
b What is the vertical component of the speed of the marble as it lands?
c What is the speed of the marble as it lands?

3 A skateboard of mass 3.00 kg and travelling at $4.00\,m\,s^{-1}$ rolls off a horizontal bench that is 1.20 m high.

a What is the magnitude and direction of the acceleration of the skateboard just before it lands?
b Using conservation of energy, what is the speed of the skateboard as it lands?

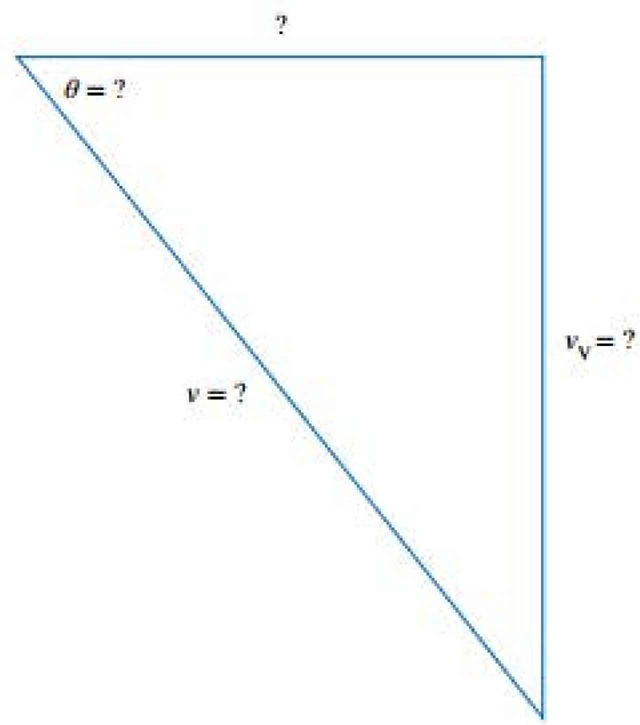
4 A golfer practising on a range with an elevated tee 4.9 m above the fairway is able to strike a ball so that it leaves the club with a horizontal velocity of $20\,m\,s^{-1}$.



a How long after the ball leaves the club will it land on the fairway?
b What horizontal distance will the ball travel before landing on the fairway?
c What is the acceleration of the ball 0.50 s after being hit?
d Calculate the speed of the ball 0.80 s after it leaves the club.
e With what speed will the ball hit the ground?

2.2 Review *continued*

5 Bill, a tourist, stands on top of a sea cliff that is 80.0 m high. Bill throws a 2.00 kg rock horizontally right at an unknown velocity into the sea. Luckily, Nicole, a physicist, happens to measure the speed of the rock from her boat below. She determines the speed of the rock to be 45.0 m s^{-1} when it is 3.00 m above the surface of the water.



a Using conservation of energy and the information provided, what is the initial velocity of Bill's throw?

b At what speed is the rock travelling as it hits the water?

6 In 1971, American astronaut Alan Shepard took a golf club to the Moon and hit a couple of balls. Which of the following answers is/are correct?

A The balls travelled in straight lines because there is no gravity.

B The balls travelled in parabolic arcs.

C The balls travelled much further than if they had been hit in an identical manner on Earth.

D The balls went into orbit.

7 Joe throws a hockey ball horizontally at 5 m s^{-1}. He then throws a polystyrene ball of identical dimensions at the same speed horizontally. If air resistance is taken into account, which of the balls will travel further? Why?

8 Two identical tennis balls X and Y are hit horizontally from a point 2.0 m above the ground with different initial speeds: ball X has an initial speed of 5.0 m s^{-1}, and ball Y has an initial speed of 10.0 m s^{-1}.

a Calculate the time it takes for ball X to strike the ground.

b Calculate the time it takes for ball Y to strike the ground.

c How much further than ball X does ball Y travel in the horizontal direction before bouncing?

9 An archer stands on top of a platform that is 20 m high and fires an arrow horizontally at 50 m s^{-1}.

a What is the speed of the arrow as it reaches the ground?

b At what angle is the arrow travelling as it reaches the ground, relative to the horizontal?

10 A bowling ball of mass 7.5 kg travelling at 10.0 m s^{-1} rolls off a horizontal table 1.0 m high.

a Calculate the ball's horizontal velocity just as it strikes the floor.

b What is the vertical velocity of the ball as it strikes the floor?

c Calculate the velocity of the ball as it reaches the floor.

d What time interval has elapsed between the ball leaving the table and striking the floor?

e Calculate the horizontal distance travelled by the ball as it falls.

f Draw a diagram showing the forces acting on the ball as it falls towards the floor.

2.3 Projectiles launched obliquely

The previous section looked at projectiles that were launched horizontally. A more common situation is projectiles that are launched obliquely (at an angle), by being thrown forwards and upwards at the same time. An example of an oblique launch is shooting for points in basketball, as shown in Figure 2.3.1. Once the ball is released, the only forces acting on it are gravity pulling it down to Earth and air resistance, which retards the ball's motion slightly.

In this section, the principles covering horizontal projectile motion will still apply, as described by Newton's first law.

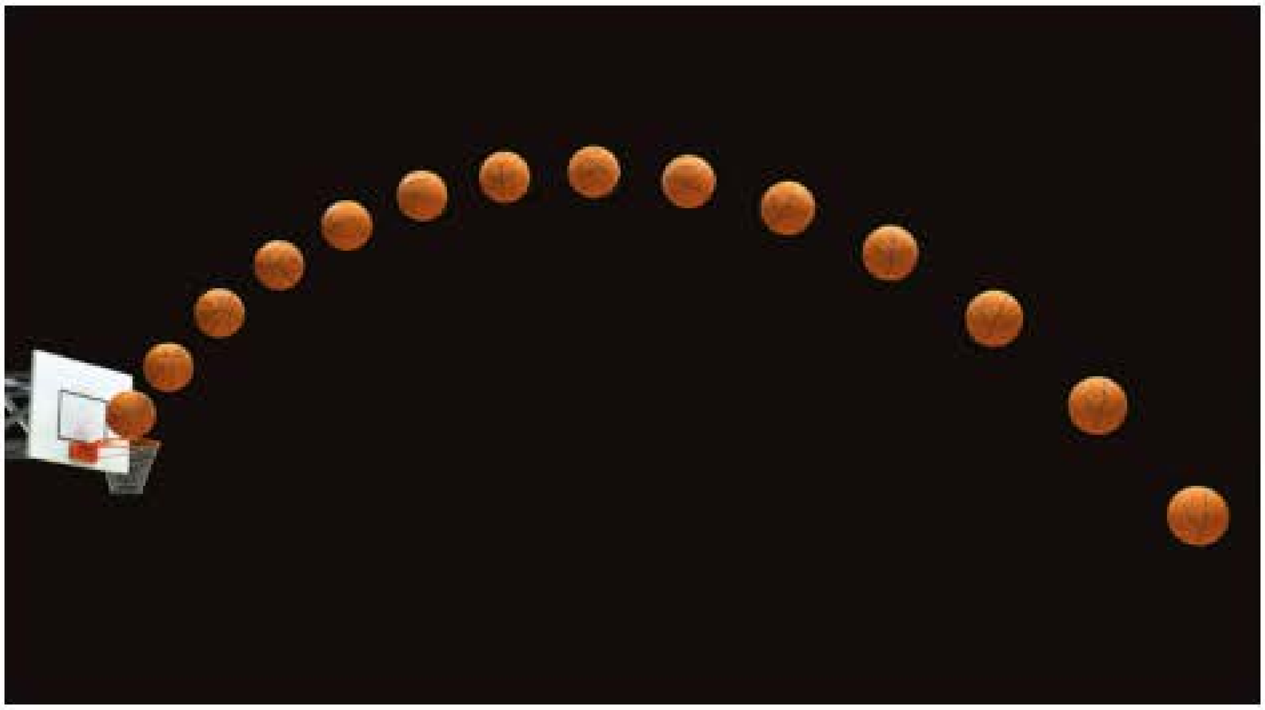


FIGURE 2.3.1 The basketball was thrown up towards the ring, travelling in a smooth parabolic path.

PROJECTILES LAUNCHED AT AN ANGLE

As with projectiles launched horizontally, if drag forces are ignored, the only force that is acting on a projectile that is launched at an angle to the horizontal is gravity, F_g.

Gravity acts vertically downwards and so it has no effect on the horizontal motion of a projectile. The vertical and horizontal components of the motion are independent of each other and must be treated separately.

In the vertical direction, a projectile accelerates due to the force of gravity; that is, at a rate of $9.80\,m\,s^{-2}$ downwards. The effect of the force due to gravity is that the vertical component of the projectile's velocity decreases as the projectile rises. It's vertical velocity is momentarily zero at the very top of the flight and then increases again as the projectile descends.

In the horizontal direction, a projectile has a uniform velocity since there are no forces acting in this direction (if air resistance is ignored).

Tips for problems involving projectile motion

General rules for solving problems involving projectile motion were given in the previous section—see pages 47–8 for a reminder.

If a projectile is launched at an angle to the horizontal, trigonometry can be used to find the initial horizontal and vertical components of velocity. Pythagoras' theorem can be used to determine the actual velocity of the projectile at any point, as well as a direction with respect to the horizontal plane.

Worked examples 2.3.1 and 2.3.2 will show you how this is done.

PHYSICSFILE

Trebuchets

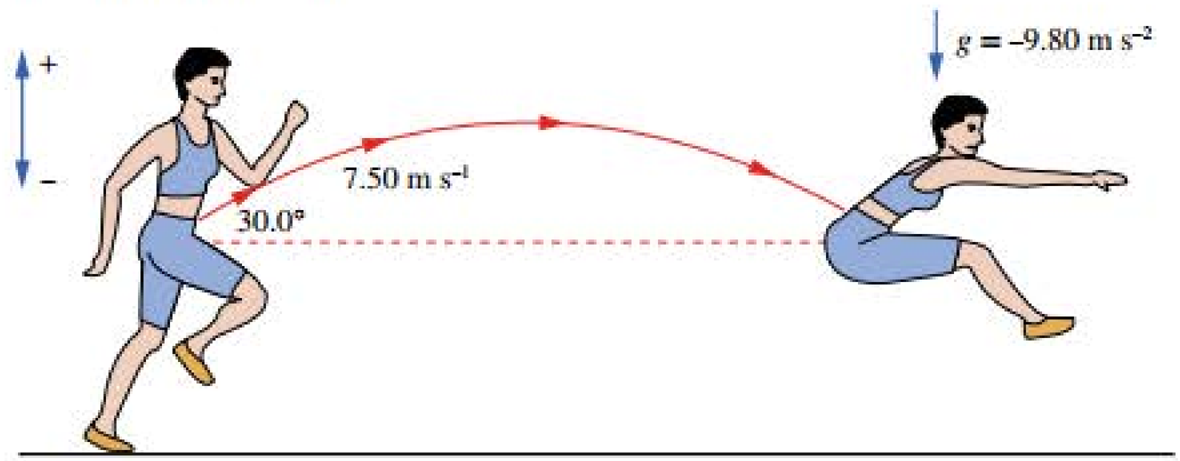
A frequently used siege weapon from the Middle Ages, a trebuchet is an example of projectile motion in action (Figure 2.3.2). The force of gravity is used to pull down a counterweight that acts on a lever, capable of launching missiles up to 150 kg. The largest known trebuchet in history, named Warwolf, was ordered to be built by King Edward I of England during the siege of Stirling Castle in the Scottish Wars of Independence. When disassembled, the weapon filled 30 wagons. The sight of it was so intimidating that it prompted the Scots to surrender. However, Edward sent the truce party back inside the castle, instead wishing to witness the power of the siege weapon.

FIGURE 2.3.2 The formidable counterweight trebuchet first came into use in the 12th century and was used by armies to fling projectiles at enemy fortifications

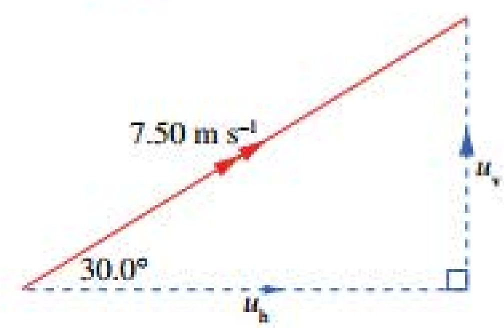
Worked example 2.3.1

LAUNCH AT AN ANGLE

A 65 kg athlete in a long-jump event leaps with a velocity of $7.50\,m\,s^{-1}$ at an angle of 30.0° to the horizontal.



For the following questions, treat the athlete as a point mass, ignore air resistance and use $g = 9.80\,m\,s^{-2}$.

a What is the athlete's velocity at the highest point?	
Thinking	**Working**
First find the horizontal and vertical components of the initial speed.	 Using trigonometry: $u_h = 7.50 \cos 30.0°$ $= 6.50\,m\,s^{-1}$ $u_v = 7.50 \sin 30.0°$ $= 3.75\,m\,s^{-1}$
Projectiles that are launched obliquely move horizontally at the highest point. The vertical component of the velocity at this point is therefore zero. The actual velocity is given by the horizontal component of the velocity throughout the motion.	At maximum height: $v = 6.50\,m\,s^{-1}$ horizontally to the right.

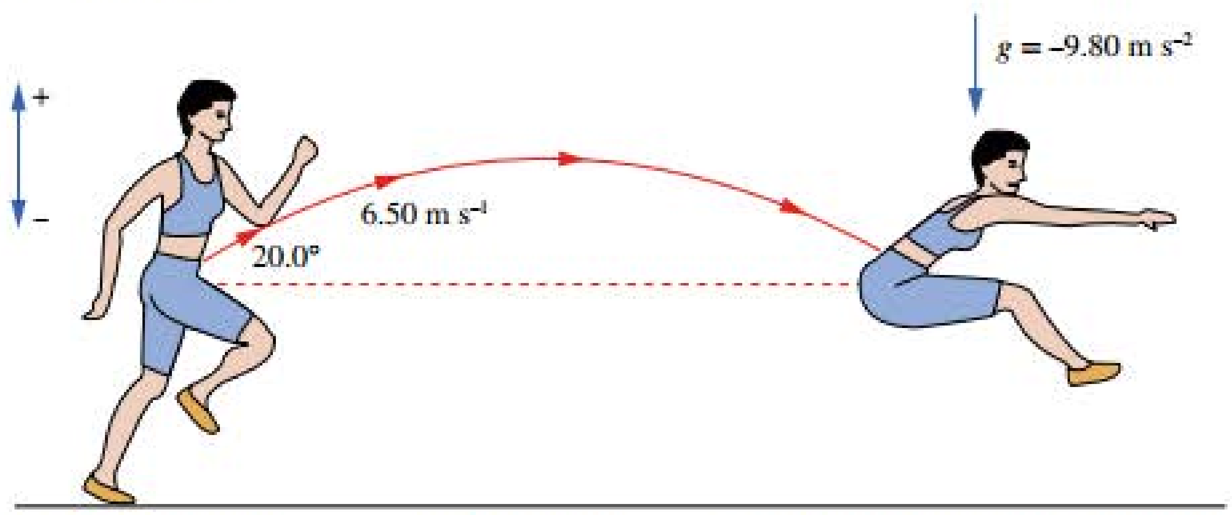
b What is the maximum height gained by the athlete's centre of mass during the jump?	
Thinking	**Working**
To find the maximum height that is gained, you must work with the vertical component. Recall that the vertical component of velocity at the highest point is zero.	Vertically, taking up as positive: $s = ?$ $u = 3.75\,m\,s^{-1}$ $v = 0$ $a = -9.80\,m\,s^{-2}$ $t = ?$

Substitute these values into an appropriate equation for uniform acceleration.	$v^2 = u^2 + 2as$ $0 = 3.75^2 + 2 \times -9.80 \times s$
Rearrange and solve for s.	$s = \frac{3.75^2}{19.6}$ $s = 0.717\text{ m}$
c Assuming a return to the original height, what is the total time the athlete is in the air?	
Thinking	**Working**
As the motion is symmetrical, the time required to complete the motion will be double that taken to reach the maximum height. First, the time it takes to reach the highest point must be found.	Vertically, taking up as positive: $s = ?$ $u = 3.75\text{ m s}^{-1}$ $v = 0$ $a = -9.80\text{ m s}^{-2}$ $t = ?$
Insert these values into an appropriate equation for uniform acceleration.	$v = u + at$ $0 = 3.75 - 9.80t$
Rearrange the formula and solve for t.	$t = \frac{3.75}{9.80}$ $= 0.383\text{ s}$
The time to complete the motion is double the time it takes to reach the maximum height.	Total time $= 2 \times 0.383$ $= 0.766\text{ s}$

Worked example: Try yourself 2.3.1

LAUNCH AT AN ANGLE

A 50 kg athlete in a long-jump event leaps with a velocity of 6.50 m s^{-1} at 20.0° to the horizontal.



For the following questions, treat the athlete as a point mass, ignore air resistance and use $g = 9.80\text{ m s}^{-2}$.

a What is the athlete's velocity at the highest point?

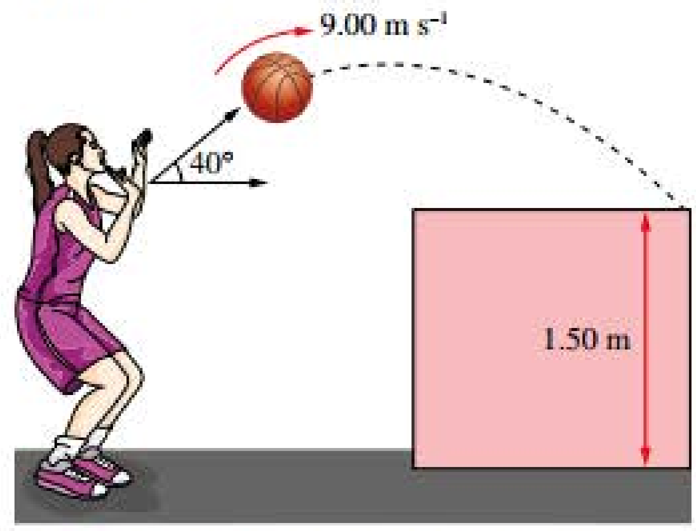
b What is the maximum height gained by the athlete's centre of mass during the jump?

c Assuming a return to the original height, what is the total time the athlete is in the air?

Worked example 2.3.2

LAUNCH AT AN ANGLE

A basketball of mass 2.00 kg is thrown with a velocity of 9.00 m s^{-1} at an angle of 40.0° to the horizontal and lands on a ledge at the same height from which it was thrown, 1.50 m above the ground.



For the following questions, treat the object as a point mass, ignore air resistance and use $g = 9.80$ m s^{-2}.

a What is the ball's velocity at the highest point?	
Thinking	**Working**
First find the horizontal and vertical components of the initial speed.	Using trigonometry: $u_h = 9.00 \cos 40.0°$ $= 6.89$ m s^{-1} $u_v = 9.00 \sin 40.0°$ $= 5.79$ m s^{-1}
Projectiles that are launched obliquely move horizontally with a constant velocity. The vertical component of the velocity at the highest point is zero. The actual velocity is therefore given by the horizontal component of the velocity.	At maximum height, $v = 6.89$ m s^{-1} horizontally to the right.
b What is the maximum height gained by the basketball during the trajectory?	
To find the maximum height that is gained, you must work with the vertical component. At the maximum height, the basketball is moving horizontally and so the vertical component of velocity at this point is zero.	Vertically, taking up as positive: $s = ?$ $u = 5.79$ m s^{-1} $v = 0$ $a = -9.80$ m s^{-2} $t = ?$
Insert these values into an appropriate equation for uniform acceleration.	$v^2 = u^2 + 2as$ $0 = 5.79^2 + 2(-9.80)s$
Rearrange the formula and solve for s.	$s = \frac{0 - 5.79^2}{2 \times (-9.80)}$ $s = 1.71$ m

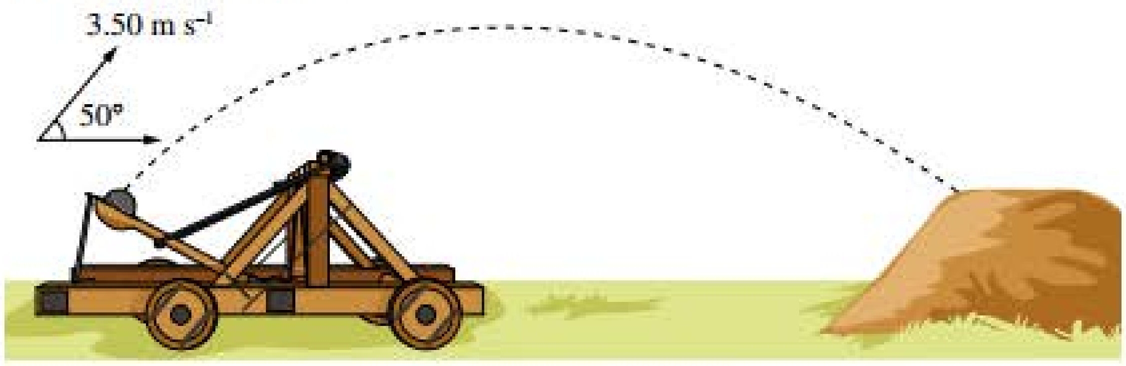
c What is the final velocity of the basketball as it strikes the ledge?	
Thinking	**Working**
Since the ball is landing at the same height above the ground as it was launched, it will have the same gravitational potential energy and hence the same kinetic energy. Its speed will therefore be the same as when it was launched.	$v = 9.00\,\text{m}\,\text{s}^{-1}$
The parabola of the ball's trajectory is symmetric about the middle. Hence, on its downwards trajectory, the ball's velocity vector will be pointed at the same angle as on its journey up at the same height, except this time it will be below the horizontal rather than above it.	$v = 9.00\,\text{m}\,\text{s}^{-1}$ at 40.0° below the horizontal

You will notice that, for any projectile motion problem that neglects air resistance, for a given height above the ground the object will be travelling at the same speed on both its upwards and downwards trajectories. This follows from the conservation of the object's energy. At the same height above the ground, the object will have the same gravitational potential energy, and hence the same kinetic energy. From this it follows that the object will have the same speed.

Worked example: Try yourself 2.3.2

LAUNCH AT AN ANGLE

A little catapult is used to launch a stone of mass 0.250 kg from ground level at a speed of $3.50\,\text{m}\,\text{s}^{-1}$ and an angle of 50.0° above the horizontal. It lands at the same level some distance away.



For the following questions, treat the object as a point mass, ignore air resistance and use $g = 9.80\,\text{m}\,\text{s}^{-2}$.

a What is the stone's velocity at the highest point?

b What is the maximum height gained by the stone during the trajectory?

c What is the final velocity of the stone as it strikes the ground?

PHYSICS IN ACTION

Physics of shot-putting

In shot-put competitions, there is an advantage in being tall. This enables the release height of the shot to be higher, which means the distance travelled by the shot will be greater. At the London Olympic Games in 2012, the men's event was won by Poland's Tomasz Majewski, with a distance of 21.89 m. Tomasz is 201 cm tall. The gold medal for women was won by Valerie Adams of New Zealand, who threw 20.70 m (Figure 2.3.3). Valerie is 193 cm tall.

FIGURE 2.3.3 Valerie Adams, of New Zealand, is a tall woman, which helps her to throw the shot long distances.

When a projectile is launched at an angle to the horizontal, the theoretical launch angle that gives the optimum range is 45°. This applies only where the projectile is launched from zero elevation—that is, when a projectile lands at the same height as it was launched. It is also possible that a projectile lands at a point lower than its launch height. For example, in the shot-put, the projectile is launched from above the ground. The theoretical launch angle for maximum range in this case is around 43°, depending on the actual release height. In reality, however, shot-putters never release at this angle. This is because the speed at which they can launch the shot decreases as the angle gets further from the horizontal. The graph in Figure 2.3.4 shows the relationship between launch speed and launch angle.

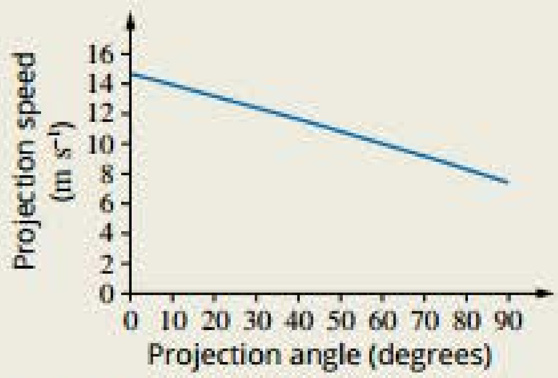


FIGURE 2.3.4 A graph showing that launch speed is greatest with a horizontal launch, and decreases as the launch angle increases.

The decrease in launch speed with increasing projection angle is caused by two factors:

- When throwing with a high projection angle, the shot-putter must expend a greater effort during the delivery phase to overcome the weight of the shot. This reduces the projection speed.
- The structure of the shoulder and arm favours the production of putting force in the horizontal direction more than in the vertical direction.

The optimum projection angle for an athlete is obtained by combining the speed–angle relation for the athlete with the equation for the range of a projectile in free flight. For these reasons, the optimum projection angle for shot-putters is actually about 34° (Figure 2.3.5).

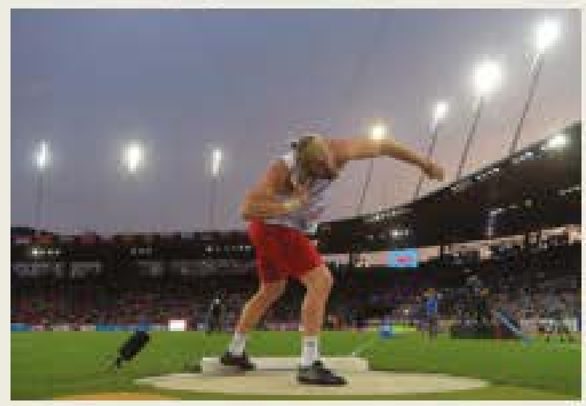


FIGURE 2.3.5 Tomasz Majewski from Poland won the gold medal for the shot-put in London 2012 with a throw of 21.87 m. He would have launched the shot at an angle of about 34°.

2.3 Review

SUMMARY

- Projectiles move in parabolic paths that can be analysed by considering the horizontal and vertical components of the motion.
- If air resistance is ignored, the only force acting on a projectile is its weight; that is, the force due to gravity, F_g. This results in the projectile having an acceleration of $9.80\,m\,s^{-2}$ vertically downwards during its flight.
- The equations for uniform acceleration
  $v = u + at$
  $s = ut + \frac{1}{2}at^2$
  $v^2 = u^2 + 2as$
  must be used for the vertical component.
- The horizontal velocity of a projectile remains constant throughout its flight if air resistance is ignored, and so $v_{av} = \frac{s}{t}$ is used.
- For objects initially launched at an angle to the horizontal, use trigonometry to calculate the initial horizontal and vertical velocities.
- At its highest point, the projectile is moving horizontally. Its velocity at this point is given by the horizontal component of its launch velocity. The vertical component of the velocity is zero at this point.
- Ignoring air resistance, the total energy of the projectile must always be conserved. Therefore, the following equations can be used:
  $E_k = \frac{1}{2}mv^2$
  $E_g = mgh$
  $E_{total} = \frac{1}{2}mv^2 + mgh$

KEY QUESTIONS

For the following questions, assume that the acceleration due to gravity is $9.80\,m\,s^{-2}$ and ignore the effects of air resistance unless otherwise stated.

1 A javelin thrower launches her javelin at 40.0° above the horizontal. Select the correct statement regarding the javelin at the highest point of its path.

A It has zero acceleration.
B It has its slowest speed.
C There are forwards and downwards forces acting on it.
D There are no forces acting on it since it is in free-fall.

2 A child is holding a garden hose at ground level and the water stream from the hose is travelling at $15\,m\,s^{-1}$. Which angle to the horizontal will result in the water stream travelling the greatest horizontal distance through the air?

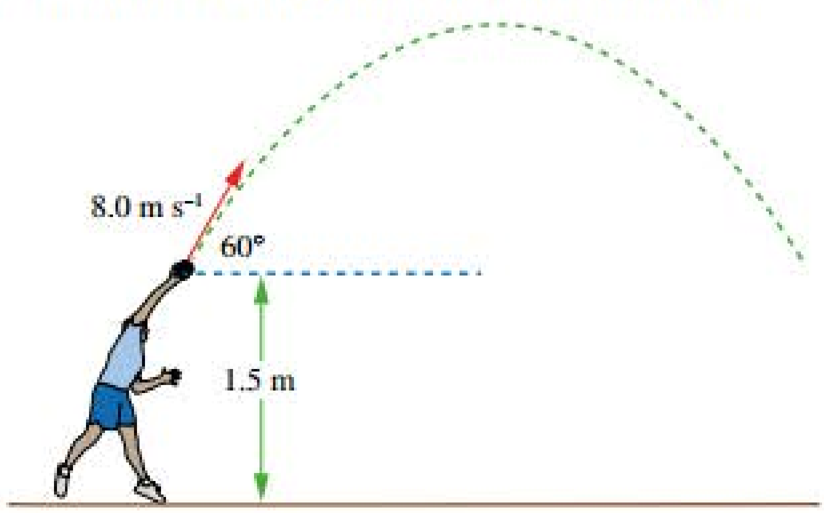
3 At the annual birdman competition, Arun dresses up in his wing-suit and jumps off the edge of a cliff at 30.0° to the horizontal at $5.00\,m\,s^{-1}$ into the water 6.00 m below. Unfortunately, the suit is poorly designed and rather than gliding, he falls with negligible air resistance on a parabolic trajectory. Use conservation of energy methods to calculate:

a Arun's speed when he is 2.70 m below the launch point
b Arun's height above the water when he is travelling at $11.1\,m\,s^{-1}$.

4 A basketballer shoots for a basket by launching the ball at $15\,m\,s^{-1}$ at 25° to the horizontal.

a Calculate the initial horizontal speed of the ball.
b What is the initial vertical speed of the ball?
c What are the magnitude and direction of the acceleration of the ball when it is at its maximum height?
d What is the speed of the ball when it is at its maximum height?

5 In a shot-put event a 2.0 kg shot is launched from a height of 1.5 m, with an initial velocity of $8.0\,m\,s^{-1}$ at an angle of 60° to the horizontal. Answer the questions below about the motion of the shot-put.



a What is the initial horizontal speed of the shot?
b What is the initial vertical speed of the shot?

2.3 Review *continued*

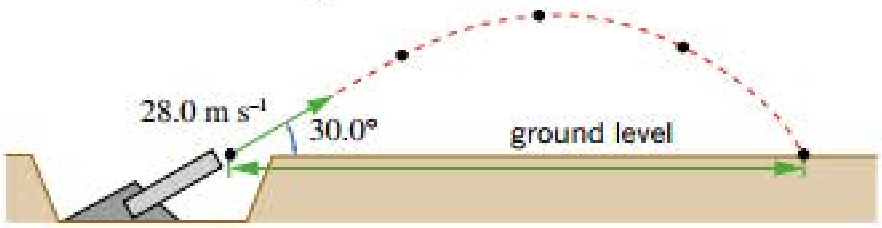
c How long does it take the shot to reach its maximum height?

d What is the maximum height from the ground that is reached by the shot?

e What is the speed of the shot when it reaches its maximum height?

The following information relates to questions 6–10.

A senior physics class conducting a research project on projectile motion constructs a device that can launch a cricket ball. The launching device is designed so that the ball can be launched at ground level with an initial velocity of 28.0 $m s^{-1}$ at an angle of 30.0° to the horizontal.



6 Calculate the horizontal component of the velocity of the ball:

a initially

b after 1.0 s

c after 2.0 s.

7 Calculate the vertical component of the velocity of the ball:

a initially

b after 1.0 s

c after 2.0 s.

8 What is the speed of the cricket ball after 2.00 s?

9 What horizontal distance does the ball travel before landing; that is, what is its range?

10 If the effects of air resistance were taken into account, which one of the following statements would be correct?

A The ball would have travelled a greater horizontal distance before striking the ground.

B The ball would have reached a greater maximum height.

C The ball's horizontal velocity would have been continually decreasing.

D The ball's vertical acceleration would have increased.

11 A girl is standing a few metres away from a wishing well when she decides to lob a coin into it. She throws the coin into the air and it lands perfectly at the bottom of the well, 6.50 m below the ground. The girl threw the coin from 1.50 m above the ground. At its highest point, it was 4.00 m above the ground and its horizontal velocity was 7.50 $m s^{-1}$. Calculate the speed at which it strikes the bottom of the well.

2.4 Circular motion in a horizontal plane

Circular motion is common throughout the universe. On a small scale, this could involve children moving in a circular path on a fair ride (Figure 2.4.1) or passengers in a car as it travels around a roundabout. In athletics, hammer throwers swing the hammer in a circular path before releasing it at high speed. On a much larger scale, the planets orbit the Sun in roughly circular paths; and on an even grander scale, stars can travel in circular paths around the centres of their galaxies. This section explains the nature of circular motion in a horizontal plane, and applies Newton's first and second laws to problems involving circular motion.

FIGURE 2.4.1 The people on this ride are travelling in a circular path at high speed.

UNIFORM CIRCULAR MOTION

Figure 2.4.2 shows an athlete in a hammer throw event, swinging a steel ball in a horizontal circle with a constant speed of $25\,m\,s^{-1}$. As the hammer travels in its circular path, its *speed is constant,* but its *velocity is continually changing.*

Remember that velocity is a vector. Since the direction of the hammer is changing, so too is its velocity, even though its speed is not changing.

The velocity of the hammer at any instant is **tangential** (at a tangent) to its path. At one instant, the hammer is travelling at $25\,m\,s^{-1}$ north, then an instant later at $25\,m\,s^{-1}$ west, then $25\,m\,s^{-1}$ south, and so on.

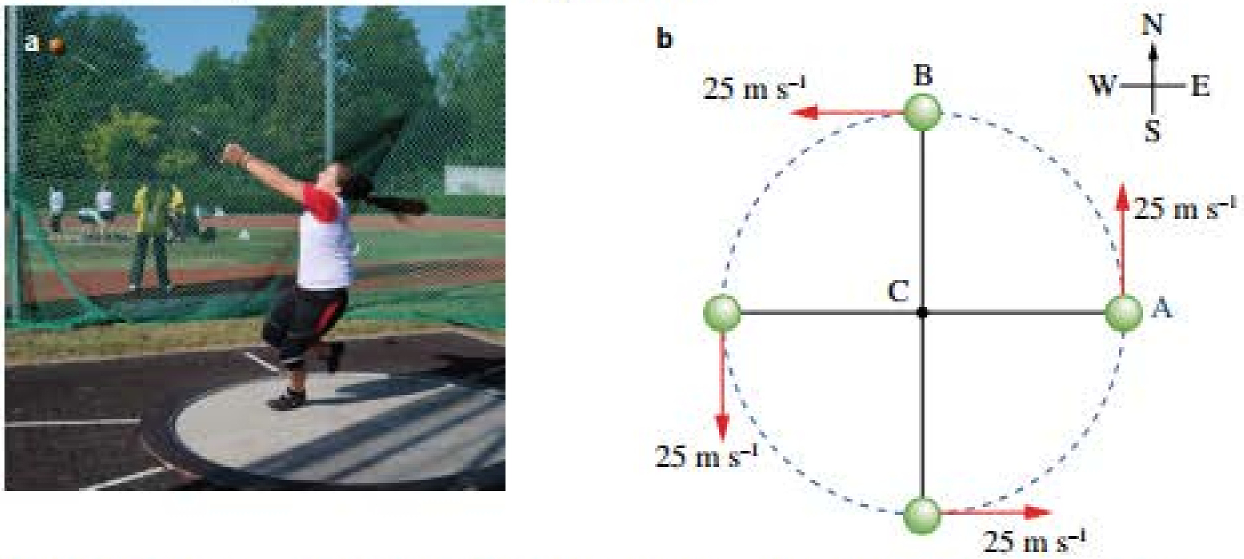


FIGURE 2.4.2 The velocity of the hammer (steel ball) at any instant is tangential to its path and is continually changing even though it has constant speed. This changing velocity means that the hammer is accelerating.

PERIOD AND FREQUENCY

Imagine that an object is moving in a circular path with a constant speed, v, and a radius of r metres, and it takes T seconds to complete one revolution. The time required to travel once around the circle is called the **period**, T, of the motion. The number of rotations each second is the **frequency**, f.

$$f = \frac{1}{T} \text{ and } T = \frac{1}{f}$$

where f is the frequency (Hz)

T is the period (s)

SPEED

An object that travels in a circle will travel a distance equal to the circumference of the circle, $C = 2\pi r$, with each revolution (Figure 2.4.3). Given that the time for each revolution is the period, T, the average speed of the object is:

$$\text{speed} = \frac{\text{distance}}{\text{time}} = \frac{\text{circumference}}{\text{period}}$$

In circular motion, this equation is represented as follows.

The average speed of an object moving in a circular path is:

$$v = \frac{2\pi r}{T}$$

where v is the speed ($m\,s^{-1}$)

r is the radius of the circle (m)

T is the period of motion (s)

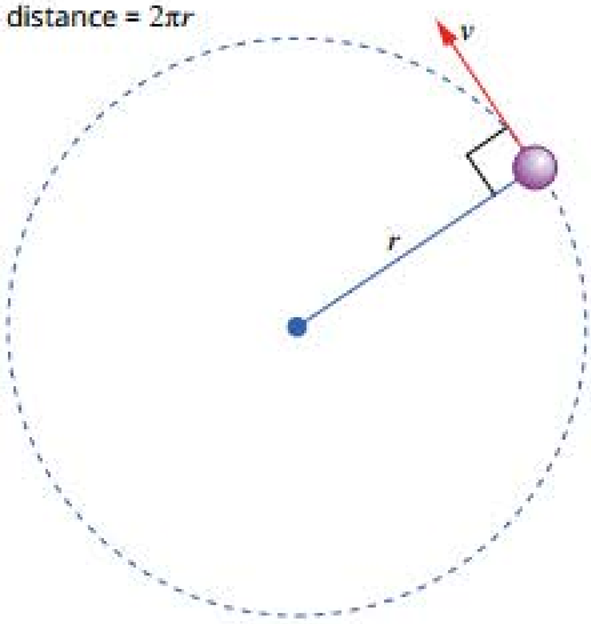


FIGURE 2.4.3 The average speed of an object moving in a circular path is given by the distance travelled in one revolution (the circumference) divided by the time taken (the period, T).

PHYSICSFILE

Wind generators

This wind generator is part of a wind farm at Albany in south-west Western Australia. The towers are 65 m high. Each blade is 35 m long and they rotate at a maximum rate of 22 revolutions per minute. From this information, you should be able to calculate that the tip of each blade is travelling at about 290 $km\,h^{-1}$.

FIGURE 2.4.4 The tips of these wind-generator blades are travelling in circular paths at speeds of around 290 $km\,h^{-1}$.

Worked example 2.4.1

CALCULATING SPEED

A wind turbine has blades 55.0 m in length that rotate at a frequency of 20 revolutions per minute. At what speed do the tips of the blades travel? Express your answer in $km\,h^{-1}$.

Thinking	Working
Calculate the period, T. Remember to express frequency in the correct units. Alternatively, recognise that 20 revolutions in 60 s means that each revolution takes 3 s.	20 revolutions per minute $= \frac{20}{60} = 0.333$ Hz $T = \frac{1}{f}$ $= \frac{1}{0.333} = 3$ s
Substitute r and T into the formula for speed and solve for v.	$v = \frac{2\pi r}{T}$ $= \frac{2 \times \pi \times 55.0}{3}$ $= 115.2\ m\,s^{-1}$
Convert $m\,s^{-1}$ into $km\,h^{-1}$ by multiplying by 3.6.	$115.2 \times 3.6 = 415\ km\,h^{-1}$

Worked example: Try yourself 2.4.1

CALCULATING SPEED

A waterwheel has blades 2.0 m in length that rotate at a frequency of 10 revolutions per minute. At what speed do the tips of the blades travel? Express your answer in $km\,h^{-1}$.

CENTRIPETAL ACCELERATION

When objects travel in circular paths, they can have a constant speed, yet have a velocity that is changing. This seeming contradiction arises because speed is a scalar quantity whereas velocity is a vector.

Since the velocity of the object is changing, it is accelerating even though its speed is not changing. The object is continually deviating inwards from its straight-line direction and so has an acceleration directed towards the centre. This acceleration is known as centripetal acceleration, a. In Figure 2.4.5, the velocity vector of an object travelling in a circular path is shown with an arrow labelled v. Notice that it is at a tangent to the circular path. The acceleration, a, is towards the centre of the circular path.

However, as Figure 2.4.5 shows, even though the object is accelerating towards the centre of the circle, it never gets any closer to the centre. This is the same principle that applies to satellites in orbit, which will be studied in Section 2.7.

The centripetal acceleration, a, of an object moving in a circular path of radius r with velocity v can be found from the relationship:

$$a = \frac{v^2}{r}$$

A substitution can be made for the speed of the object in this equation.

$$v = \frac{2\pi r}{T}$$

So:

$$a = \frac{v^2}{r}$$

$$= \frac{2\pi r}{T}^{2} \times \frac{1}{r}$$

$$= \frac{4\pi^2 r}{T^2}$$

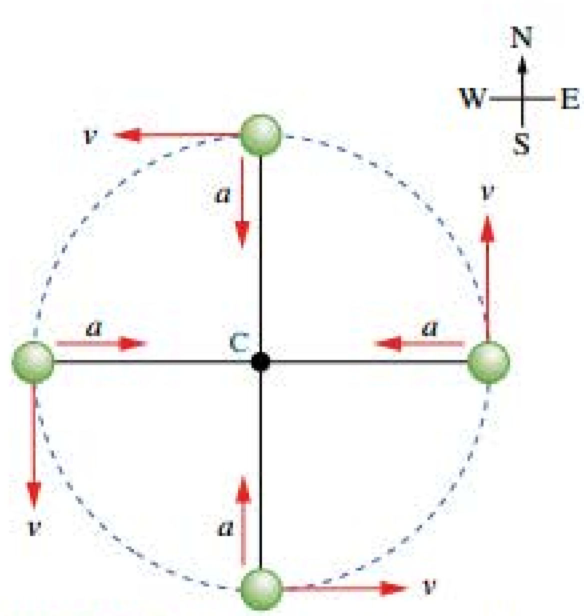


FIGURE 2.4.5 A body moving in a circular path has an acceleration towards the centre of the circle. This is known as a centripetal acceleration.

> Centripetal acceleration is always directed towards the centre of the circular path and is given by:
>
> $$a = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2}$$
>
> where a is the centripetal acceleration ($m\,s^{-2}$)
> v is the speed ($m\,s^{-1}$)
> r is the radius of the circle (m)
> T is the period of motion (s)

FORCES THAT CAUSE CIRCULAR MOTION

As with all forms of motion, an analysis of the forces that are acting is needed to understand why circular motion occurs. In the hammer-throw event described earlier in this section, the hammer ball is continually accelerating. It follows from Newton's second law that there must be a net unbalanced force continuously acting on it. The net unbalanced force that gives the hammer ball its acceleration towards the centre of the circle is known as a **centripetal force**.

In every case of circular motion, a real force is necessary to provide the centripetal force. The force acts in the same direction as the acceleration; that is, towards the centre of the circle. This centripetal force can be provided in a number of ways. For the hammer in Figure 2.4.6(a), the centripetal force is the tension force in the cable. Other examples of centripetal force are also shown in Figure 2.4.6.

FIGURE 2.4.6 (a) In a hammer throw, tension in the cable provides the centripetal force. (b) For planets and satellites, the gravitational attraction to the central body provides the centripetal force. (c) For a car on a curved road, the friction between the tyres and the road provides the centripetal force. (d) For a person in the Gravitron ride, it is the normal force from the wall that provides the centripetal force.

Now, consider the consequences if the unbalanced force ceases to act. In the example of the hammer thrower, if the tension in the wire became zero because the thrower released the ball, there is no longer a force causing the ball to change direction. The result is that the ball then moves in a straight line tangential to its circular path, as would be expected from Newton's first law.

> Centripetal force is given by:
>
> $$F_{net} = ma = \frac{mv^2}{r} = \frac{4\pi^2 rm}{T^2}$$
>
> where F_{net} is the net or centripetal force on the object (N)
> m is the mass (kg)
> a is the acceleration ($m\,s^{-2}$)
> v is the speed ($m\,s^{-1}$)
> r is the radius of the circle (m)
> T is the period of motion (s)

Worked example 2.4.2

CENTRIPETAL FORCES

An athlete in a hammer throw event is swinging a ball of mass 7.00 kg in a horizontal circular path. The ball is moving at 20.0 m s^{-1} in a circle of radius 1.60 m.

a Calculate the magnitude of the acceleration of the ball.	
Thinking	**Working**
As the object is moving in a circular path, the centripetal acceleration is towards the centre of the circle. To find the magnitude of this acceleration, write down the other variables that are given.	$v = 20.0\,\text{m s}^{-1}$ $r = 1.60\,\text{m}$ $a = ?$
Select the equation for centripetal acceleration that fits the information you have, and substitute the values.	$a = \frac{v^2}{r}$ $= \frac{20.0^2}{1.60}$ $= 250\,\text{m s}^{-2}$
Calculate the magnitude only, so no direction is needed in the answer.	The acceleration of the ball is $250\,\text{m s}^{-2}$.

b Calculate the magnitude of the tensile (tension) force acting in the wire.	
Thinking	**Working**
Identify the unbalanced force that is causing the object to move in a circular path. Write down the information that you are given.	$m = 7.00\,\text{kg}$ $a = 250\,\text{m s}^{-2}$ $F_{net} = ?$
Select the equation for centripetal force, and substitute the variables you have.	Equation for centripetal forces: $F_{net} = ma$ $= 7.00 \times 250$ $= 1.75 \times 10^3\,\text{N}$
Calculate the magnitude only, so no direction is needed in the answer.	The force of tension in the wire is the unbalanced force that is causing the ball to accelerate. Tensile force $F_T = 1.75 \times 10^3\,\text{N}$

Worked example: Try yourself 2.4.2

CENTRIPETAL FORCES

An athlete in a hammer throw event is swinging a ball of mass 7.0 kg in a horizontal circular path. The ball is moving at 25 m s^{-1} in a circle of radius 1.2 m.

a Calculate the magnitude of the acceleration of the ball.

b Calculate the magnitude of the tensile force acting in the wire.

PHYSICS IN ACTION

The Gravitron

When a car turns sharply to the left, the passengers in the car seem to sway to the right inside the car. Many mistakenly think that a force to the right is acting. In fact, the passengers are simply maintaining their motion in the original direction of the car, as described by Newton's first law; that is, they are experiencing inertia. If the passengers are (unwisely) not wearing seatbelts, they will be squashed against the right-hand door as the car turns. This will exert a large force to the left on them, which causes them to move to the left.

People moving rapidly in circular paths also mistakenly think that there is an outwards force acting on them. For example, riders on the Gravitron (also known as the Vortex or Rotor), like those in Figure 2.4.7, will 'feel' a force pushing them into the wall. This outwards force is commonly known as a centrifugal (meaning 'centre-fleeing') force. This force does not actually exist in an inertial frame of reference. The riders think that it exists because they are in the rotating frame of reference. From outside the Gravitron, it is evident that there is an inwards force (the normal reaction force) that is holding them in a circular path. If the walls disintegrated and this normal force ceased to act, they would not 'fly outwards' but move at a tangent to their circle.

FIGURE 2.4.7 There is a large inwards force from the wall (a normal reaction force) that causes these children to travel in a circular path.

The Gravitron can rotate at 24 rpm and has a radius of 7 m. The centripetal acceleration can be over 40 m s^{-2}. This is caused by a very large centripetal force from the wall (i.e. the normal force, F_N), which is greater than the weight force, F_g. Since the wall exerts such a large force, the patrons are pinned firmly to the wall as an upwards frictional force, F_f, acts to hold them up. The floor then drops away. It is important to note that there is no outwards force acting. In fact, as you can see in Figure 2.4.8, these forces are unbalanced and the net force is equal in size and direction to the normal force towards the centre of the circle.

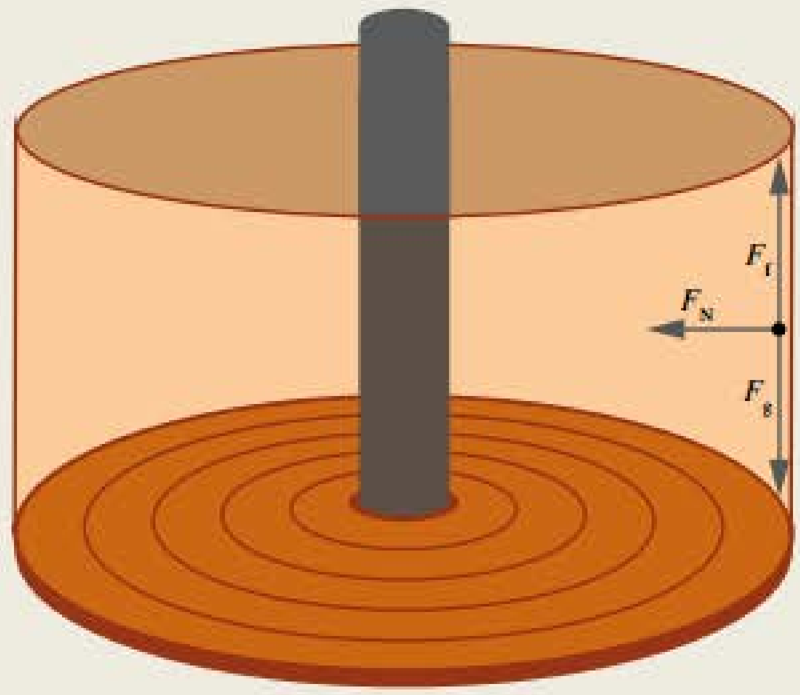


FIGURE 2.4.8 The forces acting on the person are unbalanced. There is an unbalanced force from the wall, F_N, giving the person a centripetal acceleration.

BALL ON A STRING

You may have played totem tennis at one time. This is a game in which a ball is attached to a pole by a string and can travel in a horizontal circle, although the string itself is not horizontal. This kind of motion is shown in Figure 2.4.9.

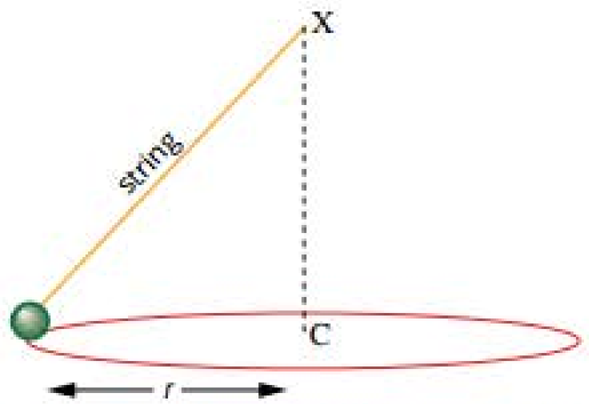


FIGURE 2.4.9 This ball is travelling in a horizontal circular path of radius r. The centre of its circular motion is at C.

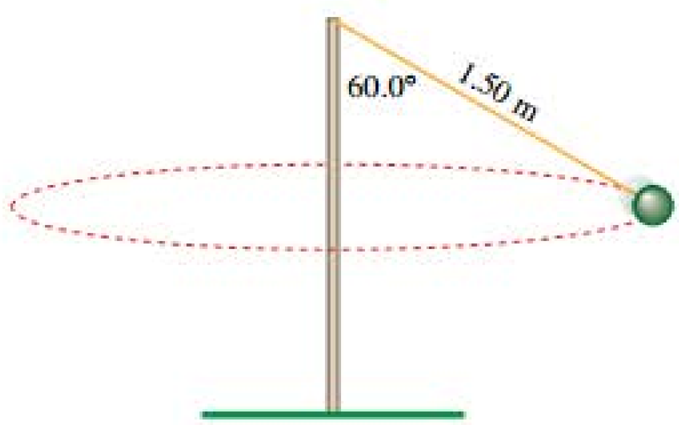
If the ball at the end of the string was swinging slowly, the string would swing down at an angle closer to the pole. If the ball was swung faster, the string would become closer to being horizontal. In fact, it is not possible for the string to be absolutely horizontal, but the greater the speed, the closer to horizontal it becomes. This system is known as a conical pendulum.

If the angle of the conical pendulum is known, trigonometry can be used to find the radius of the circle and the forces involved.

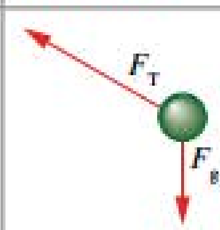
Worked example 2.4.3

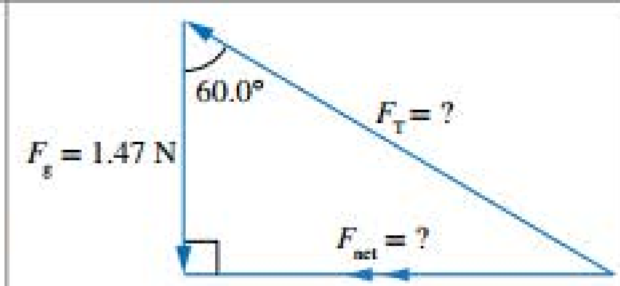
OBJECT ON THE END OF A STRING

During a game of totem tennis, the ball of mass 150 g is swinging freely in a horizontal circular path. The cord is 1.50 m long and is at an angle of 60.0° to the vertical, as shown in the diagram.



a Calculate the radius of the ball's circular path.	
Thinking	**Working**
The centre of the circular path is not the top end of the cord, but is where the pole is level with the ball. Use trigonometry to find the radius.	$r = 1.50 \sin 60.0°$ $= 1.30\,\text{m}$

b Draw and identify the forces that are acting on the ball at the instant shown in the diagram.	
Thinking	**Working**
There are two forces acting—the tension in the cord, F_T, and gravity, F_g. These forces are unbalanced.	

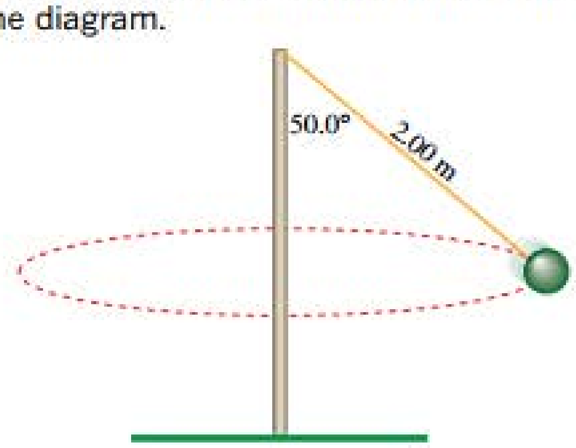
c Determine the net force that is acting on the ball at this time.	
Thinking	**Working**
First calculate the weight force, F_g.	$F_g = mg$ $= 0.150 \times 9.8$ $= 1.47\,N$
The ball has an acceleration that is towards the centre of its circular path. This is horizontal and towards the left at this instant. The net force will also lie in this direction at this instant. A force triangle and trigonometry can be used here.	 $F_{net} = 1.47 \tan 60.0°$ $= 2.55\,N$ towards the left

d Calculate the size of the tensile force in the cord.	
Thinking	**Working**
Use trigonometry to find F_T.	$F_T = \frac{1.47}{\cos 60.0°} = 2.94\,N$

Worked example: Try yourself 2.4.3

OBJECT ON THE END OF A STRING

During a game of totem tennis, the ball of mass 200 g is swinging freely in a horizontal circular path. The cord is 2.00 m long and is at an angle of 50.0° to the vertical, as shown in the diagram.



a Calculate the radius of the ball's circular path.

b Draw and identify the forces that are acting on the ball at the instant shown in the diagram.

c Determine the net force that is acting on the ball at this time.

d Calculate the size of the tensile force in the cord.

2.4 Review

SUMMARY

- Frequency, f, is the number of revolutions each second and is measured in hertz (Hz).
- Period, T, is the time for one revolution and is measured in seconds.
- The relationship between T and f is:
  $$f = \frac{1}{T} \text{ and } T = \frac{1}{f}$$
- An object moving with a uniform speed in a circular path of radius r and with a period T has an average speed that is given by:
  $$v = \frac{2\pi r}{T}$$
- The velocity of an object moving (with a constant speed) in a circular path is continually changing. The velocity vector is always directed at a tangent to the circular path.
- An object moving in a circular path (with a constant speed) has an acceleration due to its circular motion. This acceleration is directed towards the centre of the circular path and is called centripetal acceleration, a:
  $$a = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2}$$
- Centripetal acceleration is a consequence of a centripetal force acting to make an object move in a circular path.
- Centripetal forces are directed towards the centre of the circle and their magnitude can be calculated by using Newton's second law:
  $$F_{net} = \frac{mv^2}{r} = \frac{4\pi^2 rm}{T^2}$$
- Centripetal force is always supplied by a real force, the nature of which depends on the situation. The real force is commonly friction, gravity or the tension in a string or cable.

KEY QUESTIONS

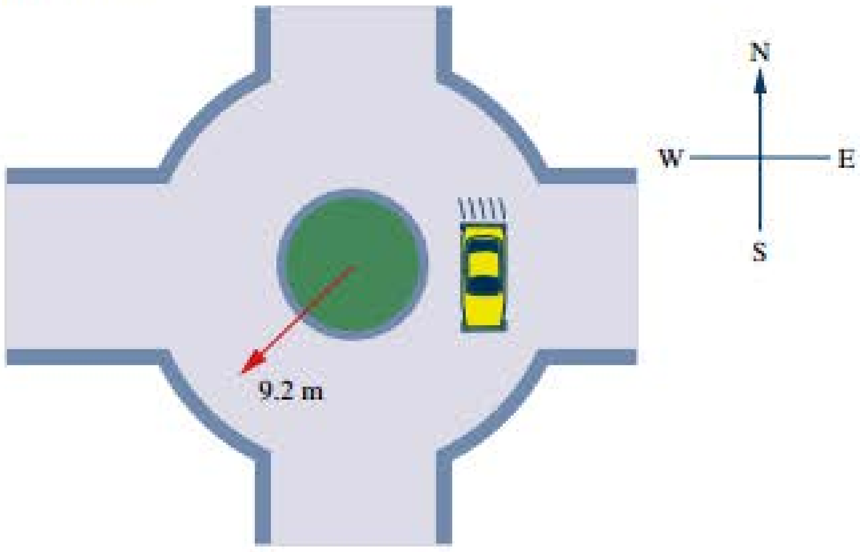
1 A car is travelling with a constant speed around a roundabout. What is the centripetal force that is causing this circular motion?

A gravity
B friction
C drag
D tension

2 A boy is swinging a yo-yo in a horizontal circle 5 times each second. What is the period of the yo-yo?

The following information applies to questions 3–7.

A car of mass 1200 kg is travelling on a roundabout in a circular path of radius 9.2 m. The car moves with a constant speed of 8.0 m s^{-1}. The direction of the car is clockwise around the roundabout when viewed from above as shown.



3 Which two of the following statements correctly describe the motion of the car as it travels around the roundabout?

A It has a constant speed.
B It has a constant velocity.
C It has zero acceleration.
D It has an acceleration that is directed towards the centre of the roundabout.

4 When the car is in the position shown in the diagram, what is the:

a speed of the car?
b velocity of the car?
c magnitude and direction of the acceleration of the car?

5 Calculate the magnitude and direction of the net force acting on the car at the position shown.

6 Some time later, the car has travelled halfway around the roundabout. What is the:

a velocity of the car at this point?
b direction of its acceleration at this point?

7 If the driver of the car kept speeding up, what would eventually happen to the car as it travelled around the roundabout? Explain your answer.

2.4 Review *continued*

8 An ice skater of mass 50 kg is skating in a horizontal circle of radius 1.5 m at a constant speed of 2.0 m s^{-1}.

a Determine the magnitude of the skater's acceleration.

b Are the forces acting on the skater balanced or unbalanced? Explain.

c Calculate the magnitude of the centripetal force acting on the skater.

9 Fiona and Mark are flying their remote-controlled model plane. It has a mass of 1.6 kg and travels in a horizontal circular path of radius 62 m with a speed of 50 km h^{-1}. A radio transmitter controls the plane so there are no strings attached. Answer the questions below about the plane's motion.

a Calculate the period of the model plane's motion.

b Determine the magnitude of the net force that is acting on the plane.

10 An athlete competing at a junior sports meet swings a 2.5 kg hammer in a horizontal circle of radius 0.80 m at 2.0 revolutions per second. Assume that the wire is horizontal at all times.

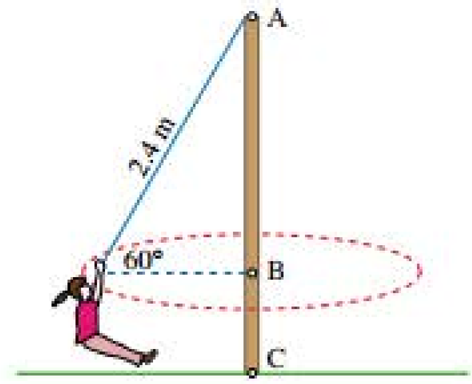
a What is the period of rotation of the hammer?

b What is the orbital speed of the hammer?

c What is the magnitude of the acceleration of the hammer?

d What is the magnitude of the net force acting on the hammer?

11 A child of mass 30 kg is playing on a maypole swing in a playground. The rope is 2.4 m long and at an angle of 60° to the horizontal as she swings freely in a circular path. Ignore the mass of the rope in your calculations.



a Calculate the radius of her circular path.

b Identify the forces that are acting on her as she swings freely.

c What is the direction of her acceleration when she is at the position shown in the diagram?

d Calculate the net force acting on the girl.

e What is her speed as she swings?

2.5 Circular motion on banked tracks

The previous section focused on relatively simple situations involving uniform circular motion in a horizontal plane. However, there are more complex situations involving this type of motion. On many road bends, for example, the road is not horizontal but at a small angle to the horizontal. This enables vehicles to travel at higher speeds without skidding. A more dramatic example of this effect is at a cycling velodrome, like that shown in Figure 2.5.1. The Speed Dome track at Midvale, Western Australia, has a velodrome that has banked or inclined corners that peak at 43°. This enables the cyclists to travel at much higher speeds than if the track was flat. This section examines the principles underlying banked cornering and applies Newton's laws to solving problems involving circular motion on banked tracks.

FIGURE 2.5.1 The Australian women's pursuit track cycling team in action on a banked velodrome track during the London Olympics in 2012.

BANKED CORNERS

Cars and bicycles can travel much faster around corners when the road or track surface is inclined or banked at some angle to the horizontal. **Banked tracks** are most obviously used at cycling velodromes or motor sport events such as NASCAR races. Road engineers also design roads to be banked in places where there are sharp corners, such as exit ramps on freeways.

When cars travel in circular paths on horizontal roads, they are relying on the force of friction between the tyres and the road to provide the sideways force that keeps the car turning in the circular path.

Consider a car travelling clockwise around a horizontal roundabout with a constant speed, v. As can be seen in Figure 2.5.2, the car has an acceleration towards C (the centre of the circle) and so the net force is also sideways on the car towards C.

The forces acting on the car are shown in Figure 2.5.3. As you can see, the vertical forces (gravity and the normal reaction force) are balanced. The only horizontal force is the sideways force that the road exerts on the car tyres. This is a force of friction, F_f, and it is unbalanced, so this is equal to the net force, F_{net}.

If the car drove over an icy patch, there would be no friction and the car would not be able to turn. It would skid in a straight line at a tangent to the circular path.

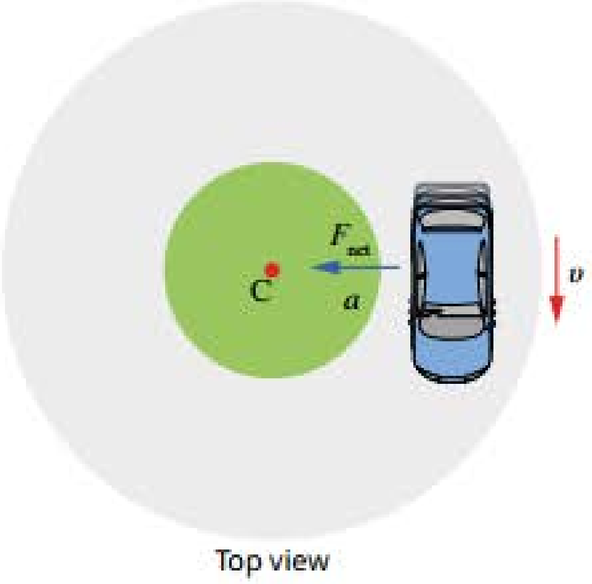


FIGURE 2.5.2 The car is travelling in a circular path on a horizontal track.

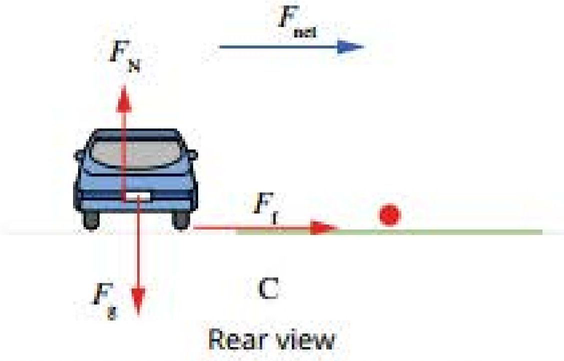


FIGURE 2.5.3 The vertical forces balance, and it is friction between the tyres and the road that enables the car to corner.

Banking the road reduces the need for a sideways frictional force and allows cars to travel faster without skidding off the road and away from the circular path. Consider the same car travelling around a circular, banked road with constant speed, v, as shown in Figure 2.5.4. It is possible for the car to travel at a speed so that there is no sideways frictional force. This is called the **design speed** and it is dependent on the angle, θ, at which the track is banked. At this speed, the car exhibits no tendency to drift higher or lower on the track.

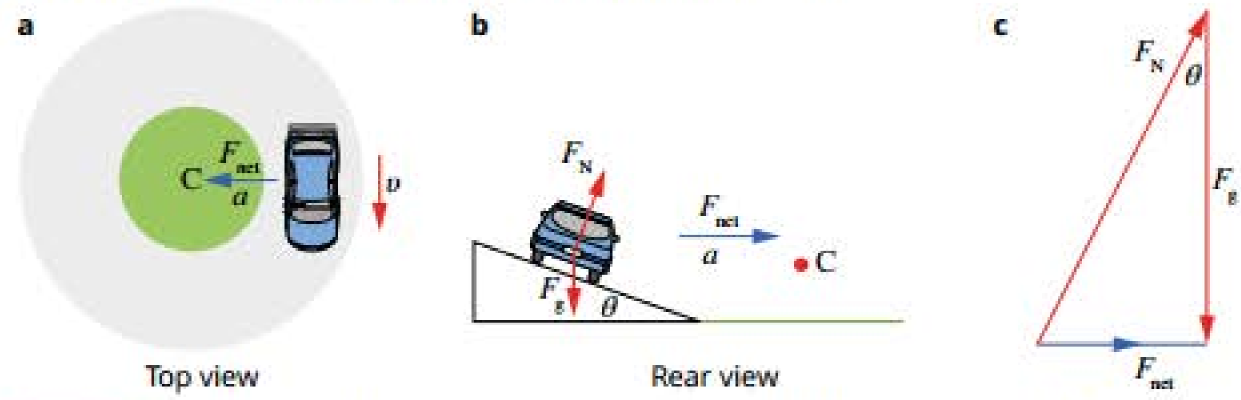


FIGURE 2.5.4 (a) The car is travelling in a circular path on a banked track. (b) The acceleration and net force are towards C. The banked track means that the normal force (F_N) has an inwards component. This is what enables the car to turn the corner. (c) Vector addition gives the net force (F_{net}) as acting horizontally towards the centre.

The car still has an acceleration towards the centre of the circle, C, and so there must be an unbalanced force in this direction. Due to the banking, there are now only two forces acting on the car: its weight, F_g, and the normal force, F_N, from the track.

As can be seen in Figure 2.5.4(b), these forces are unbalanced. They add together to give a net force that is horizontal and directed towards C. If the velocity of the car is above or below the design speed, then the centripetal force will have contributions from both friction, and the horizontal component of the normal force.

At the design speed, the angle of bank, θ, of the road or track can be found by using:

$$\tan\theta = \frac{F_{net}}{F_g}$$

where F_{net} is the force acting to the centre of the circle (N)

F_g is the force due to gravity on the object (N)

Extending this equation by substituting $F_{net} = \frac{mv^2}{r}$ and $F_g = mg$ gives:

$$\tan\theta = \frac{v^2}{rg} \text{ and hence } \theta = \tan^{-1}\frac{v^2}{rg}$$

where m is the mass of the vehicle and passengers (kg)

v is the speed of the vehicle ($m\,s^{-1}$)

r is the radius of the track (m)

θ is the angle of bank (degrees)

g is the acceleration due to gravity ($9.80\,m\,s^{-2}$ near the surface of the Earth)

If the angle and weight are known, trigonometry can be used to calculate the net force (Figure 2.5.4(c)) and therefore the design speed.

Rearranging $\tan\theta = \frac{v^2}{rg}$ to make the design speed, v, the subject gives:

$$v^2 = rg\tan\theta$$

$$v = \sqrt{rg\tan\theta}$$

It is worth noting that the normal force will be larger here than on a flat track. In the case of a cyclist, the rider and bicycle would feel a larger force acting from the road when cycling on a banked track than when cycling on a flat track (Figure 2.5.5).

Worked example 2.5.1

BANKED CORNERS

A curved section of track on an Olympic velodrome has a radius of 50 m and is banked at an angle of 42° to the horizontal. A cyclist of mass 75 kg is riding on this section of track at the design speed.

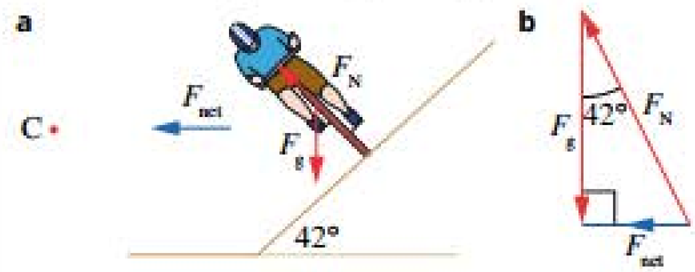
a Calculate the net force acting on a cyclist at this instant if they are riding at the design speed.	
Thinking	**Working**
Draw a force diagram and include all forces acting on the cyclist.	The forces acting on the cyclist are gravity and the normal force from the track, and these are unbalanced. The net force is horizontal and towards the centre of the circular track as shown in diagram (a) and the force triangle of diagram (b). 
Calculate the weight force, F_g.	$F_g = mg$ $= 75 \times 9.80$ $= 735\,N$
Use the force triangle and trigonometry to work out the net force, F_{net}.	$\tan\theta = \frac{F_{net}}{F_g}$ $\tan 42° = \frac{F_{net}}{735}$ $F_{net} = 0.90 \times 735$ $= 662\,N$
As force is a vector, a direction is needed in the answer.	Net force is 662 N horizontally towards the centre of the circle.
b Calculate the design speed for this section of the track.	
Thinking	**Working**
List the known values.	$m = 75\,kg$ $r = 50\,m$ $\theta = 42°$ $F_g = 735\,N$ $F_{net} = 662\,N$ $v = ?$
Use the design speed formula.	$v = \sqrt{rg\tan\theta}$ $v = \sqrt{50 \times 9.80 \times \tan 42°}$ $= 21\,m\,s^{-1}$

FIGURE 2.5.5 Australian cyclist Anna Meares on this banked velodrome track is cornering at speeds far higher than she could use on a flat track. Cyclists on a velodrome do not need to rely on friction to turn, and so experience a larger normal force than usual.

PHYSICSFILE

Wall of Death

In some amusement parks in other parts of the world, there is a ride known menacingly as the Wall of Death (Figure 2.5.6). It consists of a cylindrical enclosure with vertical walls. People on bicycles and motorbikes ride into the enclosure and around the vertical walls, so the angle of banking is 90°! The riders need to keep moving and are depending on friction to hold them up. By travelling fast, the centripetal force (the normal force from the wall) is large and this increases the size of the grip (friction) between the wall and tyres. If the rider slammed on the brakes and stopped, they would simply plummet to the ground.

FIGURE 2.5.6 For a rider to successfully conquer the Wall of Death, they need to travel reasonably fast and there must be good grip between the tyres and the track. The rider is relying on friction to maintain their motion along the wall.

Worked example: Try yourself 2.5.1

BANKED CORNERS

A curved section of track on an Olympic velodrome has radius of 40 m and is banked at an angle of 37° to the horizontal. A cyclist of mass 80 kg is riding on this section of track at the design speed.

a Calculate the net force acting on the cyclist at this instant as they are riding at the design speed.

b Calculate the design speed for this section of the track.

EXTENSION

Leaning into corners

In many sporting events, participants need to travel around corners at high speeds. As shown in Figure 2.5.7, motorbike riders lean their bikes over almost onto the track as they corner. This leaning technique is also evident in ice skating, bicycle races, skiing and even when you run round a corner. It enables the competitor to corner at high speed without falling over.

FIGURE 2.5.7 Australia's Casey Stoner won the 2012 Moto GP championship. Here he is leaning his bike as he takes a corner at Phillip Island. Leaning into the corner enables him to corner at higher speeds. In fact, the bike would go out of control if he did not lean it.

Consider a cyclist cornering on a horizontal road surface (Figure 2.5.8). The forces acting on the bicycle and rider are unbalanced. The forces are the weight force, F_g, and the force from the track. The track exerts a reaction force on the rider that acts both inwards and upwards. The inwards component is the frictional force, F_f, between the track and the tyres. The upwards component is the normal force, F_N, from the track.

The rider is travelling in a horizontal circular path at constant speed, and so has a centripetal acceleration directed towards the centre of the circle. Therefore, the net force is directed towards the centre of the circle.

By analysing the vertical and horizontal components in Figure 2.5.8, you see that the weight force, F_g, must balance the normal force, F_N. The net force that is producing the centripetal acceleration is supplied by the frictional force, F_f. In other words, the rider is depending on a sideways frictional force to turn the corner. An icy or oily patch on the track would cause the tyres to slide out from under the rider, and he or she would slide painfully along the road at a tangent to the circular path.

FIGURE 2.5.8 The forces acting as the rider turns a corner are the weight, F_g, the normal force, F_N, and the friction, F_f, between the tyres and the road. The friction supplies the unbalanced force that leads to the corner-turning motion.

2.5 Review

SUMMARY

- A banked track is one where the track is inclined at some angle to the horizontal. This enables vehicles to travel at higher speeds when cornering, compared with around a horizontal curved path.
- Banking a track eliminates the need for a sideways frictional force to turn. When the speed and angle are such that there is no sideways frictional force, the speed is known as the design speed.
- The forces acting on a vehicle travelling at the design speed on a banked track are gravity and the normal force from the track. These forces are unbalanced and add to give a net force directed towards the centre of the circular motion.
- At the design speed, the angle of bank of the track $\theta = \tan^{-1} \frac{v^2}{rg}$.
- For a given bank angle and curve radius, the design speed is given by $v = \sqrt{rg \tan\theta}$.

KEY QUESTIONS

1 A cyclist is riding along a circular section of a velodrome where the radius is 30 m and the track is inclined at 30° to the horizontal. The cyclist is riding at the design speed and maintains a constant speed. Describe the direction of the acceleration on the cyclist.

2 An architect is designing a velodrome and the original plans have semicircular sections of radius 15 m and a banking angle of 30°. The architect is asked to make changes to the plans that will increase the design speed for the velodrome. What two design elements could the architect change in order to meet this requirement?

3 A racing car is travelling around a circular banked track that has a design speed of 100 km h^{-1}. On one lap, the car travels at 150 km h^{-1}. At this higher speed, the car would tend to travel in a different position along the banked surface. Would the car travel higher or lower up the banked track? Explain your answer.

4 A racing car travels at high speed along a horizontal track and tries to turn a corner. The car skids and loses control. The racing car then travels along a banked track and is able to travel much faster around the corners without skidding at all. Complete the sentences below by choosing the correct term in bold.
On the horizontal track, the car is depending on the force of **friction/weight** to turn the corner. The **friction/normal** force is equal to the **weight/friction** of the car, so these vertical forces are **balanced/ unbalanced**. On the banked track, the **normal/weight** force is not vertical and so is not balanced by the **weight/normal** force. In both cases, the forces acting on the car are unbalanced.

5 Copy and complete the following diagram by drawing and labelling the normal force, the weight force and the net force acting on the bicycle.

The following information relates to questions 6 and 7.
A cycling velodrome has a turn that is banked at 33° to the horizontal. The radius of the track at this point is 28 m.

6 Determine the speed (in km h^{-1}) at which a cyclist of mass 55 kg would experience no sideways force on their bike as they ride this section of track.

7 **a** Calculate the size of the normal force that is acting on the cyclist.
b How would this compare with the normal force if they were riding on a flat track?

8 A race track is banked so that cars cornering at 40 m s^{-1} experience no sideways frictional forces. The track is circular with a radius of 150 m. Calculate the angle to the horizontal at which the track is banked.

The following information applies to questions 9 and 10.
A section of track at a NASCAR raceway is banked to the horizontal. The track section is circular with a radius of 80 m and it has a design speed of 18 m s^{-1}. A car of mass 1200 kg is being driven around the track at 18 m s^{-1}.

9 **a** Calculate the magnitude of the net force acting on the car (in kN).
b Calculate the angle to the horizontal at which the track is banked.

10 The driver now drives around the track at 30 m s^{-1}. What would the driver have to do to maintain a circular path around the track?

2.6 Circular motion in a vertical plane

In previous sections, the motion of objects travelling in circular paths was discussed. It was explained that a body moving with constant speed in a horizontal circular path has an acceleration that is directed towards the centre of the circle. The same applies for vertical circular paths.

If you have ever been on a rollercoaster ride, you will have travelled over humps and down through dips at high speeds and, at times, in circular arcs (Figure 2.6.1). There are also rides that travel through full 360° vertical circles. During these rides, your body will experience forces that you may or may not find pleasant.

When you travel on a rollercoaster, you can experience quite strong forces pushing you down into the seat as you fly through the dips. Also, as you travel over the humps, you tend to lift off the seat. These forces relating to circular motion in a vertical plane will be discussed in this section. As in the previous sections, Newton's laws are used to solve problems involving this type of circular motion.

FIGURE 2.6.1 This rollercoaster has a circular path in a vertical plane at this point.

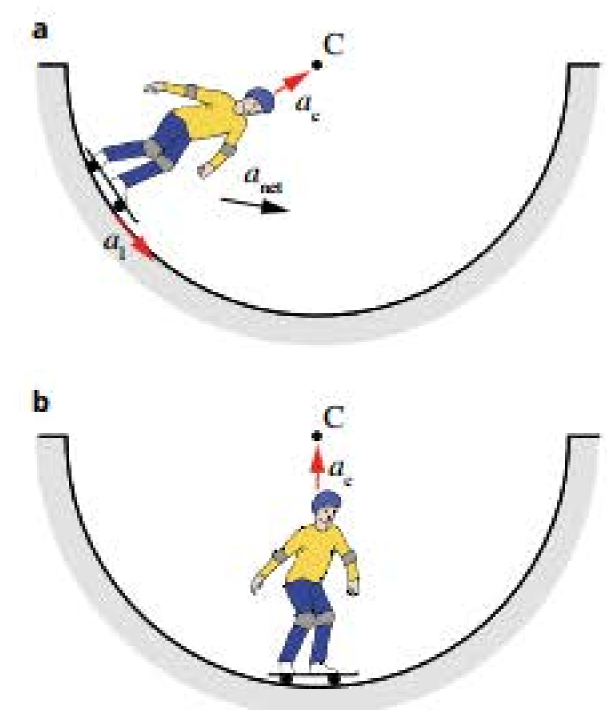


FIGURE 2.6.2 (a) High on the sides of the 'half-pipe' the skateboarder speeds up, and so has both a linear and a centripetal acceleration. The net acceleration, a_{net}, is not towards C. (b) At the lowest point the speed of the skateboarder is momentarily constant, so there is no linear acceleration. The acceleration is supplied completely by the centripetal acceleration, a_c, which is acting towards C.

MOVING IN VERTICAL CIRCLES

A body moving with constant speed in a horizontal circular path has an acceleration that is directed towards the centre of the circle. The same applies for vertical circular paths. However, circular motion in a vertical plane in real life is often more complex, as it does not usually involve constant speeds.

An example of this is illustrated in Figure 2.6.2(a). The speed of the skateboarder performing in a half-pipe will increase on the way down as gravitational potential energy is converted into kinetic energy. This means the skater will experience a linear acceleration, a_1, as well as a centripetal acceleration, a_c. The resultant acceleration is not directed towards the centre of the circular path.

At the bottom of the half-pipe, the skateboarder will be neither slowing down nor speeding up, so the acceleration is purely centripetal at this point, as shown in Figure 2.6.2(b). The same applies at the very top of a circular path. For this reason, motion at these points is easier to analyse.

Uniform horizontal motion

Theme park rides make you appreciate that the forces you experience throughout a ride can vary greatly. First, consider the case of a person in a rollercoaster cart, like that shown in Figure 2.6.3, travelling horizontally at $4.0\,m\,s^{-1}$. If the person's mass is 50 kg and the gravitational field strength is $9.80\,m\,s^{-2}$, the forces acting on the person can be calculated. These forces are the weight, F_g, and the normal reaction force, F_N, from the seat.

The person is moving in a straight line with a constant speed, so there is no unbalanced force acting. The weight force balances the normal reaction force from the seat. The normal force is therefore 490 N up, which is what usually acts upwards on this person when moving horizontally, and so they would feel their usual weight.

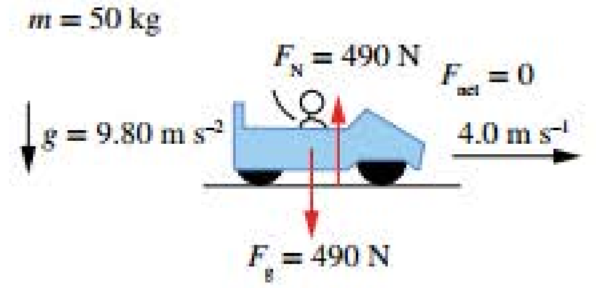


FIGURE 2.6.3 The vertical forces are in balance in this situation, i.e. $F_N = F_g$.

Circular motion: travelling through dips

Now consider the forces that act on the person after the cart has reached the bottom of a circular dip of radius 2.5 m and is moving at $8.0\,m\,s^{-1}$. Figure 2.6.4 illustrates these forces.

The person will have a centripetal acceleration due to the circular path. This centripetal acceleration is directed towards the centre, C, of the circular path—in this case, vertically upwards. The person's centripetal acceleration, a, is:

$$a = \frac{v^2}{r} = \frac{8.0^2}{2.5}$$
$= 26\,m\,s^{-2}$ towards C, or upwards

The net (centripetal) force acting on the person is given by:

$$F_{net} = ma = 50 \times 26$$
$= 1300$ N upwards

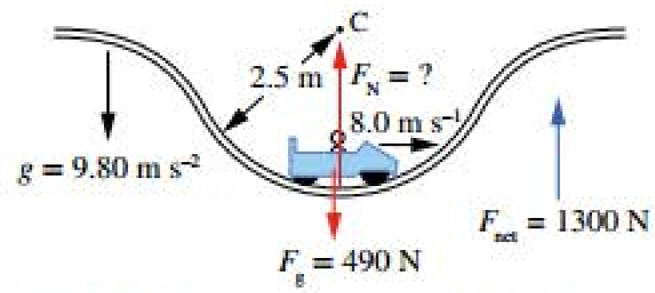


FIGURE 2.6.4 The person has a centripetal acceleration that is directed upwards towards the centre of the circle, and so the net force is also upwards. In this case, the magnitude of the normal force, F_N, is greater than the weight, F_g, and produces a situation where the rider feels heavier than usual.

The normal force, F_N, and the weight force, F_g, are not in balance anymore. They add together to give an upwards force of 1300 N. This indicates that the normal force must be greater than the weight force by 1300 N. In other words, the normal force is 490 N + 1300 N = 1790 N up. This is over three times larger than the normal force of 490 N that usually acts. That is the reason why you feel the seat pushing up against you much more strongly at this point in a ride. The normal force of 1790 N in this instance is equal to the apparent weight of the person, and indicates they would feel much heavier than usual.

Circular motion: travelling over humps

Now consider the situation as the cart moves over the top of a hump of radius 2.5 m with a lower speed of $2.0\,m\,s^{-1}$, as illustrated in Figure 2.6.5.

The person now has a centripetal acceleration that is directed vertically downwards towards the centre, C, of the circle. Therefore, the net force acting at this point is directed vertically downwards. The centripetal acceleration is:

$$a = \frac{v^2}{r} = \frac{2.0^2}{2.5}$$
$= 1.6\,m\,s^{-2}$ towards C, or downwards

The net (centripetal) force is:

$$F_{net} = ma = 50 \times 1.6$$
$= 80$ N downwards

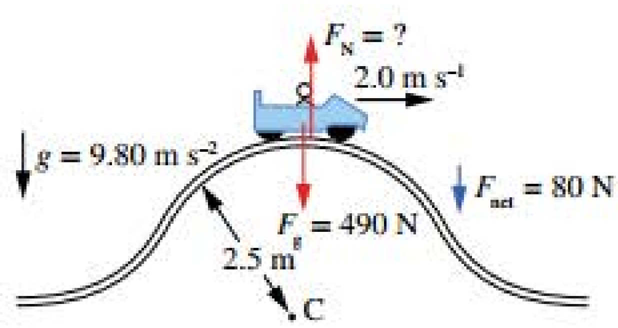


FIGURE 2.6.5 The centripetal acceleration is downwards towards the centre of the circle, and so the net force is also in that direction. At this point, the magnitude of the normal force, F_N, is less than the weight, F_g, of the person.

As in the dip, the weight force and the normal force are not in balance. They add to give a net force of 80 N down. The weight force, F_g, must therefore be 80 N greater than the normal force, F_N. This tells us that the normal force is:

490 N + (−80) N = 410 N up

How the normal force varies during the ride

It is interesting to compare the normal forces that act on the person in these three situations.

- The normal force when travelling horizontally is 490 N upwards.
- At the bottom of the dip, the normal force is 1790 N upwards. In other words, in the dip, the seat pushes into the person with a greater force than usual. This gives the person an apparent weight of 1790 N and makes the person feel much heavier than normal. If the person was sitting on weighing scales at this time, it would show a higher than usual reading.
- At the top of the hump, the normal force is 410 N upwards. In other words, over the hump, the seat pushes into the person with a smaller force than usual. This gives the person an apparent weight of 410 N and gives them the sensation of feeling lighter.

The weight of the person has not changed. F_g is 490 N throughout the duration of the ride; it is the normal force acting on the person that varies. The normal force is equal to the person's apparent weight, and this makes the person feel heavier and lighter as they travel through the dips and humps respectively.

A rollercoaster cart going through dips and over humps is always moving above the rollercoaster track, so the normal reaction force acts upwards. It is also possible for the cart to travel on the underside of the track when it goes upside down through a loop. Worked example 2.6.1 shows a toy car travelling through a vertical loop, on the inside of the loop.

PHYSICSFILE

Fighter pilots

A fighter pilot in a vertical loop manoeuvre can safely experience centripetal accelerations of up to around 5*g*, or 49 m s^{-2}. In a loop where the g-forces are greater than this, the pilot may pass out. If the pilot flies with his or her head inside the loop, the centripetal acceleration of the plane will push the pilot into their seat and make the blood flow away from their head. The resulting lack of blood in the brain may cause the pilot to 'grey out' and they may lose total consciousness ('black out'). This type of force is a called a positive-g force. Fighter pilots wear 'g-suits' that pressurise the legs to prevent blood flowing into them, which helps them to maintain consciousness (Figure 2.6.6).

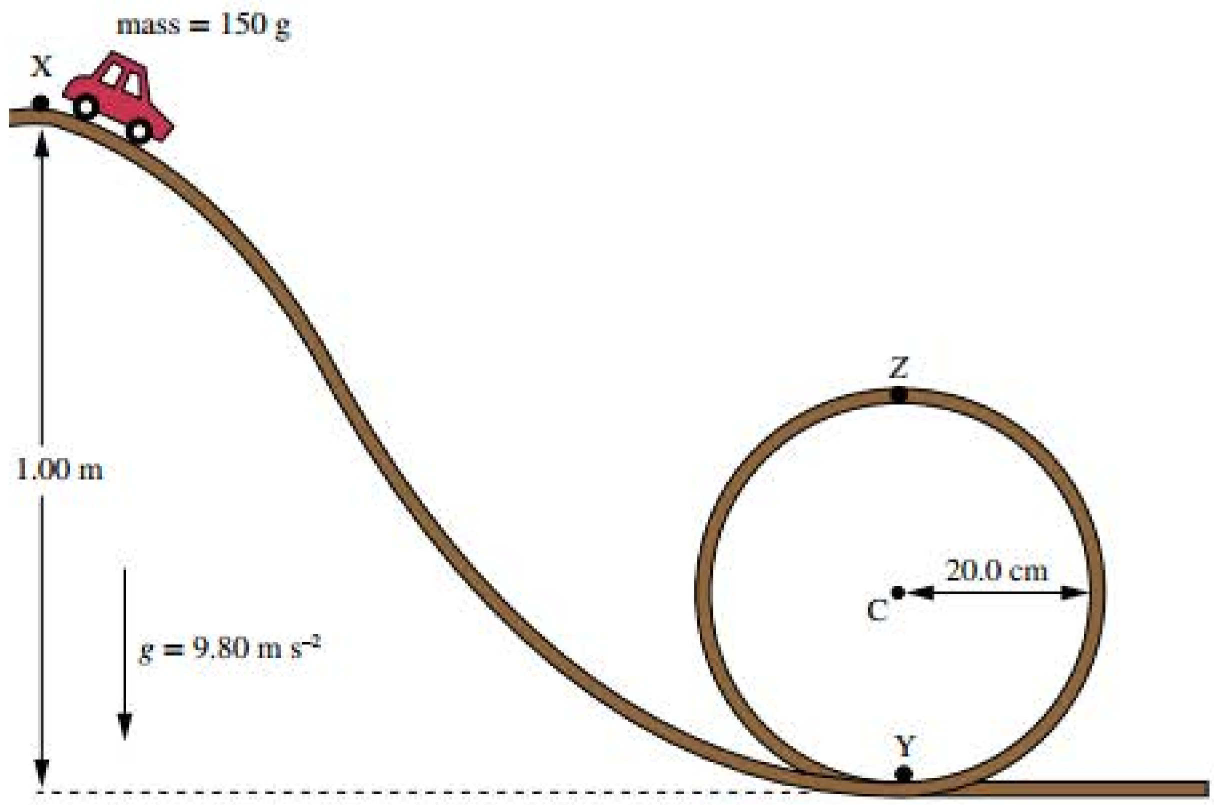
However, if the pilot's head is on the outside of the loop, the centripetal acceleration will pull the pilot onto their harness and the additional blood flow to the head can make the whites of the eyes turn red. The excess blood flow in the head may cause 'red out'. This type of force is called a negative-g force.

FIGURE 2.6.6 Fighter pilots wear pressurised suits to allow their bodies to withstand the large forces that act during tight turns.

Worked example 2.6.1

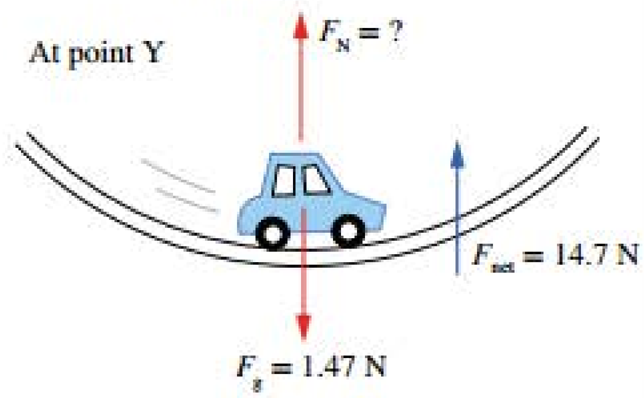
VERTICAL CIRCULAR MOTION

A student arranges a toy car track with a vertical loop of radius 20.0 cm, as shown. A toy car of mass 150 g is released from rest at point X at a height of 1.00 m. The car rolls down the track and travels inside the loop. Assume g is $9.80\,\text{m s}^{-2}$, and ignore friction.



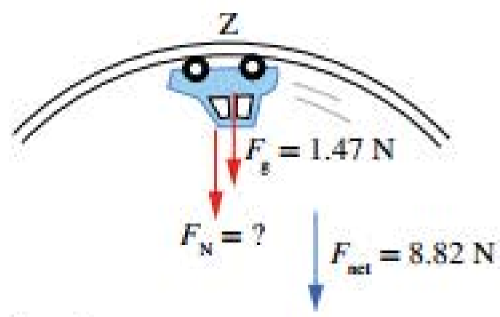
a Calculate the speed of the car as it reaches the bottom of the loop, point Y.	
Thinking	**Working**
Note all the variables given to you in the question.	At X: $m = 150\,\text{g} = 0.150\,\text{kg}$ $h = 1.00\,\text{m}$ $v = 0$ $g = 9.80\,\text{m s}^{-2}$
Use an energy approach to calculate the speed. Calculate the total mechanical energy first.	The initial speed is zero, so E_k at X is zero. The total mechanical energy, E_m, at X is: $E_m = E_k + E_g$ $= \frac{1}{2}mv^2 + mg\ h$ $= 0 + 0.150 \times 9.80 \times 1.00$ $= 1.47\,\text{J}$
Use conservation of energy ($E_m = E_k + E_g$) to determine the velocity at point Y. As the car rolls down the track, it loses its gravitational potential energy and gains kinetic energy. At the bottom of the loop (Y), the car has zero potential energy.	At Y: $E_m = 1.47\,\text{J}$ $h = 0$ $E_g = 0$ $E_m = E_k + E_g$ $= \frac{1}{2}mv^2 + mg\ h$ $1.47 = 0.5 \times 0.150v^2 + 0$ $v^2 = \frac{1.47}{0.075}$ $v = \sqrt{19.6}$ $= 4.43\,\text{m s}^{-1}$

b Calculate the normal reaction force from the track at point Y.

Thinking	Working
To solve for F_N, start by working out the net, or centripetal, force. At Y, the car has a centripetal acceleration towards C (i.e. upwards), so the net (centripetal) force must also be vertically upwards at this point.	$F_{net} = \frac{mv^2}{r}$ $= \frac{0.150 \times 4.43^2}{0.200}$ $= 14.7$ N up
Calculate the weight force, F_g, and add it to a force diagram.	$F_g = mg$ $= 0.150 \times 9.80$ $= 1.47$ N down 
Work out the normal force using vectors. Note: use up as positive and down as negative for your calculations. The forces acting are unbalanced, as the car has a centripetal acceleration upwards (towards C). The upwards (normal) force must be larger than the downwards force.	$F_{net} = F_g + F_N + 14.7 = -1.47 + F_N$ $F_N = +14.7 + 1.47$ $= 16.2$ N up Note that the force the track exerts on the car is much greater (by about ten times) than the weight force. If the car were travelling horizontally, the normal force would be just 1.47 N up.

c What is the speed of the car as it reaches point Z?

Thinking	Working
Calculate the velocity from the values you have, using $E_m = E_k + E_g$.	At Z: $m = 0.150$ kg $r = 20.0$ cm $= 0.200$ m $\Delta h = 2 \times 0.200 = 0.400$ m Mechanical energy is conserved, so $E_m = 1.47$ J At Z: $E_m = E_k + E_g$ $= \frac{1}{2}mv^2 + mg\ h$ $1.47 = \frac{1}{2}(0.150 \times v^2) + 0.150 \times 9.80 \times 0.400$ $1.47 = 0.075v^2 + 0.588$ $0.075v^2 = 1.47 - 0.588$ $v^2 = 11.76$ $v = 3.43\ \text{m s}^{-1}$

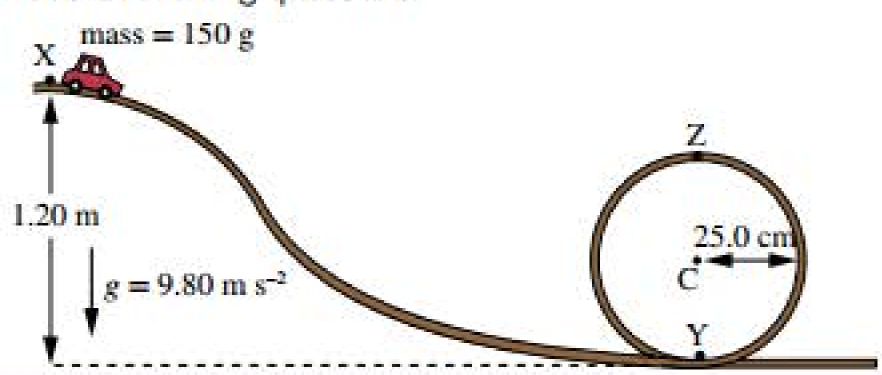
d What is the normal force acting on the car at point Z?	
Thinking	**Working**
To find F_N, start by working out the net, or centripetal, force. At Z, the car has a centripetal acceleration towards C (i.e. downwards), so the net (centripetal) force must also be vertically downwards at this point.	$F_{net} = \frac{mv^2}{r}$ $= \frac{0.150 \times 3.43^2}{0.200}$ $= 8.82$ N down
Work out the normal force using vectors. Note: use up as positive and down as negative for your calculations.	 $F_{net} = F_g + F_N$ $-8.82 = -1.47 + F_N$ $F_N = -8.82 + 1.47$ $= -7.35$ $= 7.35$ N down Note that there is still strong contact between the car and the track, as given by the normal force, but it is only about half the size of the normal force at the bottom of the track. If the car had progressively lower speeds, the normal force at Z would decrease and eventually drop to zero. At this point, the car would lose contact with the track, fall off the track and its acceleration would be equal to g.

Worked example: Try yourself 2.6.1

VERTICAL CIRCULAR MOTION

A student arranges a toy car track with a vertical loop of radius 25.0 cm, as shown.

A toy car of mass 150 g is released from rest at point X at a height of 1.20 m. The car rolls down the track and travels around the loop. Assume $g = 9.80\ \text{m s}^{-2}$, and ignore friction for the following questions.



a Calculate the speed of the car as it reaches the bottom of the loop, point Y.

b Calculate the normal reaction force from the track at point Y.

c What is the speed of the car as it reaches point Z?

d What is the normal force acting on the car at point Z?

PHYSICS IN ACTION

Travelling upside down without falling out

You might have been on a rollercoaster like the one in Figure 2.6.7, where you were actually upside down at times during the ride. These rides use their speed and the radius of their circular path to prevent the riders from falling out. In theory, the safety harnesses worn by the riders are not needed to hold the people in their seats.

FIGURE 2.6.7 The thrill seekers on this rollercoaster ride don't fall out when upside down because the centripetal acceleration of the cart is greater than $9.80\,m\,s^{-2}$ down.

The reason people don't fall out is that their centripetal acceleration while on the rollercoaster is greater than the acceleration due to gravity ($9.80\,m\,s^{-2}$). To understand the significance of this, try the following activity. Stand up, reach up with one hand, place an eraser on the palm of that hand, then turn your hand palm down and move it rapidly towards the floor.

You should find, after one or two attempts, that it is possible to keep the eraser in contact with your hand as you 'push' it down. The eraser is upside down, but it is not falling out of your hand. Your hand must, for a short time, be moving down with a downwards acceleration greater than $9.80\,m\,s^{-2}$ and continually exerting a normal force on the eraser. This acceleration of $9.80\,m\,s^{-2}$ down is the critical point in this exercise. If your hand had an acceleration less than this, the eraser would fall away from your hand to the floor. Try it to confirm that is what happens.

A similar principle holds with rollercoaster rides. The people on the ride don't fall out at the top because the motion of the rollercoaster gives them a centripetal acceleration that is greater than $9.80\,m\,s^{-2}$ down. The engineers who designed the ride would have ensured that the rollercoaster moves with sufficient speed and in a circle of the appropriate radius so that this happens.

As an example, consider a ride of radius 15 m in a simple vertical circle (Figure 2.6.8).

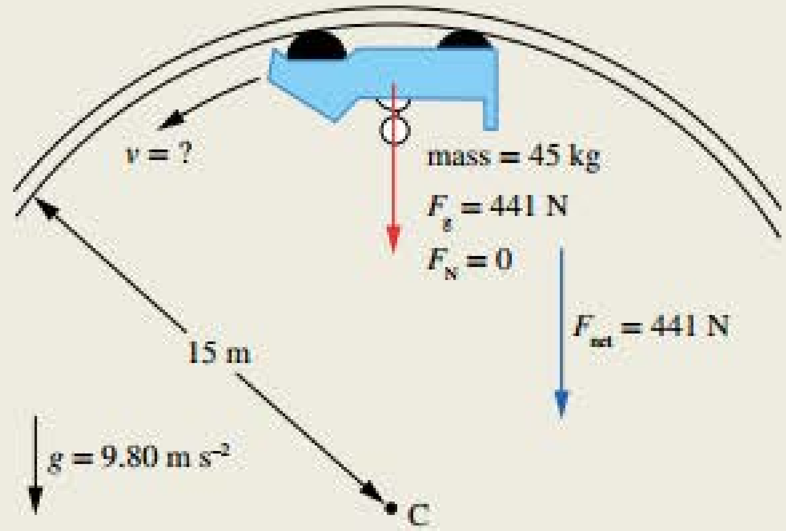


FIGURE 2.6.8 Rollercoaster travelling upside down through a loop. At the critical point where the rollercoaster just stays in contact with the track, the normal reaction force can be considered to be zero.

It is possible to calculate the speed that would ensure that a rider cannot fall out. Assume that the person has a mass of 45 kg and that g is $9.80\,m\,s^{-2}$. At the critical speed, the normal force, F_N, on the person will be zero. In other words, the seat will exert no force on the person at this speed. The centripetal force, F_{net}, is

$F_{net} = F_g + F_N$ but $F_N = 0$, so:

$F_{net} = F_g$

therefore $\frac{mv^2}{r} = mg$

$$v^2 = \frac{mgr}{m}$$
$$= gr$$
$$v = \sqrt{gr}$$
$$= \sqrt{9.80 \times 15}$$
$$= 12\,m\,s^{-1}$$

This speed is equal to $43\,km\,h^{-1}$ and is the minimum needed to prevent the riders from falling out for that radius. In practice, the rollercoaster would move with a speed much greater than this to ensure that there was a significant force between the riders and their seats. Corkscrew rollercoasters can travel at up to $110\,km\,h^{-1}$ and the riders can experience accelerations of up to $50\,m\,s^{-2}$ or (5g). So, safety harnesses are really only needed when the speed is below the critical value; their primary function is to prevent people from moving around while on the ride.

2.6 Review

SUMMARY

- The gravitational force, F_g, and normal force, F_N, must be considered when analysing the motion of an object moving in a vertical circle.
- If the normal force is greater than the gravitational force ($F_N > F_g$), the passenger or rider will feel heavier.
- If the normal force is less than the gravitational force ($F_N < F_g$), the passenger or rider will feel lighter.
- In vertical circular motion, the gravitational force always acts vertically downwards regardless of position around the circle, the net force always acts towards the centre of the circle, and the normal force always acts between the seat and the passenger or rider.
- The normal force and the gravitational force must add together as vectors in a force diagram to give the resultant as the net force.

  $F_{net} = F_N + F_g$
- At the point where a moving object lifts off from its circular path, the normal force is zero. The object will be moving with a centripetal acceleration that is equal to that due to gravity (9.80 m s^{-2} down).
- Problems relating to motion in vertical circles can also be solved using an energy approach. This involves using the equation:

  $E_m = E_k + E_g = \frac{1}{2}mv^2 + mg\Delta h$

KEY QUESTIONS

In the following questions, assume that $g = 9.80$ m s^{-2} and ignore the effects of air resistance.

The following information applies to questions 1 and 2.

A yo-yo is swung with a constant speed in a vertical circle.

1 a Describe the magnitude of the acceleration of the yo-yo along its path.

b At which point in the circular path is there the greatest amount of tension in the string?

c At which point in the circular path is there the lowest amount of tension in the string?

d At which point is the string most likely to break?

2 If the yo-yo has a mass of 80 g and the radius of the circle is 1.5 m, find the minimum speed that this yo-yo must have at the top of the circle so that the cord does not slacken.

3 A car of mass 800 kg encounters a speed hump of radius 10 m. The car drives over the hump at a constant speed of 14.4 km h^{-1}.

a Name all the vertical forces acting on the car when it is at the top of the hump.

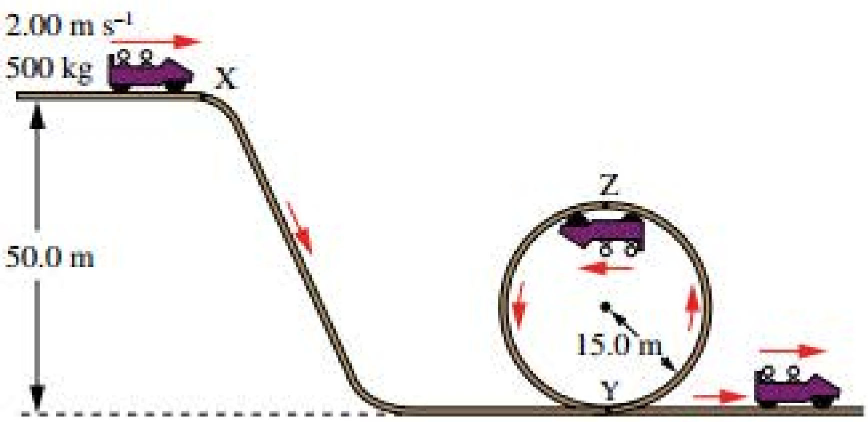
b Calculate the resultant force acting on this car when it is at the top of the hump.

c After travelling over the hump, the driver remarked to her passenger that she felt lighter as the car moved over the top of the speed hump. Is this possible? Explain your answer.

d What is the maximum speed (in km h^{-1}) that this car can have at the top of the hump and still have its wheels in contact with the road?

The following information applies to questions 4 and 5.

A popular amusement park ride is the 'loop-the-loop', in which a cart descends a steep incline at point X, enters a circular loop at point Y, and makes one complete revolution of the circular loop. A cart, whose total mass is 500 kg, carries the passengers at a speed of 2.00 m s^{-1} when it begins its descent at point X from a vertical height of 50.0 m.



4 a Calculate the speed of the cart at point Y.

b What is the speed of the cart at point Z?

c Calculate the normal force acting on the cart at Z.

5 What is the minimum speed that the cart can have at point Z and still stay in contact with the track?

2.6 Review *continued*

The following information applies to questions 6 and 7.
A stunt pilot appearing at an air show decides to perform a vertical loop so that she is upside down at the top of the loop. During the stunt she maintains a constant speed of 35 m s^{-1} while completing the 100 m radius loop.

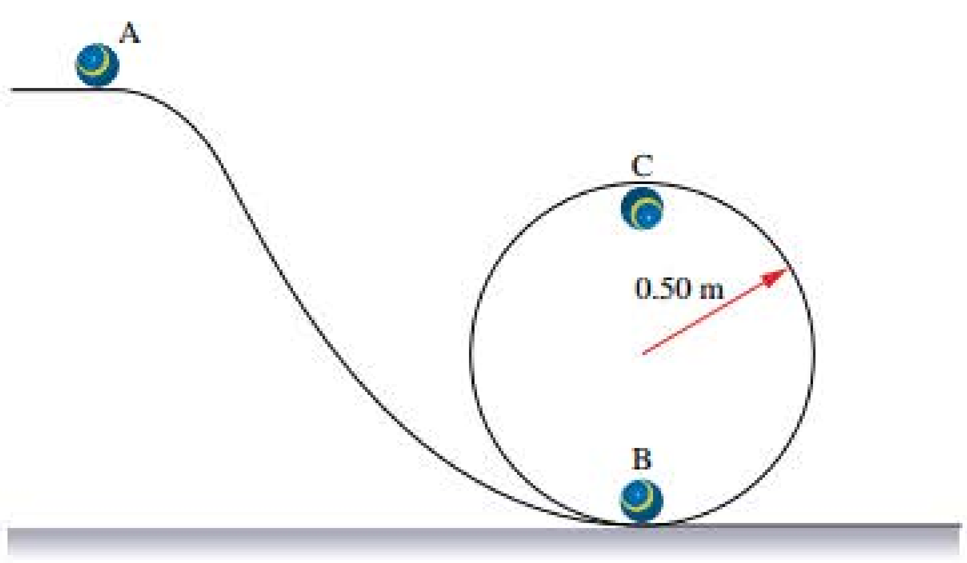
6 Calculate the apparent weight of the 80 kg pilot when she is at the top of the loop.

7 What minimum speed would the pilot need at the top of the vertical loop in order to experience zero normal force from the seat (i.e. to feel weightless)?

8 The maximum value of acceleration that the human body can safely tolerate for short time intervals is nine times that due to gravity. Calculate the maximum speed with which a pilot could safely pull out of a circular dive of radius 400 m.

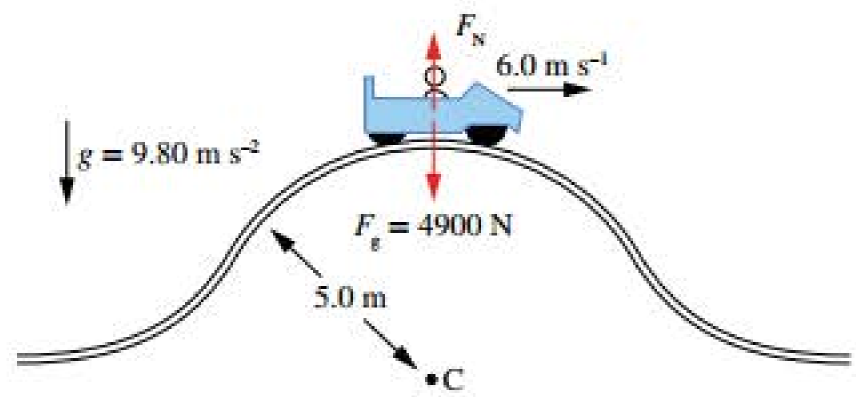
9 A skateboarder of mass 55 kg is skating on a half-pipe of radius 2.0 m. At the lowest point of the half-pipe, the speed of the skater is 6.0 m s^{-1}. (Ignore air resistance and friction.)
- **a** What is the acceleration of the skater at this point? Indicate both magnitude and direction.
- **b** Calculate the size of the normal force acting on the skater at this point.

10 A ball bearing of mass 25 g is rolled along a smooth track in the shape of a loop-the-loop. The ball bearing is given a launch speed at A so that it just maintains contact with the track as it passes through point C. Ignore air resistance and friction.



- **a** Determine the magnitude of the acceleration of the ball bearing as it passes point C.
- **b** How fast is the ball bearing travelling at point C?

11 On the Mad Mouse ride, a cart of mass 500 kg encounters a hump of radius 5.0 m. The cart's speed at the top of the hump is 6.0 m s^{-1}.



- **a** Calculate the magnitude and direction of the resultant force acting on this cart when it is at the top of the hump.
- **b** Calculate the magnitude and direction of the normal force acting on this cart when it is at the top of the hump.
- **c** What is the maximum speed that this cart can have at the top of the hump and still have its wheels in contact with the track?

2.7 Satellite motion

When Isaac Newton developed his law of universal gravitation, as discussed in Chapter 1, he was building on work previously done by Nicolaus Copernicus, Johannes Kepler and Galileo Galilei. Copernicus had proposed a Sun-centred (heliocentric) solar system. Galileo had developed laws relating to motion near the Earth's surface, and Kepler had devised rules concerned with the motion of the planets. Kepler published his laws on the motion of planets 80 years before Newton published his law of universal gravitation.

In this section, you will look at how Newton synthesised the work of Galileo and Kepler, and proposed that the force that was causing an apple to fall to Earth was the same force that was keeping the Moon in its orbit. Newton was the first to propose that satellites could be placed in orbit around Earth, almost 300 years before it was technically possible to do this. Now, thousands of artificial satellites are in orbit around Earth and are an essential part of modern life (Figure 2.7.1).

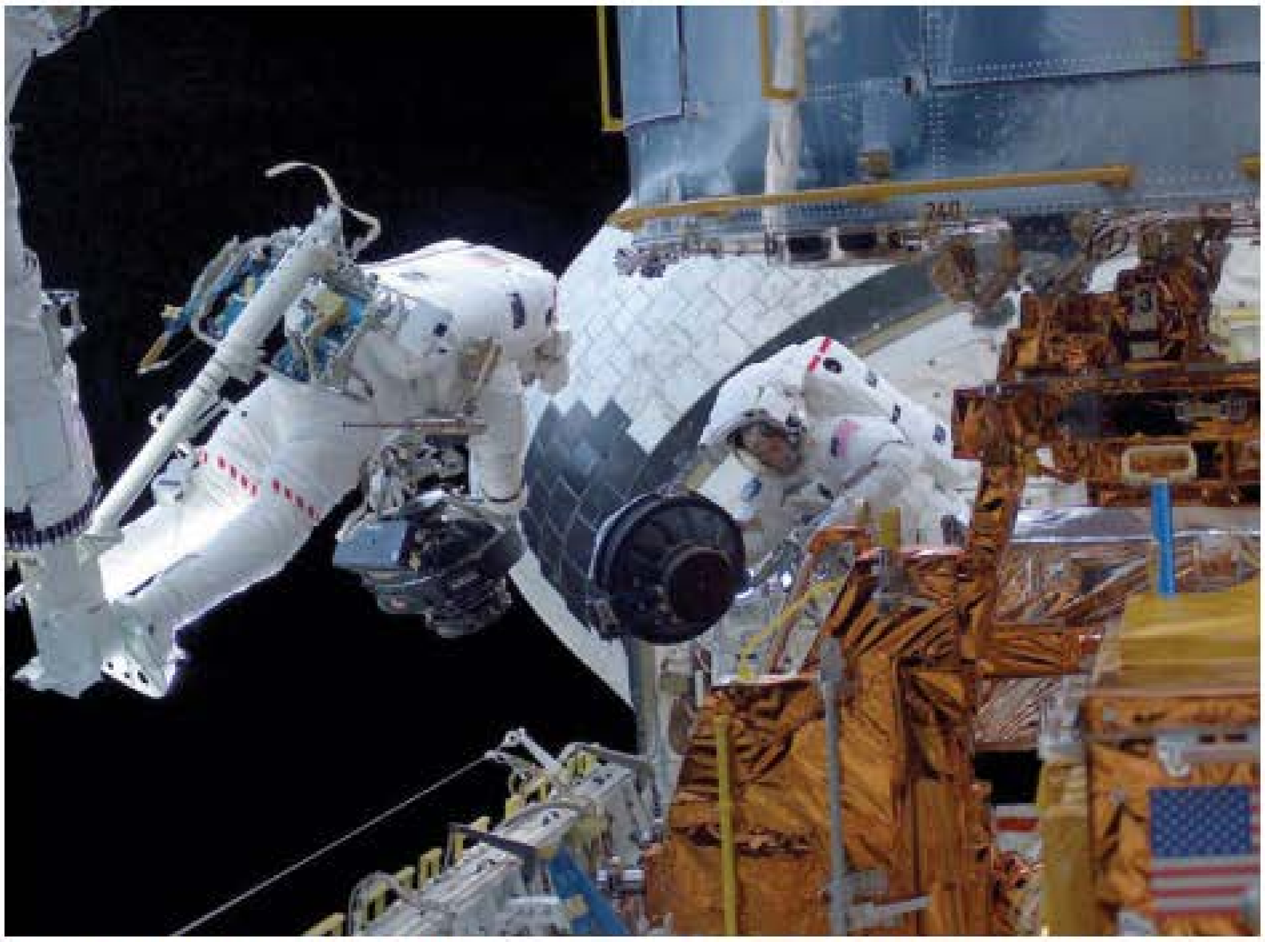


FIGURE 2.7.1 Astronauts on a repair mission to the Hubble Space Telescope (HST) in 1994. The satellite initially malfunctioned, but the repair was successful and the HST is still going strong.

NEWTON'S THOUGHT EXPERIMENT

A **satellite** is an object in a *stable orbit* around another object. Isaac Newton developed the notion of satellite motion while working on his theory of gravitation. He was comparing the motion of the Moon with the motion of a falling apple and realised that it was the gravitational force of attraction towards the Earth that determined the motion of both objects (Figure 2.7.2). He reasoned that if this force of gravity was not acting on the Moon, the Moon would move at constant speed in a straight line at a tangent to its orbit.

Newton proposed that the Moon, like the apple, was also falling. It was continuously falling to Earth without actually getting any closer to it. He devised a thought experiment in which he compared the motion of the Moon with the motion of a cannonball fired horizontally from the top of a high mountain.

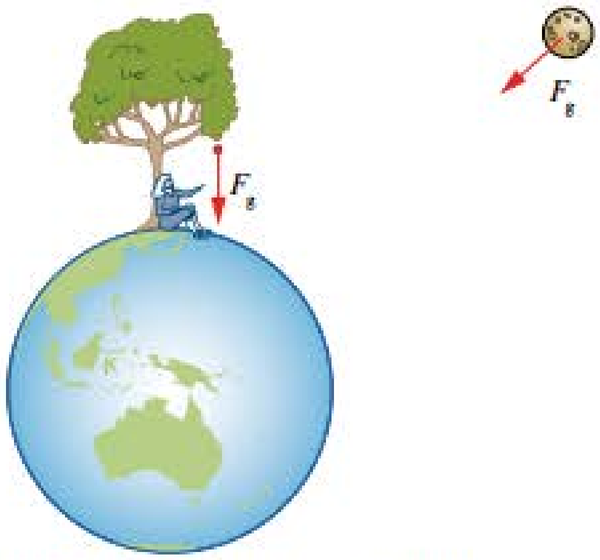


FIGURE 2.7.2 Newton realised that the gravitational attraction of the Earth (F_g) was determining the motions of both the Moon and the apple.

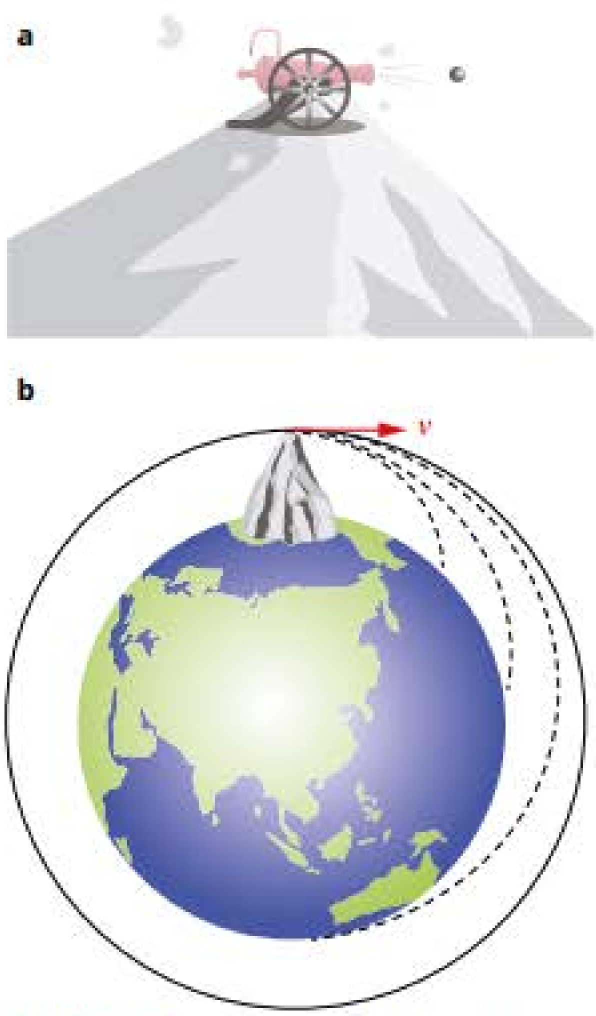


FIGURE 2.7.3 These diagrams show how a projectile that was fired fast enough from a very high mountain (a) would fall all the way around the Earth and become an Earth satellite (b).

FIGURE 2.7.4 Johannes Kepler was the first to work out that the planets do not travel in circular paths, but rather in elliptical paths.

His thought experiment is illustrated in Figure 2.7.3. In this thought experiment, if the cannonball was fired at a low speed, it would not travel a great distance before gravity pulled it to the ground (see the shortest dashed line in Figure 2.7.3(b)). If it was fired with a greater velocity, it would follow a less curved path and land a greater distance from the mountain (see the next two dashed lines in Figure 2.7.3(b)). Newton reasoned that, if air resistance was ignored and if the cannonball was fired fast enough, it could travel around the Earth and reach the place from where it had been launched (shown by the solid circular line in Figure 2.7.3(b)). At this speed, it would continue to circle the Earth indefinitely even though the cannonball has no propulsion system.

In reality, satellites could not orbit the Earth at low altitudes, because of air resistance. Nevertheless, Newton had proposed the notion of an artificial satellite hundreds of years before one was actually launched. Any object placed at the right altitude with enough speed would simply continue in its orbit.

KEPLER'S LAWS

Kepler, a German astronomer (depicted in Figure 2.7.4), published his three laws regarding the motion of planets in 1609, about 80 years before Newton's law of universal gravitation was published. Kepler was analysing the motion of the planets in orbit around the Sun, but these laws can be used for any satellite in orbit around any central mass.

Kepler's laws are as follows:

1 The planets move in elliptical orbits with the Sun at one focus.
2 The line connecting a planet to the Sun sweeps out equal areas in equal intervals of time (Figure 2.7.5).
3 For every planet orbiting the Sun, the ratio of the cube of the average orbital radius, r, to the square of the period, T, of revolution is the same; that is, $\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$ = constant, and so $\frac{r_A^3}{T_A^2} = \frac{r_B^3}{T_B^2}$ for planet A and planet B.

Kepler's first two laws proposed that planets moved in elliptical paths from the furthest point (*aphelion*) to the closest point (*perihelion*). The closer the planet was to the Sun, the faster it moved. It took Kepler many months of laborious calculations to arrive at his third law. Newton used Kepler's laws to justify the inverse square relationship. In fact, Kepler's third law can be deduced, for circular orbits, from Newton's law of universal gravitation.

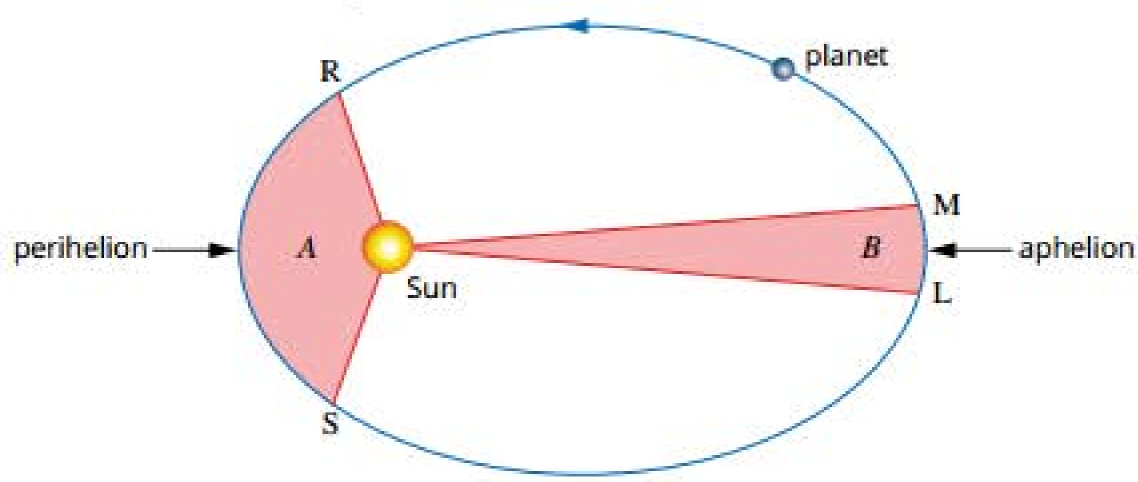


FIGURE 2.7.5 The planets, which are natural satellites of the Sun, orbit in elliptical paths with the Sun at one focus. Their speeds vary continually, and they are fastest when closest to the Sun. A line joining a planet to the Sun will sweep out equal areas in equal times. So, for example, the time it takes to move from R to S is equal to the time it takes to move from L to M, and so area *A* is the same as area *B*.

CALCULATING THE ORBITAL PROPERTIES OF SATELLITES

The speed, v, of a satellite can be calculated from its motion for one revolution. It will travel a distance equal to the circumference of the circular orbit, $2\pi r$, in the time of one period, T.

The speed, v, of a satellite in a circular orbit is given by:

$$v = \frac{\text{distance}}{\text{time}} = \frac{2\pi r}{T}$$

where r is the radius of the orbit (m)

T is the time for one revolution, or the period (s)

The centripetal acceleration of a satellite can be determined from the gravitational field strength at its location. Satellites are in free-fall; therefore, the only force acting is gravity, F_g. The International Space Station (ISS) is in orbit at a distance from Earth where g is $8.8\ \text{N kg}^{-1}$, and so it orbits with a centripetal acceleration of $8.8\ \text{m s}^{-2}$.

The centripetal acceleration, a, of a satellite can also be calculated by considering its circular motion. The equation for speed given above can be substituted into the centripetal acceleration formula to give:

$$a = \frac{v^2}{r} \text{ and since } v = \frac{2\pi r}{T}$$

$$\text{then } \frac{v^2}{r} = \frac{4\pi^2 r}{T^2}$$

The centripetal acceleration of the satellite is equal to the gravitational field strength at the location of its orbit, so, using the gravitational field strength equation from Chapter 1, we can give the following expression.

The centripetal acceleration, a, of a satellite in circular orbit is given by:

$$a = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2} = \frac{GM}{r^2} = g$$

where v is the speed of the satellite (m s^{-1})

r is the radius of the orbit from the centre of mass of the object it is orbiting (m)

T is the period of orbit (s)

M is the central mass (kg)

g is the gravitational field strength at r (N kg^{-1})

G is the gravitational constant, $6.67 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2}$

These relationships can be manipulated to determine any feature of a satellite's motion: its speed, radius of orbit or period of orbit. They can also be used to find the mass of the central body, M, around which the satellite orbits.

In the same way as with freely falling objects at the Earth's surface, the mass of the satellite itself has *no effect* on any of these orbital properties.

The gravitational force, F_g, acting on the satellite can then be found by using Newton's second law.

The gravitational force on a satellite of mass m in a stable circular orbit is given by:

$$F_g = \frac{mv^2}{r} = \frac{4\pi^2 rm}{T^2} = \frac{GMm}{r^2} = mg$$

Worked example 2.7.1

WORKING WITH KEPLER'S LAWS

Determine the orbital speed of the Moon, assuming it is in a circular orbit of radius 384 000 km from the centre of the Earth. Take the mass of the Earth to be 5.97×10^{24} kg and use $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

Thinking	Working
Ensure that the variables are in their standard units.	$r = 384\,000$ km $= 3.84 \times 10^8$ m
Choose the appropriate relationship between the orbital speed, *v*, and the data that has been provided.	$a = g = \frac{GM}{r^2} = \frac{v^2}{r}$
Make *v*, the orbital speed, the subject of the equation.	$v = \sqrt{\frac{GM}{r}}$
Substitute in values and solve for the orbital speed, *v*.	$v = \sqrt{\frac{(6.67 \times 10^{-11}) \times (5.97 \times 10^{24})}{3.84 \times 10^8}}$ $= 1.02 \times 10^3$ m s^{-1}

Worked example: Try yourself 2.7.1

WORKING WITH KEPLER'S LAWS

Determine the orbital speed of a satellite, assuming it is in a circular orbit of radius of 42 100 km from the centre of the Earth. Take the mass of the Earth to be 5.97×10^{24} kg and use $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

HOW NEWTON DERIVED KEPLER'S THIRD LAW USING ALGEBRA

It took Kepler many months of trial-and-error calculations to arrive at his third law:

$$\frac{r^3}{T^2} = \text{constant}$$

Newton was able to use some clever algebra to derive this from his law of universal gravitation and the centripetal force equation:

$$F_c = F_g$$

$$\frac{m4\pi^2 r}{T^2} = \frac{GMm}{r^2}$$

$$\therefore \frac{r^3}{T^2} = \frac{GM}{4\pi^2}$$

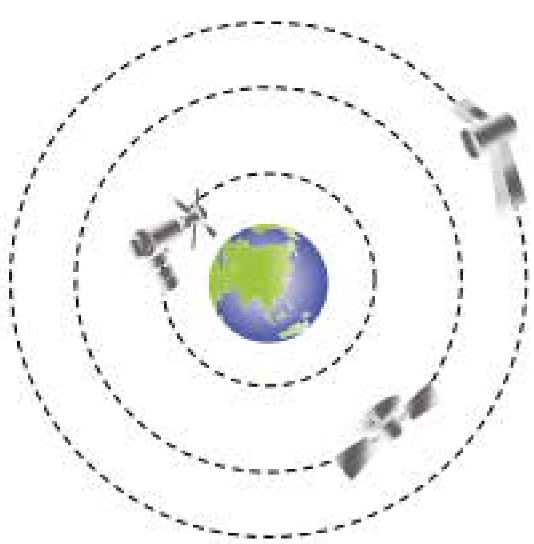


FIGURE 2.7.6 These three satellites are at different distances from the centre of the Earth and hence according to Kepler's third law will have different orbital periods. For all three, the ratio $\frac{r^3}{T^2}$ will equal the same constant value.

For any central mass, M, the term $\frac{GM}{4\pi^2}$ is constant and the ratio $\frac{r^3}{T^2}$ is equal to this constant value for all its satellites (Figure 2.7.6).

So, for example, if you know the orbital radius, r, and period, T, of one of the moons of Saturn, you could calculate $\frac{r^3}{T^2}$ and use this as a constant value for all of Saturn's moons. If you knew the period, T, of a different satellite of Saturn, it would then be straightforward to calculate its orbital radius, r.

Worked example 2.7.2

SATELLITES IN ORBIT

Ganymede is the largest of Jupiter's moons. It has a mass of 1.66×10^{23} kg, an orbital radius of 1.07×10^6 km and an orbital period of 6.18×10^5 s (7.15 days).

a Use Kepler's third law to calculate the orbital radius (in km) of Europa, another moon of Jupiter, which has an orbital period of 3.55 days.

Thinking	Working
Note down the values for the known satellite. You can work in days and km as this question involves ratio.	Ganymede: $r = 1.07 \times 10^6$ km $T = 7.15$ days
For all satellites of a central mass, $\frac{r^3}{T^2}$ = constant. Work out this ratio for the known satellite.	$\frac{r^3}{T^2}$ = constant $= \frac{(1.07 \times 10^6)^3}{7.15^2}$ $= 2.40 \times 10^{16}$
Use this constant value with the ratio for the satellite in question. Make sure T is in days to match the ratio calculated in the previous step.	Europa: $T = 3.55$ days, $r = ?$ $\frac{r^3}{T^2}$ = constant $\frac{r^3}{3.55^2} = 2.40 \times 10^{16}$
Make r^3 the subject of the equation.	$r^3 = 2.40 \times 10^{16} \times 3.55^2$ $= 3.02 \times 10^{17}$
Solve for r. The unit for r is km as the original ratio was calculated using km.	$r = \sqrt[3]{3.02 \times 10^{17}}$ $= 6.71 \times 10^5$ km Note: Europa has a shorter period than Ganymede, so you should expect Europa to have a smaller orbit than Ganymede.

b Use the orbital data for Ganymede to calculate the mass of Jupiter.

Thinking	Working
Note down the values for the known satellite. You must work in SI units to find the mass value in kg.	$r_G = 1.07 \times 10^9$ m $T_G = 6.18 \times 10^5$ s $m_G = 1.66 \times 10^{23}$ kg $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2} $M_J = ?$
Select the expressions from the equation for centripetal acceleration that best suit your data. $a = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2} = \frac{GM}{r^2} = g$	Use the 3rd and 4th terms of the expression. $\frac{4\pi^2 r}{T^2} = \frac{GM}{r^2}$ These two expressions use the given variables r and T, and the constant G, so that a solution may be found for M.

PHYSICSFILE

Ganymede

Jupiter is orbited by more than 60 known satellites, the biggest of which is Ganymede. Ganymede is very large. It is the biggest of all the moons in the solar system and is even bigger than the planet Mercury.

FIGURE 2.7.7 Ganymede

Transpose to make M the subject.	$M = \frac{4\pi^2 r^3}{GT^2}$
Substitute values and solve.	$M = \frac{4\pi^2(1.07\times10^9)^3}{6.67\times10^{-11}\times(6.18\times10^5)^2}$ $= 1.90 \times 10^{27}$ kg
c Calculate the orbital speed of Ganymede in km s^{-1}.	
Thinking	**Working**
Note values you will need to use in the equation $v = \frac{2\pi r}{T}$.	Ganymede: $r = 1.07 \times 10^6$ km $T = 6.18 \times 10^5$ s $v = ?$
Substitute values and solve. The answer will be in km s^{-1} if r is expressed in km.	$v = \frac{2\pi r}{T}$ $= \frac{2\pi\times1.07\times10^6}{6.18\times10^5}$ $= 10.9$ km s^{-1}

Worked example: Try yourself 2.7.2

SATELLITES IN ORBIT

Callisto is the second largest of Jupiter's moons. It is about the same size as the planet Mercury. Callisto has a mass of 1.08×10^{23} kg, an orbital radius of 1.88×10^6 km and an orbital period of 1.44×10^6 s (16.7 days).

a Use Kepler's third law to calculate the orbital radius (in km) of Europa, another moon of Jupiter, which has an orbital period of 3.55 days.

b Use the orbital data for Callisto to calculate the mass of Jupiter.

c Calculate the orbital speed of Callisto in km s^{-1}.

2.7 Review

SUMMARY

- A satellite is an object that is in a stable orbit around a larger central mass.
- The only force acting on a satellite is the gravitational attraction between it and the central body.
- Satellites are in continual free-fall. They move with a centripetal acceleration that is equal to the gravitational field strength at the location of their orbit.
- The speed of a satellite, v, is given by:
  $$v = \frac{2\pi r}{T}$$
- For a satellite in a circular orbit:
  $$a = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2} = \frac{GM}{r^2} = g$$
- The gravitational force acting on a satellite in a circular orbit is given by:
  $$F_g = \frac{mv^2}{r} = \frac{4\pi^2 rm}{T^2} = \frac{GMm}{r^2} = mg$$
- For any central body of mass M:
  $$\frac{r^3}{T^2} = \frac{GM}{4\pi^2} = \text{a constant}$$
  so knowing another satellite's orbital radius from the centre of mass of the central body, r, enables its period, T, to be determined.

KEY QUESTIONS

1 Which of the following is correct?
  A Earth is a satellite of Mars.
  B The Moon is a satellite of the Sun.
  C Earth is a satellite of the Sun.
  D The Sun is a satellite of Earth.

2 A satellite of mass M is in a circular orbit around the Moon. A module of mass M then attaches to the original satellite so that the combined mass is now $2M$. How does this affect the orbital properties of the satellite?
  A The speed of the satellite will decrease and the period will increase.
  B Both the speed and period of the satellite will decrease.
  C Both the speed and period of the satellite will increase.
  D Nothing will change.

3 The gravitational field strength at the location where the Optus D1 satellite is in stable orbit around the Earth is equal to $0.22\,\text{N kg}^{-1}$. The mass of this satellite is $2.3 \times 10^3\,\text{kg}$.
  a Using only the information given, calculate the magnitude of the acceleration of this satellite as it orbits.
  b Calculate the net force acting on this satellite as it orbits.

4 One of Saturn's moons is Atlas, which has an orbital radius of $1.37 \times 10^5\,\text{km}$ and a period of 0.60 days. The largest of Saturn's moons is Titan. It has an orbital radius of $1.20 \times 10^6\,\text{km}$. What is the orbital period of Titan in days?

Chapter review

02

KEY TERMS

air resistance
banked track
centripetal force
conserved
design speed
frequency
friction
inclined plane
law of conservation of energy
mechanical energy
period
projectile
satellite
tangential

1 Diana rolls a bowling ball down a smooth straight ramp. Choose the option below that best describes the way the ball will travel.
  A with constant speed
  B with constant acceleration
  C with decreasing speed
  D with increasing acceleration

2 A bowling ball is rolling down a smooth track that is inclined at 30° to the horizontal.
  a What is the magnitude of the acceleration of the ball?
  b How does the magnitude of the normal force that acts on the ball compare to its weight?

3 A marble is rolled from rest down a smooth slide that is 2.5 m long. The slide is inclined at an angle of 30° to the horizontal.
  a Calculate the acceleration of the marble.
  b What is the speed of the marble as it reaches the end of the slide?

4 Marshall has a mass of 54 kg and he is riding his 3.0 kg skateboard down a 5.0 m long ramp that is inclined at an angle of 65° to the horizontal. Ignore friction when answering parts **a–d**.
  a Calculate the magnitude of the normal force acting on Marshall and his skateboard.
  b What is the acceleration of Marshall on his skateboard as he travels down the ramp?
  c What is the net force acting on Marshall and his board when no friction acts?
  d If Marshall's initial speed is zero at the top of the ramp, calculate his final speed as he reaches the bottom of the ramp.
  e Marshall now stands halfway up the incline while holding his board in his hands. Friction now acts on Marshall. Calculate the frictional force acting on Marshall while he is standing stationary on the ramp.

5 A very high waterslide is 50.0 m tall and is inclined at an angle of 70° to the horizontal. It is known that riders reach a speed of 100 km h^{-1} on this slide. Do not assume friction is negligible.
  a For a 70.0 kg teenager using the slide, calculate the net force on the teenager as he slides.
  b For the same teenager, calculate the magnitude of the average frictional force opposing the motion.
  c If the friction acts on the teenager to slow him down, what is the reaction force to this?
  d What is the reaction force to the teenager's weight force?

6 A toy car is moving at 2.5 m s^{-1} as it rolls off a horizontal table. The car takes 1.0 s to reach the floor.
  a How far does the car land from the foot of the table?
  b What is the magnitude and direction of acceleration when the car is halfway to the floor?

7 A bowling ball of mass 7.5 kg travelling at 10 m s^{-1} rolls off a horizontal table that is 0.97 m high.
  a What is the horizontal speed of the ball as it strikes the floor?
  b What is the vertical speed of the ball as it strikes the floor?
  c Calculate the speed of the ball as it reaches the floor.

8 In a tennis match, a tennis ball is hit from a height of 1.2 m with an initial velocity of 16 m s^{-1} at an angle of 50° to the horizontal. Ignore drag forces for the following questions.
  a What is the initial horizontal speed of the ball?
  b What is the initial vertical speed of the ball?
  c What is the maximum height that the ball reaches above the court surface?

9 Three physics friends, Tait, Fred and Neil, are having an argument while discussing projectile motion. They consider a soccer ball kicked straight off the ground at an angle of 50.0° and a speed of $35.0\,m\,s^{-1}$. The ball hits the ground at the same level some distance away; there is negligible air resistance. Fred argues that, at the end of its journey, the ball will have fallen from a great height and so should be travelling faster than when it was kicked, due to the acceleration of gravity. Neil, however, claims that, since the force provided by the kicker is no longer there, the ball will slow down by the time it lands. Finally, Tait says that the ball will be travelling at the same speed when it lands as when it was kicked due to the conservation of energy. Which of the three friends is correct and why? Why are the others wrong?

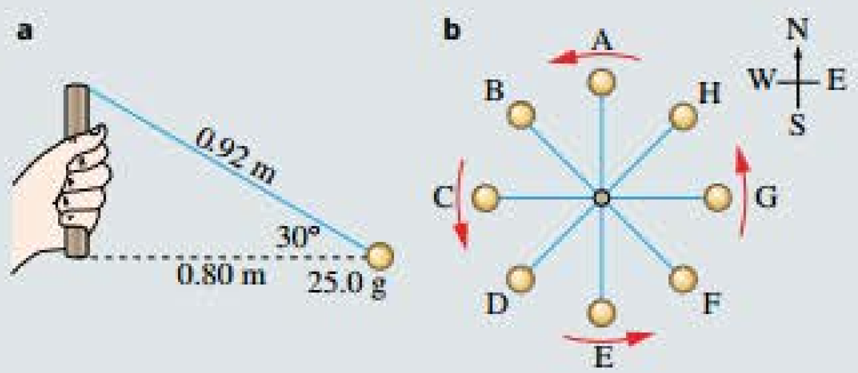
10 An orange of mass 100 g is tossed horizontally at $6.0\,m\,s^{-1}$ from a height of 2.0 m. Ignore air resistance and use $g = 9.80\,m\,s^{-2}$ when answering these questions.
- **a** What is the initial kinetic energy of the orange?
- **b** Calculate the initial potential energy of the orange.
- **c** What is the speed of the orange as it lands?

11 Calculate the gravitational potential energy that a 40.0 g bullet has when it has travelled 1000 m up into the air after having been fired from a gun.

12 A boy throws a 157 g cricket ball up into the air. It leaves his hand at a speed of $20.5\,m\,s^{-1}$.
- **a** Calculate the kinetic energy of the ball as it leaves the boy's hand.
- **b** What is the gravitational potential energy at the top of its flight, if air resistance is ignored?
- **c** Calculate the maximum height reached, if air resistance is ignored.

The following information applies to questions 13 and 14.

During a high-school physics experiment, a copper ball of mass 25.0 g was attached to a very light piece of steel wire 0.920 m long and whirled in a circle at 30.0° to the horizontal, as shown in diagram (a). The ball moves in a circular path of radius 0.800 m with a period of 1.36 s. The top view of the resulting motion of the ball is shown in diagram (b).



13 **a** Calculate the orbital speed of the ball.
- **b** What is the centripetal acceleration of the ball?
- **c** What is the magnitude of the centripetal force acting on the ball?

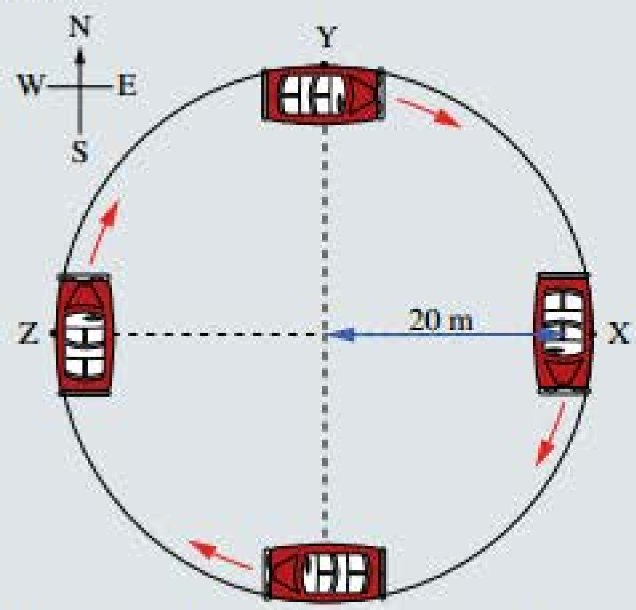
14 **a** Draw a diagram similar to diagram (a) that shows all the forces acting on the ball at this time.
- **b** What is the magnitude of the tension in the wire?

15 A radio-controlled car is travelling in a circular path of radius 10.0 m at a constant speed of $5.00\,m\,s^{-1}$.
- **a** What is the acceleration of the toy car?
- **b** What force is creating the circular motion of the car?

16 The Moon orbits the Earth once in 27.3 days in a circular orbit of radius 3.84×10^8 m.
- **a** Calculate the orbital speed of the Moon.
- **b** Calculate the net force keeping the Moon in orbit if the mass of the Moon is 7.36×10^{22} kg.

17 A geostationary communications satellite is at an altitude of 3.60×10^4 m. The Earth has an average radius 6.37×10^6 m and a period of rotation of 23 hours, 56 minutes and 5 seconds. Calculate the centripetal acceleration of the satellite.

18 A car of mass 1500 kg is driven at constant speed of $10\,m\,s^{-1}$ around a level, circular roundabout. The centre of mass of the car is always 20.0 m from the centre of the track.



- **a** What is the velocity of the car at point X?
- **b** What is the speed of the car at point Y?
- **c** What is the period of revolution for this car?
- **d** What is the acceleration of the car at point X?
- **e** Determine the size and direction of the unbalanced frictional force acting on the tyres at point X.

19 A track cyclist is riding at high speed on the steeply banked section of a velodrome ($\theta = 37°$). Which statement describes the size of the normal force acting on the cyclist at this point?
- **A** It is greater than the weight of the cyclist.
- **B** It is zero.
- **C** It is less than the weight of the cyclist.
- **D** It is equal to the weight of the cyclist.

20 Liam and Joe are riding on a cycling track with a turn that is banked at 40.0° to the horizontal. The radius of the track at this point is 30.0 m. Joe has a mass of 90.0 kg and is travelling at the track's design speed, while Liam has a mass of 70.0 kg and is travelling at less than the track's design speed.

- **a** Calculate the design speed of the track at this point.
- **b** Does Liam have any sideways friction acting on him from the track? If so, in what direction is the vector pointing, up or down the track?
- **c** Calculate the net force acting on Joe. Then, calculate the horizontal component of the normal force acting on Joe. Why is this result significant?
- **d** If Joe speeds up even further, will he necessarily skid off the track? Why or why not?

21 The Ferris wheel at an amusement park has an arm radius of 10 m and its compartments move with a constant speed of 5.0 m s^{-1}.

- **a** Calculate the normal force that a 50 kg boy would experience from the seat when at the:
  - **i** top of the ride
  - **ii** bottom of the ride.
- **b** After getting off the ride, the boy remarks to a friend that he felt lighter than usual at the top of the ride. Which option explains why he might feel lighter at the top of the ride?
  - **A** He lost weight during the ride.
  - **B** The strength of the gravitational field was weaker at the top of the ride.
  - **C** The normal force there was larger than the gravitational force.
  - **D** The normal force there was smaller than the gravitational force.

22 As part of a demonstration, a teacher swings a bucket half-filled with water in a vertical circle at high speed. No water spills from the bucket even when it passes the overhead position. Discuss the forces acting on the water when the bucket is directly overhead and indicate their directions. Indicate the direction and relative size of the water's acceleration as it passes this position.

23 Which description best describes the motion of astronauts when orbiting the Earth?

- **A** They float in a zero gravity environment.
- **B** They float in a reduced gravity environment.
- **C** They fall down very slowly due to the very small gravity.
- **D** They fall in a reduced gravity environment.

24 Select the statement below that correctly states how a satellite in a stable circular orbit 200 km above the Earth will move.

- **A** It will have an acceleration of 9.80 m s^{-2}.
- **B** It will have constant velocity.
- **C** It will have zero acceleration.
- **D** It will have acceleration of less than 9.80 m s^{-2}.

25 What can be said about an object if it is orbiting the Earth in space and appears to be weightless?

- **A** It is in free-fall.
- **B** It is in zero gravity.
- **C** It has no mass.
- **D** It is floating.

26 A low-Earth-orbit satellite X has an orbital radius of r and period T. A high-Earth-orbit satellite Y has orbital radius of $5r$. In terms of T, what is the orbital period of Y?

27 The planet Neptune has a mass of 1.02×10^{26} kg. One of its moons, Triton, has a mass of 2.14×10^{22} kg and an orbital radius equal to 3.55×10^{8} m.

- **a** Calculate the orbital acceleration of Triton.
- **b** Calculate the orbital speed of Triton.
- **c** Calculate the orbital period of Triton (in days).

28 Ceres, the first asteroid to be discovered, was found by Giuseppe Piazzi in 1801. Ceres has a mass of 7.0×10^{20} kg and a radius of 385 km.

- **a** What is the gravitational field strength at the surface of Ceres?
- **b** Determine the speed required by a satellite in order to remain in orbit 10.0 km above the surface of Ceres.

CHAPTER 03

Equilibrium of forces

Engineers and architects must use their physics knowledge to determine the forces that act within the buildings and other structures they design.

When all the forces acting on an object add up to a zero net force, the body is said to be in translational equilibrium. There will be no translational acceleration if the forces acting on the object are balanced. For a structure to be completely at rest and stable, it is not enough that it is simply in translational equilibrium. The structure or system must also be in rotational equilibrium. So the sum of all the torques acting about a point must be zero. When an object is in translational and rotational equilibrium it is said to be in static equilibrium.

This chapter centres on the concept of equilibrium, which describes the situation in which forces and torques are balanced. If there is equilibrium of translational forces then there will be no net translational force. If there is equilibrium of torque or moments then the object will not be rotating.

Science Understanding

- when an object experiences a net force at a distance from a pivot and at an angle to the lever arm, it will experience a torque or moment about that point

  This includes applying the relationship

  $\tau = rF\sin\theta$

- for a rigid body to be in equilibrium, the sum of the forces and the sum of the moments must be zero

  This includes applying the relationships

  $\Sigma F = 0$

  $\tau = rF\sin\theta$

  $\Sigma\tau = 0$

3.1 Torque

Many real-life situations involve objects that rotate about a pivot point, such as closing a door, using a spanner or turning a steering wheel (Figure 3.1.1). In these situations, a force acts to provide a turning effect or a **torque** (τ). Newton's laws use the concept of a force to help explain the motion of a body in a straight line. The concept of torque is used in exactly the same way to explain a change in the rotational motion of an object.

FIGURE 3.1.1 Applying a torque to the rim or arms of a steering wheel will cause it to turn.

TORQUE

When a turning effect is applied to an object, a number of factors must work together to cause the object to turn. There must be a **pivot point** around which the object will rotate; this is also referred to as the **axis of rotation**. There must be a force (F) applied to the object in such a way as to cause the object to rotate. For the force applied to cause rotation, the force must not be aligned with the pivot point. This condition also implies that there must be some distance from the **line of action** of the force and the pivot point.

EXTENSION

Torque and moment of a force

The terms 'torque' (τ) and 'moment of a force' (M), which is usually shortened to moment, are terms that are used interchangeably in physics at the high school level.

The difference between these terms is in the conventions of their use. Typically, torque is used for dynamic problems in which there is an angular acceleration, and so the speed at which an object is rotating is changing. Moments are typically used in static problems, often as reaction forces or internal forces in an object, and are balanced so that there is no angular acceleration. An example of how a torque might be used would be in the force applied by an axle to the wheel of a car to increase its speed of rotation. A moment might be used when calculating the reaction force at the attached point for a beam that is attached to a wall at one end and has a force applied on the other.

Both torque and moment are calculated using the equation $\tau = rF$.

Force and the pivot point

When analysing a rotating system, the position of the pivot point or axis of rotation must be considered (Figure 3.1.2). A wheel, for example, moves in a circular path around its axle. The imaginary line along the length of the axle is the axis of rotation.

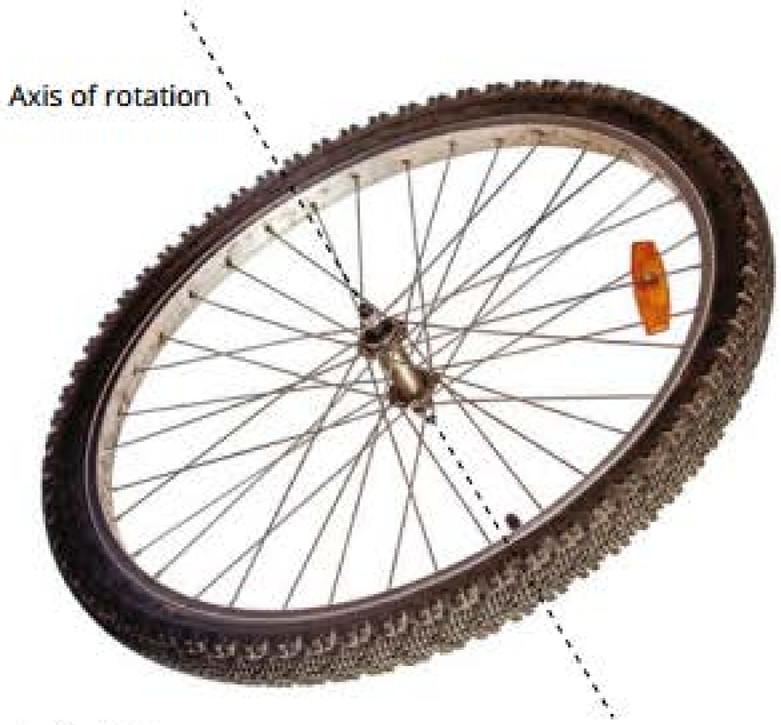


FIGURE 3.1.2 The axis of rotation.

A force applied directly towards or away from the axis of rotation of the wheel will not create a turning effect on the wheel (Figure 3.1.3).

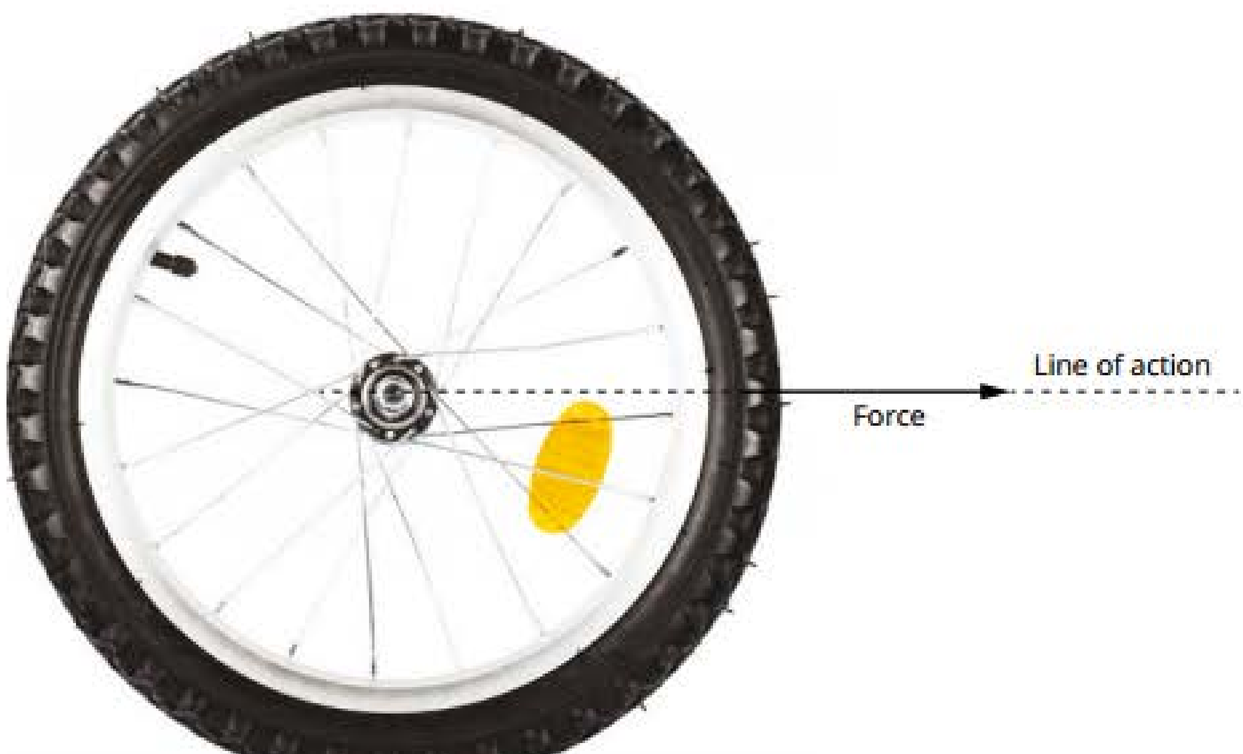


FIGURE 3.1.3 When the line of action of the force passes through the pivot point the wheel will not experience a torque.

Torque will be achieved by applying a force on the wheel at a point where the line of action of the force does not pass through the axis of rotation or the pivot point. The maximum effect is achieved when the force applied is at 90° to a line drawn from the axis of rotation to the point at which the force is applied (Figure 3.1.4).

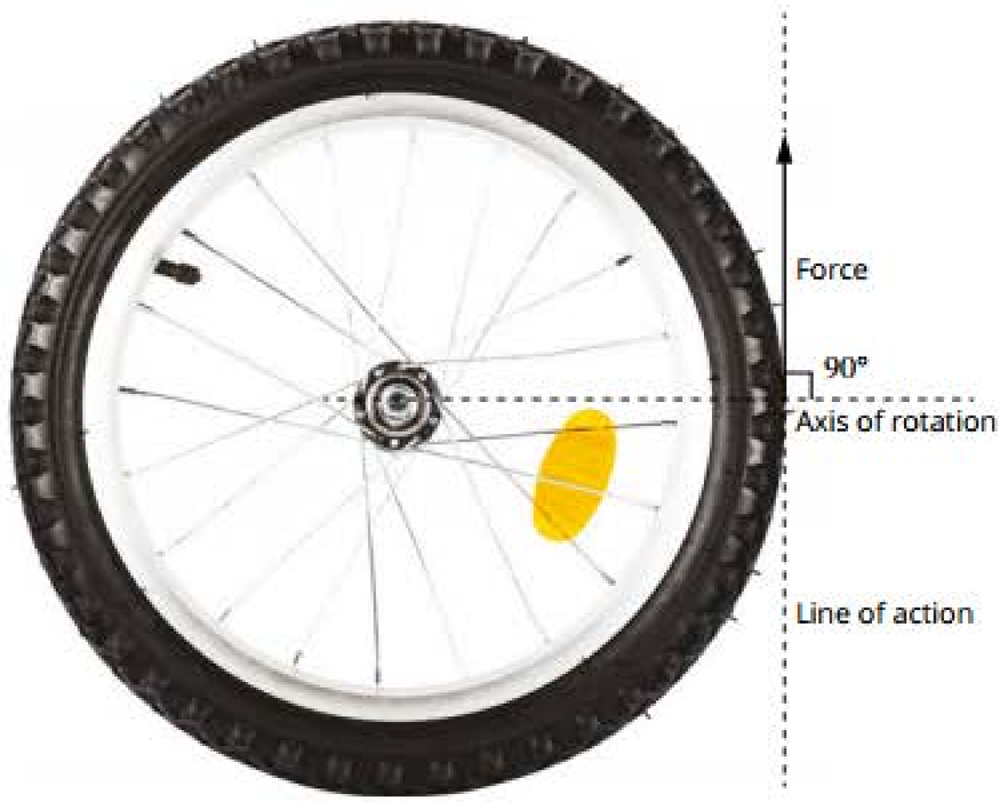


FIGURE 3.1.4 Maximum torque occurs when the force applied is perpendicular (90°) to the axis of rotation.

Magnitude of the force and the torque

The torque (τ) on an object is directly proportional to the magnitude of the force (F). If all other things are equal, a larger force will result in a larger torque (Figure 3.15).

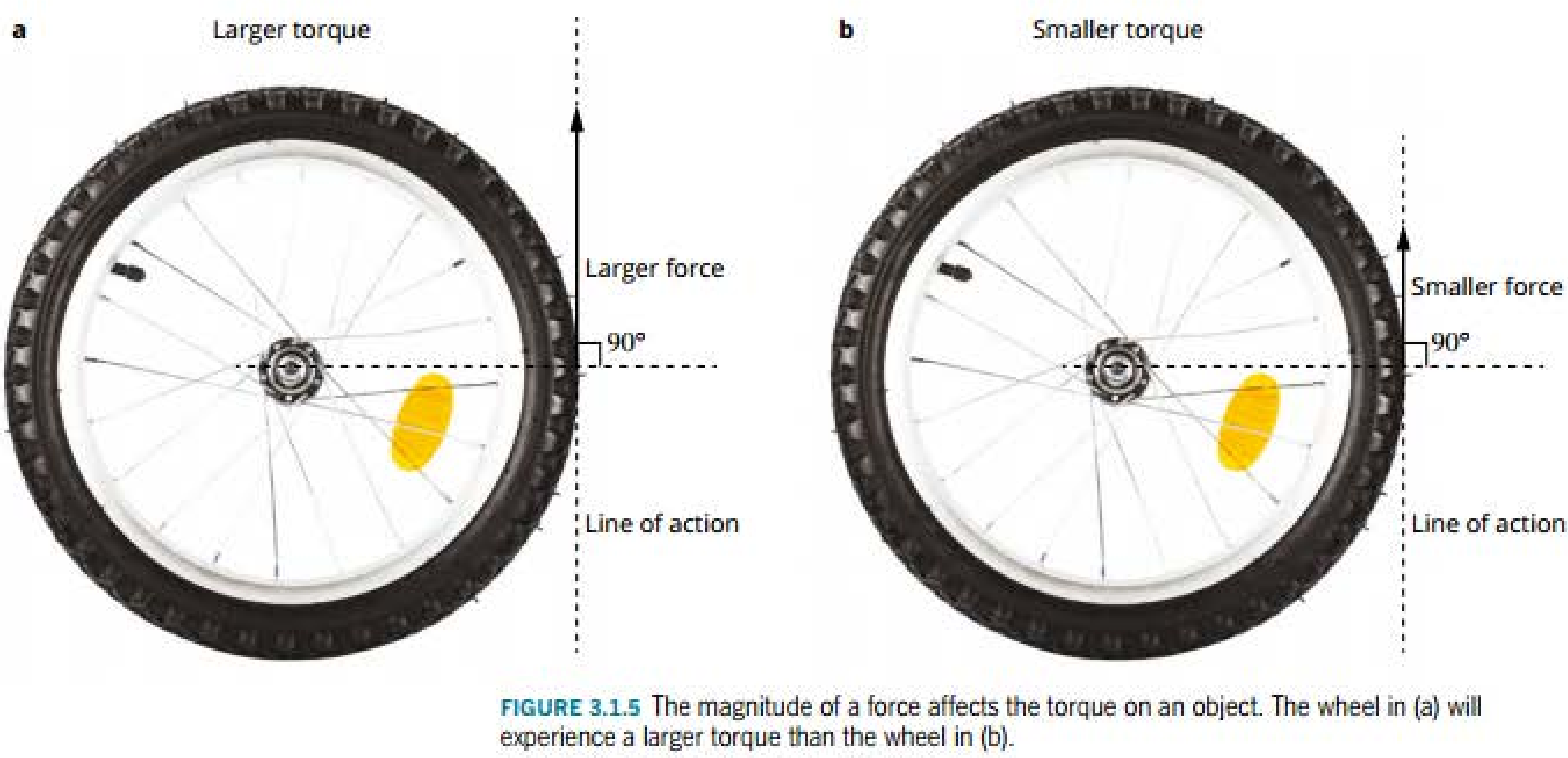


FIGURE 3.1.5 The magnitude of a force affects the torque on an object. The wheel in (a) will experience a larger torque than the wheel in (b).

Distance from the pivot point and torque

In addition to the size of the force and the angle at which it is acting, the amount of torque created is also directly proportional to the perpendicular distance between the axis of rotation and the line of action of the force. This perpendicular distance is called the **force arm** and is given the symbol $r_{\perp}$. Given that everything else is constant, then the larger the force arm or perpendicular distance ($r_{\perp}$), the larger the torque (τ) (Figure 3.1.6).

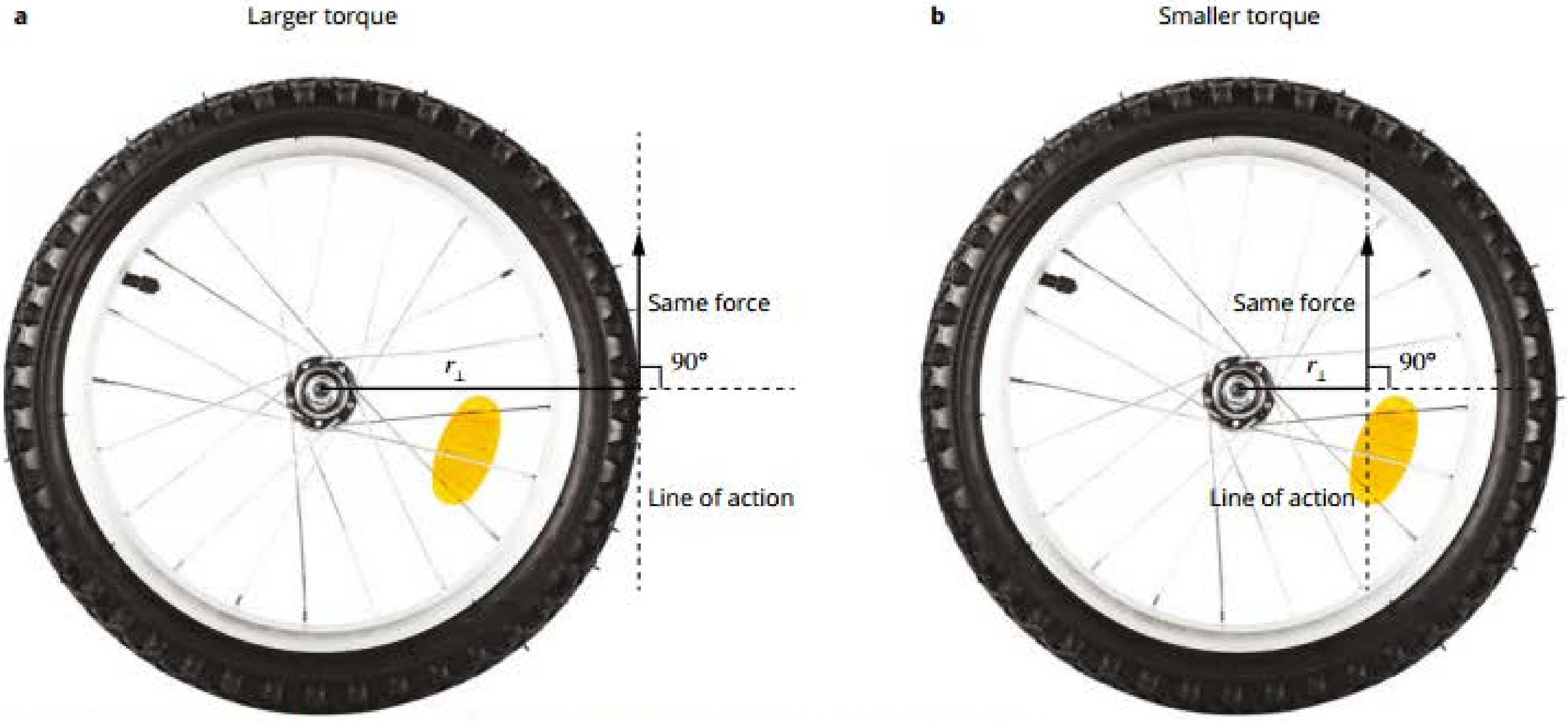


FIGURE 3.1.6 The perpendicular distance from the pivot point to the line of action of the force affects the torque on an object. The wheel in (a) will experience a larger torque than the wheel in (b).

The torque equation

The magnitude of the torque (τ) increases or decreases as the force (F) increases or decreases, and increases or decreases as the force arm or the perpendicular distance from the pivot to the line of action of the force ($r_{\perp}$) increases or decreases.

Hence, the formula for calculating the magnitude of the torque is:

$$\tau = r_{\perp} F$$

where τ = torque (N m)

$r_{\perp}$ = force arm (m)

F = force (N)

A rotating body rotates either clockwise or anticlockwise. You can determine the direction of the rotation as clockwise or anticlockwise by considering the context of the question. When two or more torques are acting, a sign convention can be applied to each torque, where clockwise is considered to be positive and anticlockwise is considered as negative. In this way, the torques can be added to determine the resultant torque.

Worked example 3.1.1

CALCULATING TORQUE

A bus driver applies a force of 45.0 N on the steering wheel of a bus as it turns a right-hand corner. The radius of the steering wheel is 30.0 cm. If the force is applied at 90° to the radius, calculate the torque on the steering wheel.	
Thinking	**Working**
Identify the variables involved and state them in their standard form.	$\tau = ?$ $r_{\perp} = 0.300\,m$ $F = 45.0\,N$
Apply the equation for torque. Rearrange as necessary.	$\tau = r_{\perp}F$ $= 0.300 \times 45.0$ $\tau = 13.5\,N$
State the answer including the appropriate direction (clockwise or anticlockwise).	$\tau = 13.5\,N$ clockwise

Worked example: Try yourself 3.1.1

CALCULATING TORQUE

A force of 255 N is required to apply a torque on the steering wheel of a sports car as it turns left. The force is applied at 90° to the 15.5 cm radius of the steering wheel. Calculate the torque on the steering wheel.

Torque on different objects

A torque doesn't need to be acting on a circular object. Any object can rotate about a point if a force is applied where the line of action of the force is not acting through the pivot point.

Spanners apply a torque to a nut or bolt; the pivot point is the bolt and a force is applied at right angles to the spanner (Figure 3.1.7).

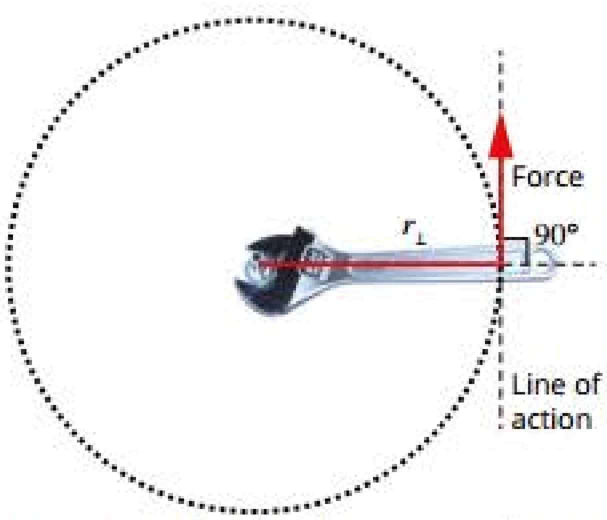


FIGURE 3.1.7 An adjustable spanner applies a torque to the nut of a nut and bolt system.

Longer spanners can apply a greater torque on a nut than a shorter spanner. This can be useful when trying to remove tight nuts or bolts, such as wheel nuts on a car (Figure 3.1.8). Some wheel-nut spanners have extendable handles, which enables extra torque for loosening very tight nuts or tightening the nuts sufficiently that they won't come loose.

FIGURE 3.1.8 Loosening the wheel nuts with an extended handle spanner will increase the torque on the nut.

Worked example 3.1.2

CALCULATING PERPENDICULAR DISTANCE

A car driver can apply a maximum force of 845 N on a wheel-nut spanner that is adjustable up to 30.0 cm in length. The force is applied at 90° to the radius. If the wheel nuts need a torque of 224 N to remove them, what is the minimum length of the adjustable spanner so that the nuts can be loosened? State whether the spanner is long enough.

Thinking	Working
Identify the variables involved and state them in their standard form.	$\tau = 224\,\text{N m}$ $r_{\perp} = ?$ $F = 845.0\,\text{N}$
Apply the equation for torque. Rearrange as necessary.	$\tau = r_{\perp}F$ $r_{\perp} = \frac{\tau}{F}$ $= \frac{224}{845}$ $= 0.265\,\text{m}$
State the answer with comparable units to the question.	$r_{\perp} = 26.5\,\text{cm}$
Compare the answer with the length of the spanner and state whether it is or isn't appropriate for this task.	As the spanner can be extended to 30.0 cm this is long enough to provide the minimum perpendicular distance of 26.5 cm.

Worked example: Try yourself 3.1.2

CALCULATING PERPENDICULAR DISTANCE

A truck driver can apply a maximum force of 1022 N on a large truck wheel-nut spanner that has a length of 80.0 cm. The force is applied at 90° to the radius. If the truck's wheel nuts need a torque of 635 N m to make them secure, determine if the length of this spanner is sufficient for the job.

Doors are also good examples of torque in action; the hinges form the axis of rotation and the force is applied to the door at a perpendicular distance from the axis of rotation (Figure 3.1.9).

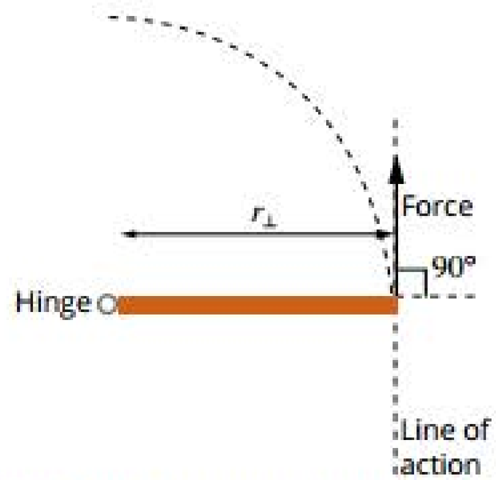


FIGURE 3.1.9 A door can have a torque applied to it, as long as the line of action of the force is not through the axis of rotation.

NON-PERPENDICULAR CALCULATIONS OF TORQUE

When the force causing a torque acts along a line that is at an angle other than 90° to an object (such as a door), then the torque is reduced (Figure 3.1.10). In these circumstances, you can calculate the torque by two approaches: either by finding the component of the force acting perpendicular to the door, or by finding the perpendicular distance from the pivot point to the line of action of the force.

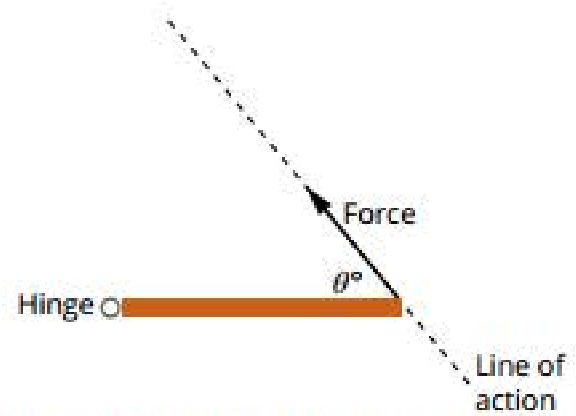


FIGURE 3.1.10 When the force causing a torque is not perpendicular to the door, the torque is reduced.

Recall that the formula for torque (τ) on an object is:

$$\tau = r_{\perp}F$$

This equation calculates the torque (τ) when the force (F) and the distance from the pivot to the line of action of the force (r) are perpendicular to each other. It really doesn't matter whether the radius is perpendicular to the line of action of the force, or if the force is perpendicular to the radius. You will explore both methods, and see that the methods are equivalent.

Calculating torque using perpendicular force

The component of any force can be calculated using trigonometry. To find the component of the force that is perpendicular to a door, for example, you can use the magnitude of the force and the angle between the door and the line of action of the force (Figure 3.1.11).

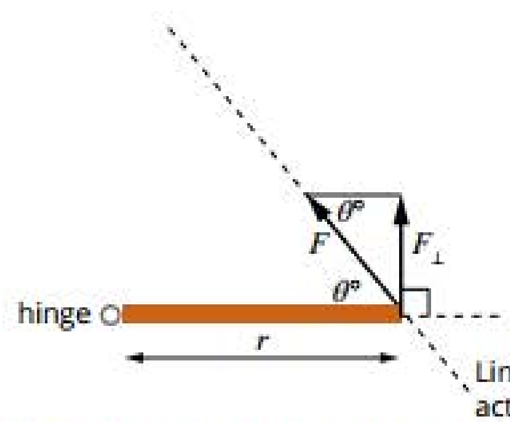


FIGURE 3.1.11 Determining the components of a force.

In this case $\tau = rF_{\perp}$ and $F_{\perp} = F\sin\theta$

Combining the two equations:

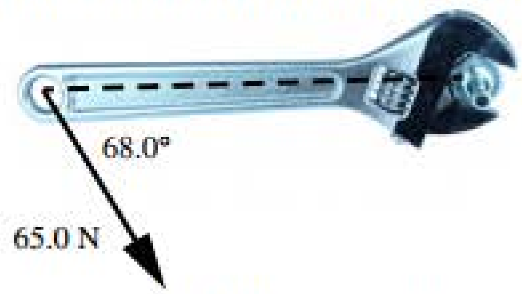
$\tau = rF\sin\theta$

The strategy for solving questions of this type may be either to calculate the perpendicular force and then apply the torque equation, or to use the combined equation. It is recommended that you begin by calculating the perpendicular component of the force, and then use the torque equation. When you have gained confidence with that strategy, try using the combined equation.

Worked example 3.1.3

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF FORCE

A student uses a 42.0 cm long adjustable spanner to loosen a nut on her bike. She applies a force of 65.0 N at an angle of 68.0° to the spanner.



Calculate the anticlockwise torque that the student applies to the nut.

Thinking	Working
Use the trigonometric relationship $F_{\perp} = F\sin\theta$ to determine the force perpendicular to the spanner.	$F_{\perp} = F\sin\theta$ $= 65.0\sin 68.0°$ $= 60.3$ N
Convert variables to their standard units.	$r = 42.0$ cm $= 0.420$ m
Apply the equation for torque: $\tau = rF_{\perp}$	$\tau = rF_{\perp}$ $= 0.420 \times 60.3$ $= 25.3$ N m
State the answer with the appropriate units including the direction, since torque is a vector.	$\tau = 25.3$ N m anticlockwise

Worked example: Try yourself 3.1.3

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF FORCE

A mechanic uses a 17.0 cm long spanner to tighten a nut on a winch. He applies a force of 104 N at an angle of 75.0° to the spanner.

Calculate the clockwise torque that the mechanic applies to the nut. Give your answer to three significant figures.

Calculating torque using perpendicular radius

The perpendicular component of any distance can be calculated using Pythagoras' theory or trigonometry. To find the component of a length that is perpendicular to the line of action of the force acting on a door, construct a line from the pivot point to the line of action of the force so that it intersects the line of action at right angles (Figure 3.1.12).

In this case

$\tau = r_{\perp}F$ and $r_{\perp} = r\sin\theta$

combine to give

$\tau = r\sin\theta F$

The equation $\tau = rF\sin\theta$ is identical to $\tau = r\sin\theta F$, so either method would be appropriate for calculating the torque on an object when the force is not at right angles to the object. In either method, the component of the distance or the component of the force is always going to be less than the distance or the force itself. This decrease in value will result in a smaller torque being applied to the object. The maximum torque will always be when the line of action of the force is perpendicular to the maximum distance from the pivot point to the line of action of the force.

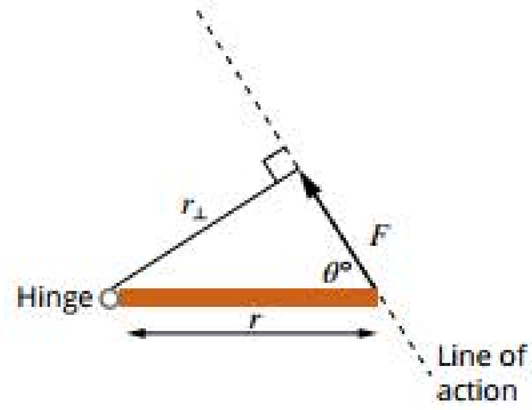
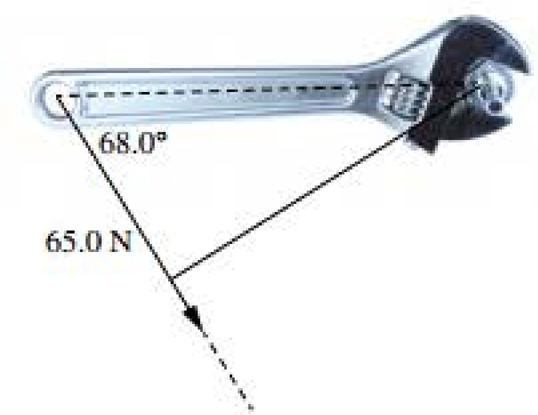


FIGURE 3.1.12 Determining the components of a distance.

Worked example 3.1.4

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF DISTANCE

A student uses a 42.0 cm long adjustable spanner to loosen a nut on her bike. She applies a force of 65.0 N at an angle of 68.0° to the spanner.



Using the perpendicular distance, calculate the anticlockwise torque that the student applies to the nut.

Thinking	Working
Convert variables to their standard units.	$r = 42.0\,\text{cm}$ $= 0.420\,\text{m}$
Use the trigonometric relationship $r_{\perp} = r\sin\theta$ to determine the perpendicular distance from the pivot point to the line of action of the force.	$r_{\perp} = r\sin\theta$ $= 0.420 \times \sin 68°$ $= 0.389\,\text{m}$
Apply the equation for torque $\tau = r_{\perp}F$.	$\tau = r_{\perp}F$ $= 0.389 \times 65.0$ $= 25.3$
State the answer with the appropriate unit and direction. Note that this is the same as the answer as for Worked example 3.1.3.	$\tau = 25.3\,\text{N m}$ anticlockwise

Worked example: Try yourself 3.1.4

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF DISTANCE

A mechanic uses a 17.0 cm long spanner to tighten a nut on a winch. He applies a force of 104 N at an angle of 75.0° to the spanner.

Using the perpendicular distance, calculate the clockwise torque that the mechanic applies to the nut. Give your answer to three significant figures.

PHYSICS IN ACTION

The torque wrench

The extent to which a nut or bolt is tightened can be critical to the safe operation of machinery or motors, a good example being the wheel nuts on a car. If a nut or bolt is too loose then it could fall out, but if it is too tight then it could either distort the part or the bolt could break off. Both of these situations could result in expensive repairs. To avoid nuts and bolts being too loose or too tight, manufacturers use different tools and methods to estimate the amount of torque required to tighten a nut or bolt to the correct tightness.

Figure 3.1.13 illustrates different types of torque wrench. The beam wrench is the simplest type of torque wrench. It has a flexible lever arm with a bar and scale, separating the wrench head and handle. When a torque is applied, a pointer on the scale moves to indicate the amount of torque being applied in newton metres (N m).

The click-type torque wrench can be set to apply a fixed amount of torque. When the required amount of torque has been achieved, the wrench 'clicks' and releases itself, preventing any further tightening from being applied.

More recently, electronic torque wrenches have been developed. A force beam within the wrench converts the compressive or tensile force applied to an electrical signal that is calibrated to show the force in newton metres on a digital readout. Measurements can also be stored within the instrument's memory, and further transferred to a computer. The force sensors you will most likely use in physics work on this same principle.

FIGURE 3.1.13 Three types of wrenches commonly used to measure the torque applied to a nut or bolt: (a) beam torque wrench, (b) click-type torque wrench, (c) digital torque wrench.

3.1 Review

SUMMARY

- Torque is a measurement of the amount of force applied to an object to make it rotate around an axis.
- The formula for calculating the magnitude of the torque is $\tau = r_{\perp}F$.
- Torque occurs when the acting force is not applied directly through the pivot point of the object.
- Maximum torque occurs when the acting force is applied at a perpendicular angle of 90° to the force arm.
- The larger the force acting on the object, the larger the torque will be.
- The longer the force arm, the greater the torque will be.
- If torque is generated by a force that is not perpendicular to the lever arm of the object, then either:
  - the component force perpendicular to the length of the object is used to calculate torque $\tau = rF_{\perp}$

  or

  - the distance from the pivot point perpendicular to the line of action of the force is used to calculate torque $\tau = r_{\perp}F$.
- Both strategies for determining torque from non-perpendicular situations equate to $\tau = rF\sin\theta$.

KEY QUESTIONS

1 Select the correct option below to complete the following statement about torque.
When a force acts such that the line of action of the force is not directed through the pivot point of the object then:
A a torque will result
B no torque will result

2 Which of the options listed below best explains the factors affecting the torque acting on an object?
A Torque is only increased by increasing the force applied to the object.
B Torque increases when either the force or the force arm increases, or both increase.
C Torque only increases when the force arm decreases.
D Torque only increases when the force arm increases.

3 Calculate the torque applied to a bolt by a 20.0 cm long spanner that has a force of 25.0 N acting at 90° to its length and at the end of the spanner.

4 Calculate the radius of the wheel on a pressure valve that supplies a torque on the valve of magnitude 3.47 N m when a force of 12.0 N is applied.

5 When opening a 1 m wide door, the torque of the greatest magnitude is provided by:
A pushing with a force of 33 N at right angles to the door, in the middle of the door
B pushing with a force of 25 N at right angles to the door, 25 cm from the handle edge of the door
C pushing with a force of 50 N at right angles 25 cm from the hinges of the door
D pulling with a force of 25 N at right angles to the door, at the handle edge of the door

6 Demolishers wish to knock over a concrete wall. They plan to use a wrecking-ball that exerts 5000 N as it hits the wall. If it hits at a point that is 3.0 m above the ground, calculate the magnitude of the torque that is developed on the wall if the wall pivots at its base.

7 Calculate the length of a spanner that is used to tighten a nut to a torque of magnitude 32.1 N m when a force of 24.0 N is applied at right angles to the spanner, at the end of the spanner.

8 A camper ties a rope from the top of a 2.00 m tent pole to a peg on the ground. The rope is tightened so that the rope applies a 30.0 N force at an angle of 40.0° to the pole. Calculate the magnitude of the torque that is developed on the tent pole due to the rope if it pivots at its base.

3.1 Review *continued*

9 Penny is assembling some flat-pack furniture with an Allen key. Calculate the torque supplied on a screw by a 7.00 cm long Allen key that has a force of 8.50 N acting at 90.0° to its length, and at the end of the Allen key.

10 A mechanic uses a trolley jack to jack up a car by pushing vertically down on the end of a 90.0 cm lever with a force of 82.0 N, as shown in the diagram below. The lever is shown at an angle of 40.0° up from horizontal. Calculate the magnitude of the torque acting on the pivot.

The following information applies to questions 11 and 12.

A rope is attached at 30.0° to a freshly planted tree. The line of action of the force is along the same line as the rope, and the rope is attached 1.50 m from the base of the tree. Assume the base of the tree is the pivot point.

11 Calculate the length of the perpendicular force arm of the rope.

12 Calculate the torque on the tree if the force applied to the tree by the rope is 12.5 N and it is tensioned clockwise about the base of the tree.

3.2 Equilibrium of forces

Newton's first law states that an object will continue with its motion unless acted upon by an external unbalanced force. The situation in which the object continues with its motion, or doesn't change its velocity, occurs when the forces on the object are balanced. In situations where the forces are balanced, the forces are said to be in translational equilibrium, such as during a tug-o-war (Figure 3.2.1). When a tug-o-war starts there is an equilibrium of forces as both teams take the strain. Winning a tug-o-war match involves one team applying a greater force so that there is a net force on the rope, causing the rope and teams to accelerate in the winning team's direction. Once the rope and the teams are moving at a constant velocity then an equilibrium of forces exists once again.

FIGURE 3.2.1 The game tug-o-war is an example of translational equilibrium.

CENTRE OF MASS AND STABILITY

Think about an athlete running in a 100 m sprint. In simple terms, the athlete runs in a straight line along the track, and the displacement and velocity at any time can be calculated using the principles discussed in the 'Linear motion' chapter in Year 11. In reality, however, the motion of the various parts of the athlete's body will differ significantly during the run. The athlete's arms and legs move in a complex manner that requires a significantly more complex analysis.

The analysis of the motion of complicated systems, such as a sprinter or high-jumper, can be simplified to the motion of a single point. The mass of the sprinter can be considered to be 'concentrated' into a single point, which has travelled in a straight line. This single point is called the **centre of mass**. There is as much mass above the centre of mass as there is below it, as much mass to the left as there is to the right, and as much mass in front as there is behind it.

If an object is uniform in one dimension only (e.g. a straight piece of wire), its centre of mass will be exactly half way along its length. The centre of mass of an object in two dimensions (e.g. a flat piece of card), will be the point that is half way along and half way across the object (Figure 3.2.2). It is even possible for the centre of mass to lie outside the body; for example, the centre of mass of a doughnut is located at the centre of the hole. A person's centre of mass is typically in the lower part of the back, but it will vary with the positions of the arms and legs.

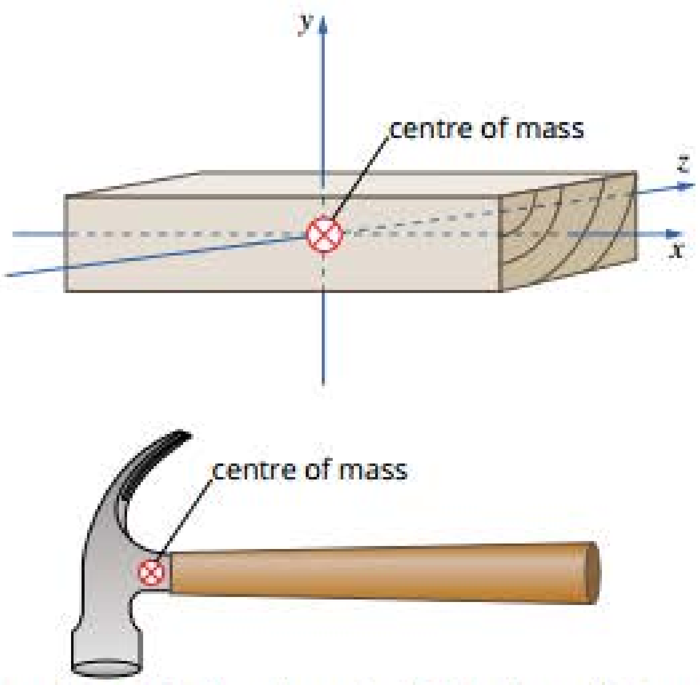


FIGURE 3.2.2 The centre of mass of a three-dimensional block of wood is shown with arrows drawn in the *x*, *y* and *z* dimensions illustrating the lines where there is equal mass on either side of the lines. The centre of mass of a non-uniform hammer shows that the centre of mass will be closer to the end with a higher density than the end with the lower density.

A concept that is closely related to centre of mass is **centre of gravity**. Instead of being a point particle whose motion equates to the whole extended body or system, the centre of gravity is the position from which the entire weight of the body or system is considered to act. As a consequence of this, the centre of gravity is the position at which the body will balance. For all practical purposes, the centre of gravity is exactly at the centre of mass. It is only when a body is so large that it is in a non-uniform gravitational field that the centre of gravity no longer coincides with the centre of mass.

There are three types of equilibrium related to the stability of an object:

- **stable equilibrium**—The object will return to its equilibrium position even when a force is applied to it, as long as the centre of mass is not moved outside the **base** of support.
- **unstable equilibrium**—The object will accelerate and will not return to its equilibrium position when a force is applied to it, since the centre of mass is readily moved outside the base of support.
- **neutral equilibrium**—The object will remain stationary no matter where it is placed, as any force applied has no effect on the relationship between the centre of mass and the base or point of support.

Figure 3.2.3 illustrates how this applies to real life objects.

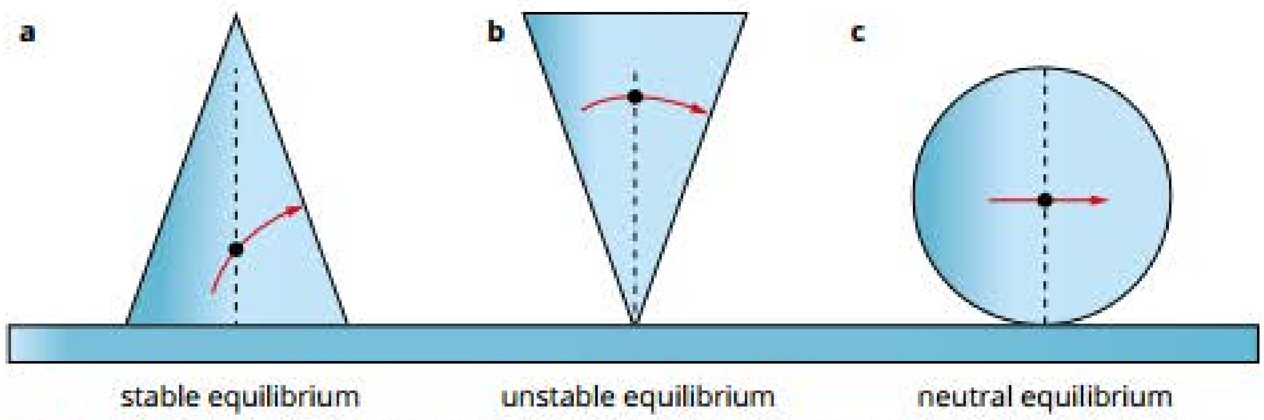
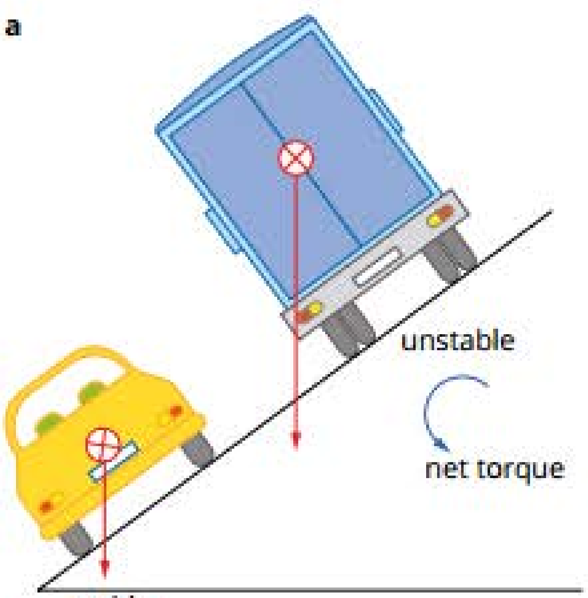


FIGURE 3.2.3 Three objects illustrating the difference between (a) stable, (b) unstable and (c) neutral equilibrium. As (a) has a broad base of support it will return to its original position even when a force is applied to it, as long as the centre of mass isn't moved outside the base of support.

a



b

FIGURE 3.2.4 (a) The car on the incline is in stable equilibrium, while the heavily laden truck on the same incline could topple. The weight vector is outside the lower point of support for the truck, so there is no reaction force from the road to the higher wheel. (b) Modern four-wheel drives and tractors have inclinometers to warn the driver if the vehicle is in danger of tipping.

When designing structures, engineers and architects want to ensure that balance and stability are maintained. Whether this occurs depends on the relative positions of the centre of gravity and the base. When a vertical line downwards from the centre of gravity passes through the base of support, the object is stable. The vertical line from the centre of gravity represents the direction of the force of gravity on the object.

In Figure 3.2.4(a), the weight of the car passes through the car's support base, between the wheels. The torque acting on the car is clockwise in this case and therefore does not cause the car to tip over. In the case of the truck, however, the weight is directed outside the truck's base of support, so an anticlockwise torque acts to tip the truck over.

The stability of an object or structure can be increased in a number of ways:

- the centre of gravity is lowered
- the width of the support base is increased
- the angle from the centre of gravity to the edge of the base is increased.

As a result of any of these, the object has to be tipped further to make the force of gravity act outside the support base. Racing cars have a very low centre of gravity to increase their stability when cornering at high speed. In a similar way, training wheels on a child's bicycle widen the support base, making it harder to tip the bicycle sideways.

These conditions also apply to objects that are supported from above. A child on a playground swing will be in stable equilibrium when the swing is hanging straight down. A gentle push will cause the swing to move but it will return to its original position once the outside force is removed.

If, on the other hand, the swing were raised to a significant height and held there, then it would be in unstable equilibrium. The swing is in equilibrium only because the forces on it are balanced while it's being held. If the holding force (say a parent's hands) were removed, then the swing will start to move, swinging across the point of stable equilibrium until eventually coming to rest in a position of stable equilibrium.

TRANSLATIONAL EQUILIBRIUM

A **translational equilibrium** of forces occurs when the sum of the forces acting through the centre of mass of an object add to give a zero resultant or zero net translational force. As a net translational force causes acceleration in one direction, then a zero net translational force causes no acceleration of the object. This condition is the defining aspect of a translational equilibrium of forces.

When analysing a situation involving more than one force acting on an object, translational equilibrium will exist if the sum of the forces is equal to zero: $\Sigma F = 0$. This can also be written as $F_{net} = 0$.

Vector diagrams of an equilibrium of forces in one dimension

Vector diagrams can be drawn of the forces acting on an object, when the forces are acting in one dimension. For example, if three people are pulling to the right and three people are pulling to the left in a game of tug-o-war, as shown in Figure 3.2.5, then the forces are all in one dimension—left and right. You can add these forces using a vector diagram by drawing all the forces from each person head to tail as described in Chapter 6 'Scalars and vectors' in *Pearson Physics 11 Western Australia*. If the tug-o-war is in translational equilibrium, then the all the forces should add to give a zero net force.

FIGURE 3.2.5 Tug-o-war with a vector addition diagram showing a net force of zero, indicating an equilibrium of forces exists.

Calculating an equilibrium of forces in one dimension

Whether a situation is in translational equilibrium or not can be determined using a consistent sign convention to represent the direction of the force vectors in one dimension. Typically, in the x-dimension, left is treated as negative and right as positive; similarly in the y-dimension, up is treated as positive and down as negative. In the z-dimension, forwards is positive and backwards is negative. By applying a consistent sign convention to the forces acting on an object, the addition of those forces, with their signs, will give a vector sum of zero if the situation is in translational equilibrium.

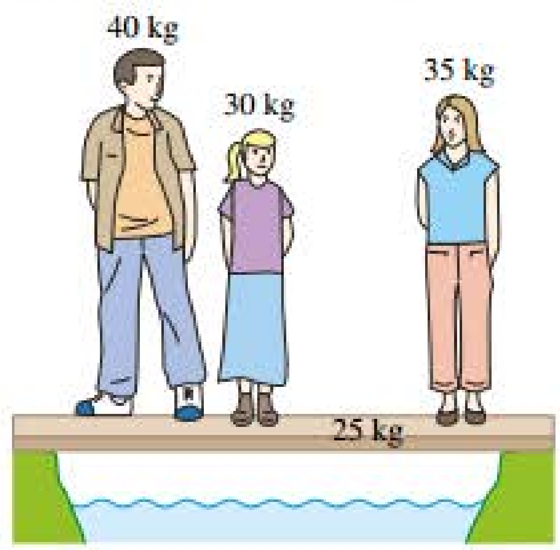
That is:

$\Sigma F_{\text{left–right}} = 0$
$\Sigma F_{\text{up–down}} = 0$
$\Sigma F_{\text{forwards–backwards}} = 0$

Worked example 3.2.1

CALCULATING TRANSLATIONAL EQUILIBRIUM IN ONE DIMENSION

Three students are standing on a plank that is bridging a small stream. The plank is supported at each end by the ground. The plank has a mass of 25.0 kg and the students have masses of 40.0 kg, 30.0 kg and 35.0 kg. There is an upwards force of the left bank on the plank of 700 N. If the plank is in translational equilibrium, then calculate the force of the right bank on the plank. Use $g = 9.80\,\text{N kg}^{-1}$ when answering this question.

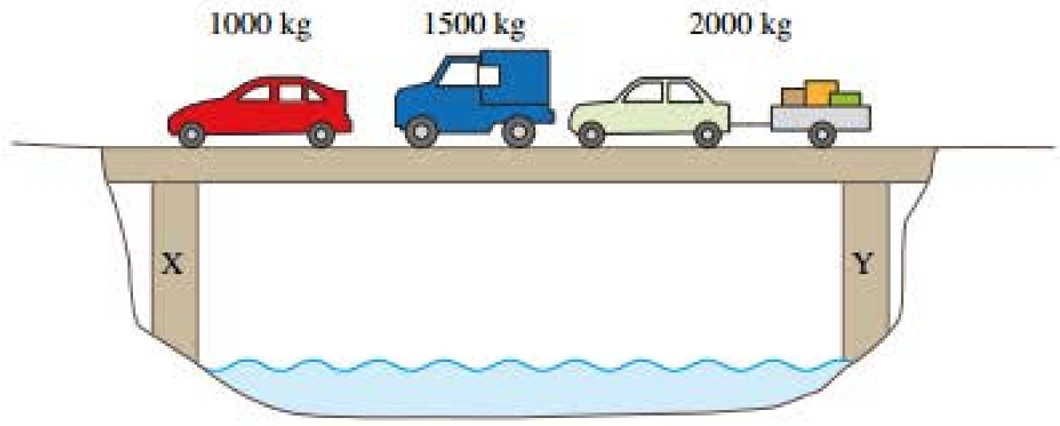


Thinking	Working
Identify the variables involved and state them with their directions in their standard form.	$m_1 = 40.0\,\text{kg}$ $m_2 = 30.0\,\text{kg}$ $m_3 = 35.0\,\text{kg}$ $m_P = 25.0\,\text{kg}$ $F_{LB} = 700\,\text{N}$ up $g = 9.80\,\text{N kg}^{-1}$ down
Apply a sign convention to the vector data.	$F_{LB} = +700\,\text{N}$ $g = -9.80\,\text{N kg}^{-1}$
The object experiencing translational equilibrium is the plank.	$\Sigma F_{\text{up–down}} = 0$
Expand the equation to include each of the forces acting on the plank.	$F_1 + F_2 + F_3 + F_{WP} + F_{LB} + F_{RB} = 0$ $m_1g + m_2g + m_3g + m_Pg + F_{LB} + F_{RB} = 0$
Substitute the data into the equation and solve for the unknown.	$m_1g + m_2g + m_3g + m_Pg + F_{LB} + F_{RB} = 0$ $(40.0 \times -9.80) + (30.0 \times -9.80) + (35.0 \times -9.80) + (25.0 \times -9.80) + 700 + F_{RB} = 0$ $-392 + -294 + -343 + -245 + 700 + F_{RB} = 0$ $-574 + F_{RB} = 0$ $F_{RB} = 574\,\text{N}$
State the answer with the appropriate direction.	$F_{RB} = 574\,\text{N}$ up

Worked example: Try yourself 3.2.1

CALCULATING TRANSLATIONAL EQUILIBRIUM IN ONE DIMENSION

Three cars are parked on a beam bridge that has a mass of 500 kg. Pillar X applies a force of 200×10^4 N upwards. If the situation is in translational equilibrium, then calculate the force provided by pillar Y. Use $g = 9.80\,\mathrm{N\,kg^{-1}}$ when answering this question. Car 1 (C_1) is red; car 2 (C_2) is blue; car 3 (C_3) is light green.



SOLVING EQUILIBRIUM IN TWO DIMENSIONS

If the forces involved in the equilibrium situation are in two dimensions, there are two strategies for determining if the sum of the forces is zero: using vector diagrams or vector components.

Vector diagrams

You can draw **vector diagrams** in which vectors are added in two dimensions. This was covered in Chapter 6 'Scalars and vectors' *in Pearson Physics 11 Western Australia*. In any vector addition diagram, the individual vectors are added head to tail, with the net vector drawn from the tail of the first vector to the head of the last vector. In a situation in which the forces are in equilibrium, the vector addition diagram should end up with the head of the last vector finishing at the tail of the first vector. This means that the vector addition diagram ends up in a closed loop and therefore there is no net force (Figure 3.2.6).

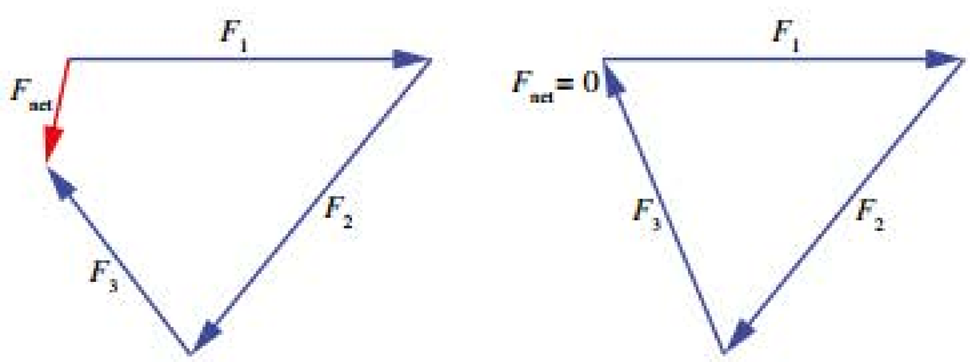


FIGURE 3.2.6 The vector diagram on the left shows three vectors added head to tail, and the resultant net force vector F_{net} in red. The closed loop vector addition diagram on the right shows an equilibrium of forces, where $F_{net} = 0$.

Vector components

A vector diagram that results in a closed loop fulfils the conditions of a translational equilibrium of forces. If the three forces form a right triangle, you can use Pythagoras' theorem and trigonometry to determine the magnitude and direction of a third force that will be in equilibrium with two other forces. If the three forces are in any other triangle, then the unknown force can be found using the sine law or the cosine law, which is a more complex method. However, there is an easier method that involves determining the **components** of the forces that are in two perpendicular dimensions.

These force components are then added in each of their dimensions, which results in two perpendicular resultant vectors that can be added using Pythagoras' theorem and trigonometry (Figure 3.2.7).

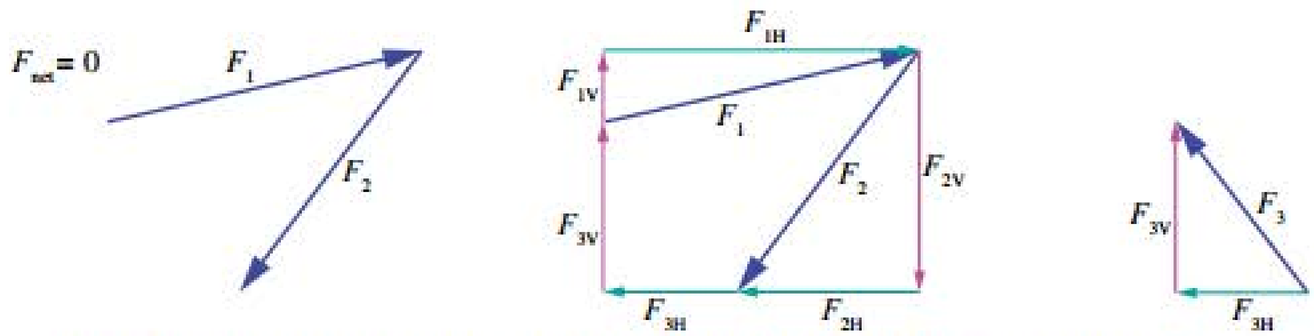


FIGURE 3.2.7 A method for finding F_3 that acts to make an equilibrium of forces with F_1 and F_2.

The guiding principle behind this method is that for the forces to be in equilibrium the sum of the *x*-axis (horizontal or left–right) forces must equal zero, and the sum of the *y*-axis (vertical or up–down) forces must also equal zero.

$\Sigma F_{x\text{-axis}} = 0$ or $\Sigma F_{\text{vertical}} = 0$ or $\Sigma F_{\text{left–right}} = 0$

$\Sigma F_{y\text{-axis}} = 0$ $\quad\Sigma F_{\text{horizontal}} = 0$ $\quad\Sigma F_{\text{up–down}} = 0$

To draw a vector diagram (Figure 3.2.8):

- identify all of the forces that apply (e.g. tension, weight force)
- draw the vectors for each force separately and mark all known angles
- reposition the vectors so they are head to tail, adding all of the forces together.

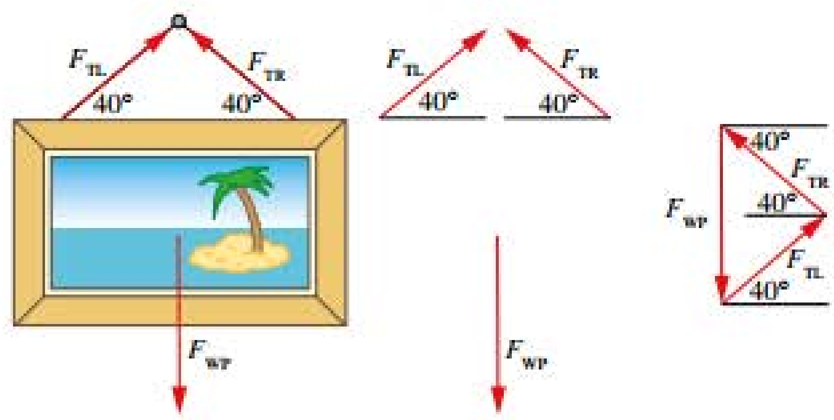
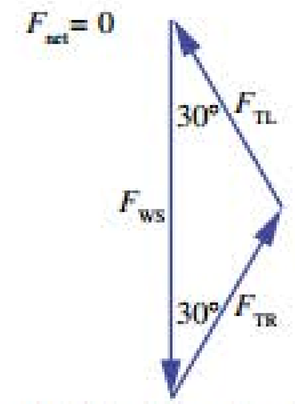
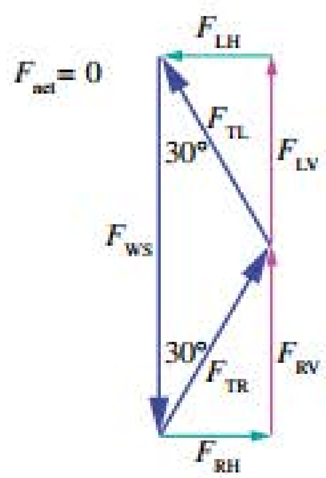
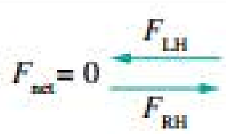
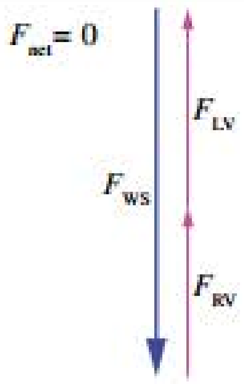


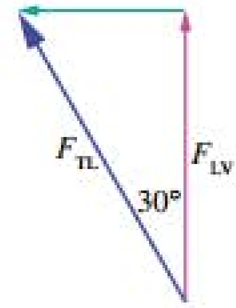
FIGURE 3.2.8 A vector diagram is drawn by identifying the vectors and then adding the vectors together, head to tail.

Worked example 3.2.2

CALCULATING TRANSLATIONAL EQUILIBRIUM IN TWO DIMENSIONS

An advertising sign is hung by two cables to the ceiling of a shop. The sign has a mass of 45.0 kg and the cables are at an angle of 30° to the vertical as shown in the image below. If the mass of the cables is ignored, calculate the tension in each cable when the sign is suspended. Use $g = 9.80\,\text{N kg}^{-1}$ when answering this question.

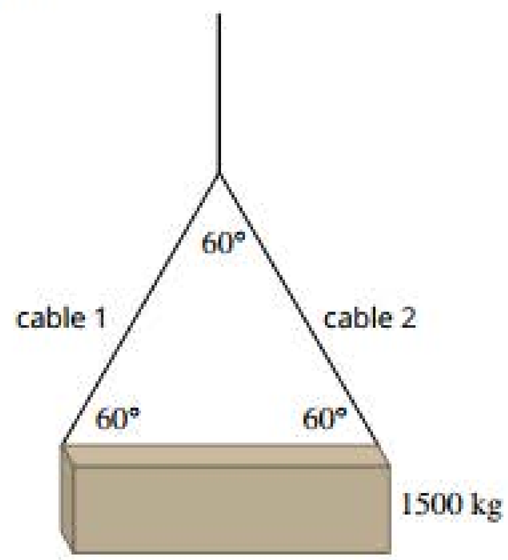
Thinking	**Working**
Construct a vector diagram adding all of the forces together.	 F_{WS} is the weight of the sign, F_{TL} is the tension of the left cable and F_{TR} is the tension of the right cable.
Apply left and right components and up and down components.	
In the horizontal dimension, F_{LH} is in equilibrium with F_{RH}.	
In the vertical dimension, F_{WS} is in equilibrium with F_{LV} and F_{RV}.	
Apply the equation for translational equilibrium in one dimension. F_{RV} and F_{LV} are equal in magnitude and therefore each is half of F_{WS}.	$\Sigma F_{up\text{–}down} = 0$ $F_{RV} = F_{LV}$
Expand the equation to include each of the vertical forces acting on the sign.	$F_{WS} + F_{RV} + F_{LV} = 0$ $m_S g + F_{RV} + F_{RV} = 0$
Substitute the data into the equation and solve for the unknown.	$(45.0 \times -9.80) + 2F_{RV} = 0$ $-441 + 2F_{RV} = 0$ $F_{RV} = \frac{441}{2}$ $F_{RV} = F_{LV} = 220.5\,\text{N}$

Draw the right triangle with one of the vertical components of the tension and the angle.	
Use trigonometry to solve for the tension in one of the cables, which will equal the tension in the other cable as well.	$\cos\theta = \frac{F_{LV}}{F_{TL}}$ $F_{TL} = \frac{F_{LV}}{\cos\theta}$ $= \frac{220.5}{\cos 30°}$ $= 255\,N$ $F_{TL} = F_{TR} = 255\,N$

Worked example: Try yourself 3.2.2

CALCULATING TRANSLATIONAL EQUILIBRIUM IN TWO DIMENSIONS

A concrete beam of mass 1500 kg is being lifted by cables labelled 1 and 2, as shown in the diagram. The beam is moving upwards with a constant velocity of $2.0\,m\,s^{-1}$. Calculate the tension in cable 1 and cable 2. Ignore the mass of the cables and use $g = 9.80\,N\,kg^{-1}$ when answering this question. Give your answers to three significant figures.



3.2 Review

SUMMARY

- Translational equilibrium in one dimension can be represented mathematically as $\Sigma F = 0$.
- Translational equilibrium in two dimensions can be represented mathematically as $\Sigma F_x = 0$ and $\Sigma F_y = 0$.
- In two dimensions, an equilibrium of forces can be represented in a closed vector diagram.
- By calculating x and y components of the forces in equilibrium, the forces in equilibrium in each dimension can be found.

KEY QUESTIONS

1 Select the correct statement explaining when translational equilibrium occurs.

A A net force acts through the centre of mass and causes no acceleration.

B A net force acts through the centre of mass and causes an acceleration.

C No net force acts through the centre of mass resulting in an acceleration.

D No net force acts through the centre of mass and so there is no acceleration.

2 An object is in translational equilibrium. Which of the following statements applies?

A The object must be stationary.

B The object must be moving at constant velocity.

C The object must be experiencing translational acceleration.

D The object must not be experiencing translational acceleration.

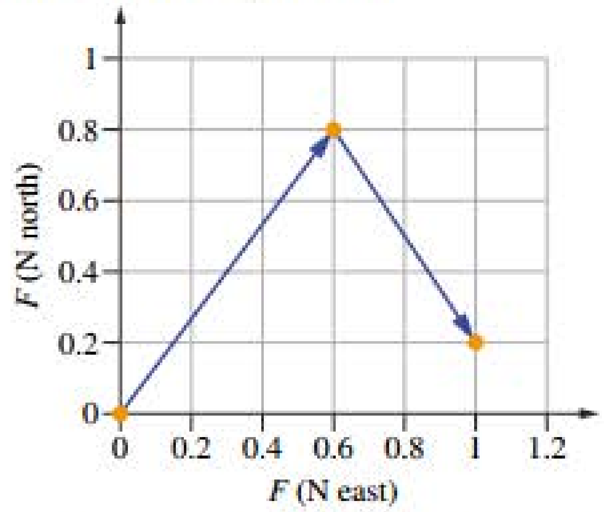
3 A string supports a bird-feeder that has a mass of 355 g. If the bird-feeder is in translational equilibrium, calculate the tension in the string supporting the bird-feeder.

4 Calculate the mass of a pendulum bob that hangs stationary from a chain when the tension in the chain is 7.50 N.

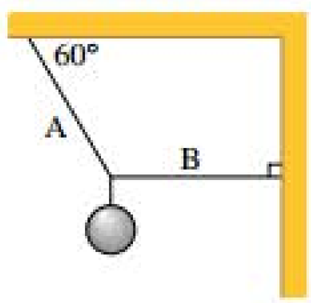
5 A chef wants to make sure that the wire cables attached to each end of a kitchen rail will be able to hold the weight of the rail, two tools and a frying pan that must hang from it. The mass of the rail is 3.05 kg, the mass of each tool is 350 g and the mass of the frying pan is 2.25 kg. Calculate the tension force in each wire cable if the load is spread evenly along the rail. State your answer to three significant figures.

6 Two window cleaners working on the windows of a high-rise building work on a platform suspended by four cables. Tom has a mass of 79.0 kg, Jack has a mass of 68.0 kg and the mass of the platform is 225 kg. The platform is moving down the side of the building at a constant speed and all cables provide the same tension. Calculate the tension in one of the cables.

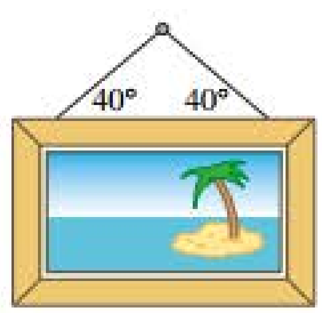
7 A shopping trolley is pushed in different directions by three children. The force vectors of two of the children are shown in the diagram. Draw the force vector from the third child that would cause the shopping trolley to be in translational equilibrium.



8 A bowling alley wants to promote its business by suspending a 100 kg fibreglass bowling ball on a frame outside its alley. The bowling ball is supported by two steel cables capable of withstanding up to 1500 N of tension before breaking. Cable A is at an angle of 60° to the horizontal frame member and cable B is perpendicular to the vertical frame member. Assuming the mass of the cables is negligible, calculate the tension in cable A and in cable B. Give your answer to three significant figures.



9 A picture is hung on a wall as shown in the diagram below. If the hanging wire is capable of supporting a maximum force of 40.0 N, what is the maximum mass of the picture that can be supported before the wire snaps?



10 A street performer stands on a rope tied between two posts. When the performer is standing at the centre, the rope makes an angle of 10.0° to the horizontal. Assuming the mass of the rope is negligible and that the mass of the performer is 75.0 kg, calculate the tension in the rope. Give your answer to three significant figures.

3.3 Static equilibrium

FIGURE 3.3.1 When a cyclist doesn't need to pedal, they can stand on both pedals with equal force. This causes an equal torque in the clockwise and anticlockwise directions. The pedals are in rotational equilibrium.

Objects are in translational equilibrium when the sum of the forces acting through their centre of mass equals zero. However, not all forces act through the centre of mass of an object; some forces may act in ways that generate torque or moment on an object. In order for an object to be in rotational equilibrium, the sum of the moments in a clockwise direction must balance the sum of the moments in an anticlockwise direction (Figure 3.3.1). This relationship is called the **principle of moments**.

By combining the conditions for translational equilibrium and rotational equilibrium, an object can be made to be in a state of static equilibrium. Static equilibrium will be explored in more detail in this section.

ROTATIONAL EQUILIBRIUM

A **rotational equilibrium** of torque or moments occurs about a **reference point** when the sum of all the torques acting on an object in the clockwise direction is equal to the sum of all the torques acting in the anticlockwise direction.

When analysing a situation involving more than one torque acting on an object, the principle of moments can be said to apply when rotational equilibrium exists, or the sum of the torques about a reference point is equal to zero.

That is:

$\Sigma\tau = 0$

This is also represented as:

$\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$

Figure 3.3.2 illustrates the masts of a boat that are in rotational equilibrium around their base of support in the boat.

FIGURE 3.3.2 To keep each mast in rotational equilibrium in relation to its base, the stainless steel cables (stays) attached to it must provide opposing torques on the mast.

When the sum of the torques is not balanced, the object will experience an unbalanced torque and will rotate about the reference point.

CONDITIONS FOR STATIC EQUILIBRIUM

When a body or a system is not accelerating or rotating, it must be in both translational and rotational equilibrium. This situation then satisfies the conditions for **static equilibrium**. This can be represented by:

$\Sigma F = 0$ and $\Sigma\tau = 0$

This can also be shown as:

$\Sigma F_x = 0$

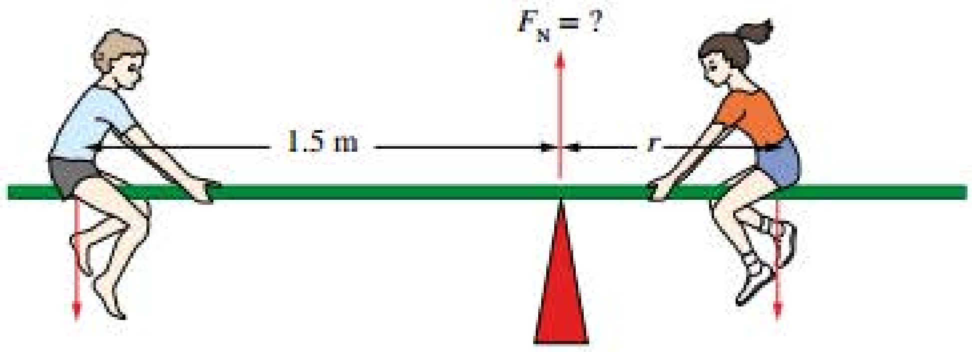
$\Sigma F_y = 0$

$\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$

Worked example 3.3.1

CALCULATING STATIC EQUILIBRIUM

Two young children make a seesaw with a long plank. The boy sits on the seesaw 1.50 m from the pivot. The masses of the boy and the girl are 20.0 kg and 30.0 kg respectively. Assume that the mass of the plank is negligible. Use $g = 9.80\ \text{N kg}^{-1}$ in your calculations where required.



a Calculate the force applied to the plank by the pivot point when the children are sitting on the seesaw.	
Thinking	**Working**
Identify the variables involved and state them with their directions in their standard form.	$m_g = 30.0\ \text{kg}$ $m_b = 20.0\ \text{kg}$ $g = 9.80\ \text{N kg}^{-1}$ down
Apply a sign convention to the vector force data.	$g = -9.80\ \text{N kg}^{-1}$
Identify the object that is in translational equilibrium. This is the object on which all the forces are acting.	The object experiencing translational equilibrium is the plank.
Apply the equation for translational equilibrium in one dimension.	$\Sigma F_y = 0$
Expand the equation to include each of the forces acting on the plank.	$F_{Wg} + F_{Wb} + F_P = 0$ $m_g g + m_b g + F_P = 0$
Substitute the data into the equation and solve for the unknown	$(30.0 \times -9.80) + (20.0 \times -9.80) + F_P = 0$ $(-294) + (-196) + F_P = 0$ $(-490) + F_P = 0$ $F_P = 490\ \text{N}$
State the answer, giving the direction.	$F_P = 490\ \text{N}$ up

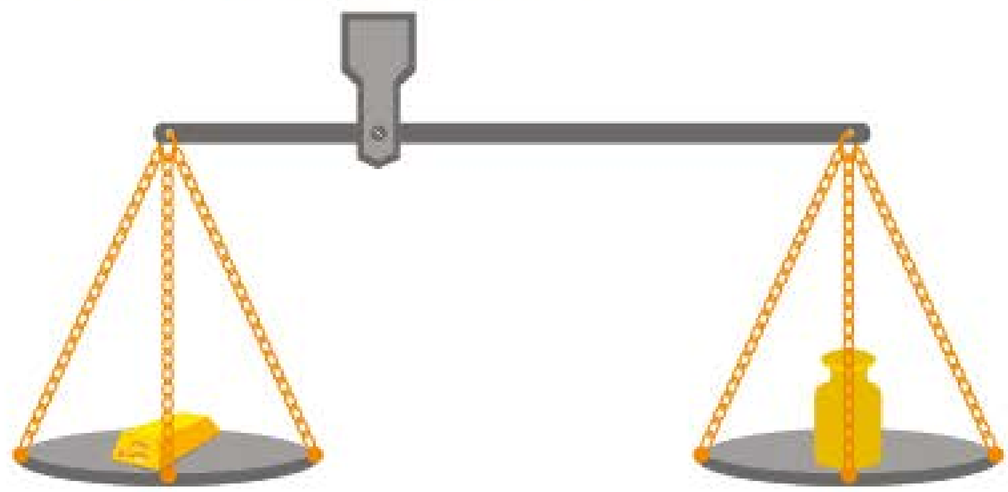
b Calculate where the girl has to sit in order to balance the boy.	
Thinking	**Working**
Identify the variables involved and state them in their standard form.	$m_g = 30.0\ \text{kg}$ $m_b = 20.0\ \text{kg}$ $r_{\perp b} = 1.5\ \text{m}$ $g = 9.80\ \text{N kg}^{-1}$
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the seesaw.
Decide the reference point about which the torques will be calculated.	The reference point is the pivot of the seesaw.
Decide which force causes the clockwise torque and which force causes the anticlockwise torque around the chosen reference point.	The force of the girl on the seesaw provides the clockwise torque. The force of the boy on the seesaw provides the anticlockwise torque.

Apply the equation for rotational equilibrium.	$\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$
Expand the equation to include each of the torques acting on the seesaw	$F_{Wg}r_{\perp g} = F_{Wb}r_{\perp b}$
Substitute the data into the equation and solve for the unknown.	$F_{W}r_{\perp g} = F_{Wb}r_{\perp b}$ $m_g g r_{\perp g} = m_b g r_{\perp b}$ $30.0 \times 9.80 r_{\perp g} = 20.0 \times 9.80 \times 1.50$ $r_{\perp g} = \frac{20.0 \times 9.80 \times 1.50}{30.0 \times 9.80}$ $= 1.00\,\text{m}$

Worked example: Try yourself 3.3.1

CALCULATING STATIC EQUILIBRIUM

A set of scales (with one longer arm) is used to measure the mass of gold, when it is larger than the standard set of masses. A lump of gold weighing 150 g is placed on the short arm, which is 10.0 cm long, and the standard masses are placed on the long arm. Use $g = 9.80\,\text{N}\,\text{kg}^{-1}$ in your calculations where required. Give your answers to three significant figures.



a Calculate the force applied to the scale's arm due to the pivot point if a standard mass of 50.0 g exactly balances the gold.

b Calculate the length the long arm should have in order to balance the gold.

In Worked example 3.3.1, the seesaw is in equilibrium because all the forces and torques are balanced. In solving the problem, it seemed obvious to choose the pivot as the reference point, around which the torques are determined. But because the seesaw plank is in equilibrium, any point could have been chosen as the reference point. For example, you could take the reference point to be where the girl is sitting (Figure 3.3.3). This will mean that the torques acting on the plank would be due to the boy, and due to the seesaw pivot point. The torque due to the girl becomes zero as the lever arm distance for her will be zero. The solutions to the questions will still work out the same.

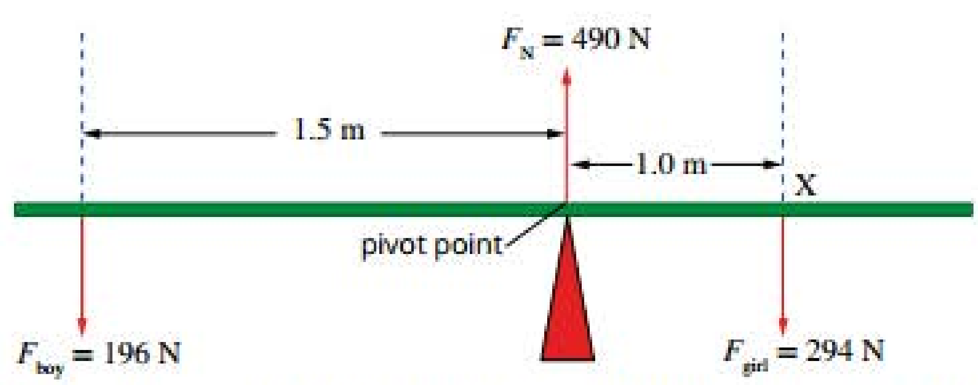


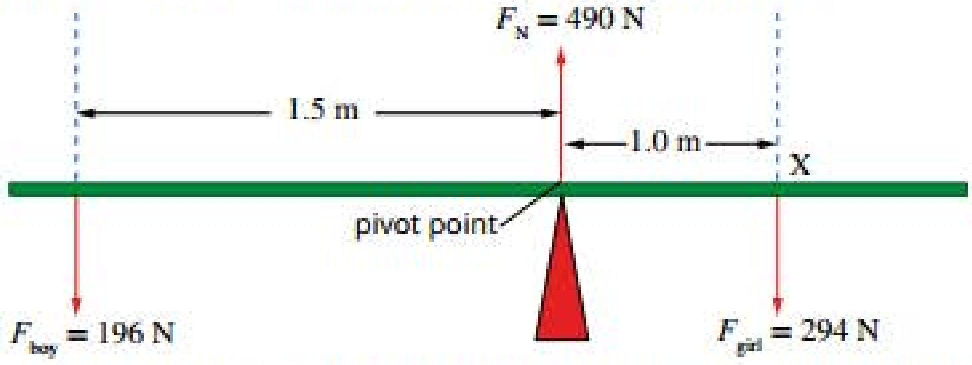
FIGURE 3.3.3 A force diagram for the seesaw problem from Worked example 3.3.1 in which the point at which the girl sits (labelled X) has been chosen as the reference point.

The boy will create an anticlockwise torque around the girl, and the normal force at the pivot for the seesaw creates an equal clockwise torque around the girl. In the following worked example it can be verified that the seesaw is in rotational equilibrium by calculating the torques around the position of the girl. The plank will be in rotational equilibrium if the clockwise torque equals the anticlockwise torque.

Worked example 3.3.2

CALCULATING STATIC EQUILIBRIUM USING A DIFFERENT REFERENCE POINT

Verify that the seesaw plank in the image below is in rotational equilibrium about the reference point X, where the girl is sitting. The weights of the boy and girl are 196 N and 294 N respectively, and the force of the pivot on the plank is 490 N upwards. Assume that the mass of the plank is negligible.



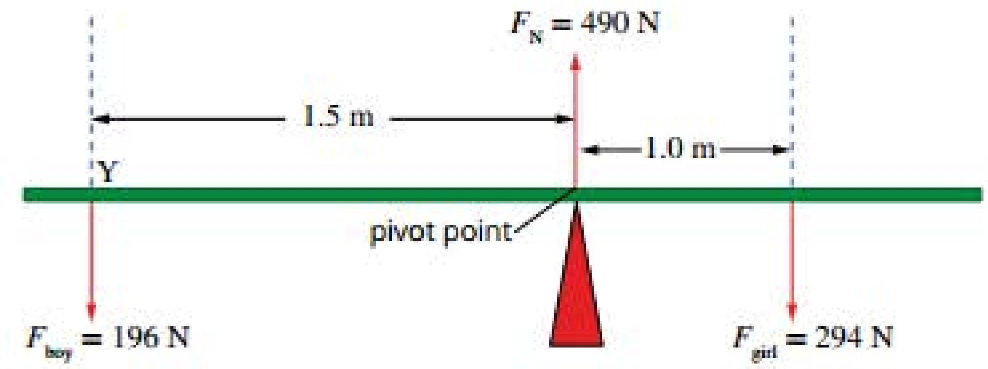
Here the point at which the girl is sitting (labelled X) has been chosen as the reference point. The force due to the boy acts downwards on the plank and the force due to the pivot point acts upwards on the plank.

Thinking	**Working**
Identify the variables involved and state them in their standard form.	$F_P = 490\,N$ $F_g = 294\,N$ $F_b = 196\,N$ $r_{\perp b} = 2.50\,m$ $r_{\perp P} = 1.00\,m$
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the seesaw.
Decide the reference point about which the torques will be calculated.	The reference point is the position of the girl at X.
Decide which force causes the clockwise torque, and which force causes the anticlockwise torque around the chosen reference point.	The force of the pivot on the plank provides the clockwise torque. The force of the boy on the plank provides the anticlockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$
Expand the equation to include each of the torques acting on the seesaw. Note that the girl's torque is not included here as, being the reference point, her torque is zero.	$F_P r_{\perp P} = F_b r_{\perp b}$
Substitute the data into the equation and solve.	$F_P r_{\perp P} = F_b r_{\perp b}$ $490 \times 1.00 = 196 \times 2.5$ $490 = 490$
Identify the magnitude of the clockwise torque compared to the magnitude of the anticlockwise torque.	Around reference point X (the position of the girl), the clockwise torque due to the pivot point on the plank is equal to the anticlockwise torque due to the boy on the plank. So the plank is in rotational equilibrium.

Worked example: Try yourself 3.3.2

CALCULATING STATIC EQUILIBRIUM USING A DIFFERENT REFERENCE POINT

Verify that the seesaw plank in the figure below is also in rotational equilibrium about the reference point Y, where the boy is sitting. The weights of the boy and girl are 196 N and 294 N respectively, and the force of the pivot on the plank is 490 N upwards. Assume that the plank's mass is negligible.



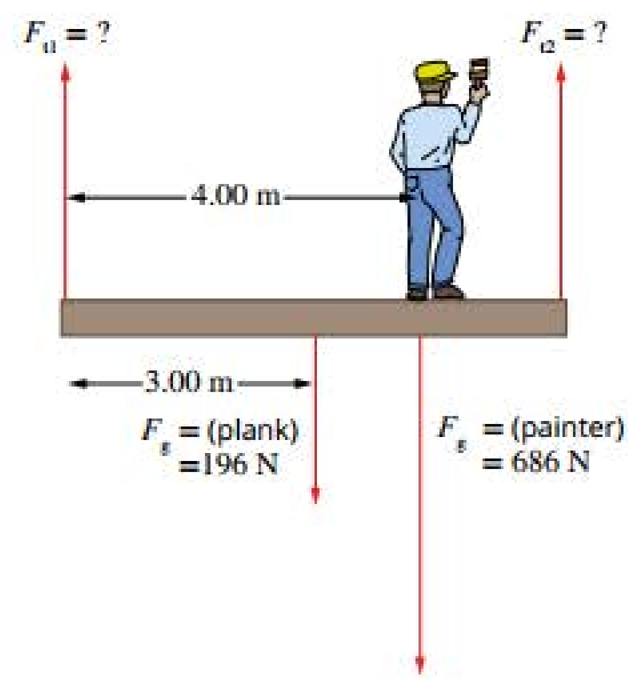
STATIC EQUILIBRIUM WITH TWO UNKNOWN FORCES

The seesaw problem is relatively straightforward since in each situation there is only one unknown force. If there are two unknown forces, the reference point can be chosen to coincide with one of the forces. By using this strategy, it means that the force acting at the reference point contributes no torque (since $r = 0$). The remaining unknown force can be found using the relationship $\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$. Worked example 3.3.3 uses this strategy.

Worked example 3.3.3

CALCULATING STATIC EQUILIBRIUM WITH TWO UNKNOWNS

A 70.0 kg painter stands 4.00 m from the end of a 6.00 m long plank that is supported by a rope at each end. The plank has a mass of 20.0 kg. Determine the tension on the left-hand rope (F_{t1}). Use $g = 9.80\ \text{N kg}^{-1}$ in your calculations where required.



Thinking	Working
Identify the variables involved and state them in their standard form.	$m_{pl} = 20.0\,kg$ $m_{pa} = 70.0\,kg$ $r_{\perp F_{t1}-F_{t2}} = 6.00\,m$ $r_{\perp c-F_{t2}} = 3.00\,m$ $r_{\perp pa-F_{t2}} = 2.00\,m$ $g = 9.80\,N\,kg^{-1}$
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the plank.
Decide the reference point about which the torques will be calculated. Always choose a point at which one of the unknown forces acts, in order to form an equation with just one unknown. This avoids the need to solve simultaneous equations.	The reference point is the point at which the rope providing the tension force F_{t2} is attached.
Decide which force causes the clockwise torques and which forces cause the anticlockwise torques around the chosen reference point.	The tension force of the left-hand rope on the plank provides the clockwise torque. The force of the painter on the plank provides an anticlockwise torque. The force of gravity on the plank provides another anticlockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$
Expand the equation to include each of the torques acting on the plank. The torque of the right-hand rope is not included as it acts through the reference point.	$F_{t1}r_{\perp F_{t1}-F_{t2}} = F_{pl}r_{\perp c-F_{t2}} + F_{pa}r_{\perp pa-F_{t2}}$
Substitute the data into the equation and solve for the unknown force.	$F_{t1} \times 6.00 = 20 \times 9.80 \times 3.00 + 70 \times 9.80 \times 2.00$ $F_{t1} = \frac{20 \times 9.80 \times 3.00 + 70 \times 9.80 \times 2.00}{6.00}$ $= \frac{588 + 1372}{6.00}$ $F_{t1} = 327\,N$

Work example: Try yourself 3.3.3

CALCULATING STATIC EQUILIBRIUM WITH TWO UNKNOWNS

For the painter on the plank scenario in Worked example 3.3.3, determine the tension on the right-hand rope (F_{t2}).

Another way to determine the second unknown force is to apply the conditions for translational equilibrium. You can check these values:

$\Sigma F = 0$: the sum of the two upwards forces (tensions) 555 N + 327 N = 882 N.

This balances the sum of the two downwards forces: 196 N + 686 N = 882 N.

OTHER STATIC EQUILIBRIUM SCENARIOS

The seesaw scenario and the supported plank are just two of many situations in which static equilibrium occur. Other conditions in which static equilibrium exists include the following scenarios.

Cantilevers

A beam that extends beyond its support structure is called a **cantilever**. Cantilevers are common structural elements. For example, the diving board at the local pool is a cantilever.

A cantilever bridge might be used to span a river or valley (Figure 3.3.4). Pillars are built at regular intervals across a river in order to support a number of beams. The cantilever beams are then joined at the centre of each span. The forces on the pillars are not affected by joining the beams; these are the same as if the beams were not connected. All the support for the cantilever is supplied by pillars. Other structures that can involve cantilevers include shelving, awnings over the footpath outside shops, and the wings of a plane.

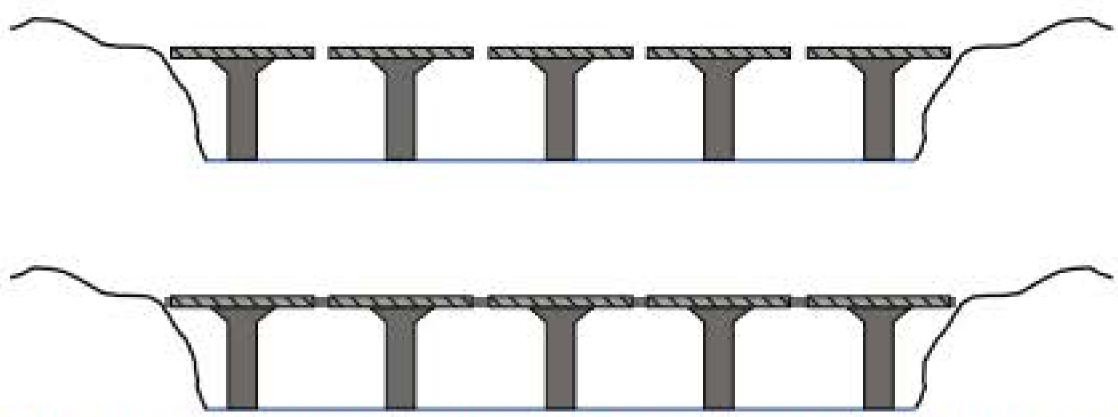


FIGURE 3.3.4 Each beam in the cantilever bridge is fully supported by the pillar below the beam. No added support is provided by connecting the beams.

PHYSICS IN ACTION

Cantilever ruler

- Is it possible to support a ruler with two fingers, when one finger is not in the middle of the ruler? Try this using a 1-metre ruler. A smaller ruler will do if that is all that is available.
- Place your pointer fingers underneath each end of the ruler to support it as shown in Figure 3.3.5.
- Move your fingers towards each other to find the centre of mass of the ruler.
- Place your right hand at the 30 cm mark of the ruler. Now determine where you need to place your left hand if it must be placed at a point that is less than the 50 cm mark.
- You'll find that, regardless of where you start and provided that the ruler is balanced, your fingers will come together at the centre of mass of the ruler. The force downwards from the ruler on each finger will vary depending upon its position and the friction between finger and ruler will vary as a result of the change in that downwards force. The end result is that both fingers end up either side of the centre of mass.

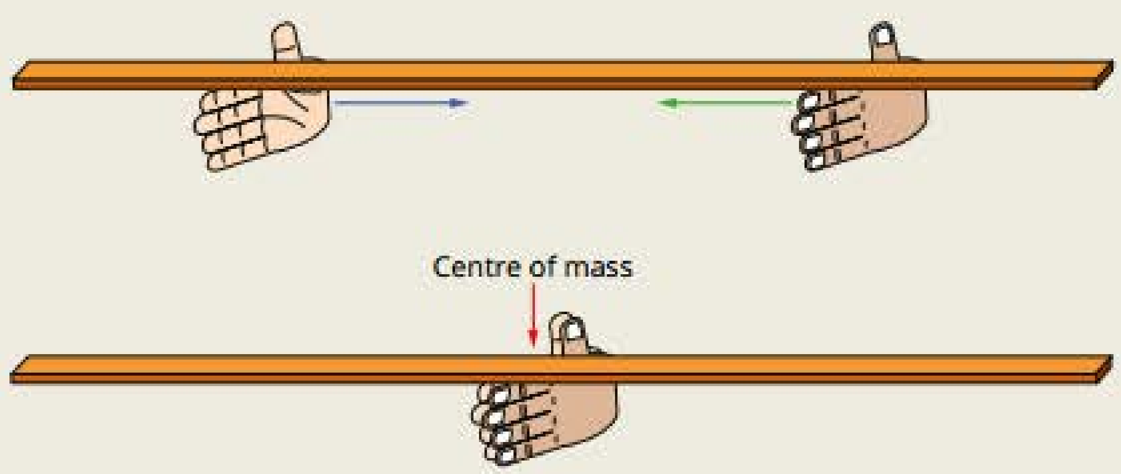


FIGURE 3.3.5 Finding the centre of mass of a ruler with your fingers.

When the centre of mass of the beam is not directly above a support, the force due to gravity that acts on the centre of mass will provide a torque that must be factored into the conditions of rotational equilibrium. In this case two supports are usually required, with one support providing a force causing an upwards force on the beam and the other support providing a downwards force on the beam.

Figure 3.3.6 shows a swimming pool diving board, with one support providing an upwards force on the board and the other support providing a downwards force on the board. A diving board must have a downwards force on the fixed end of the board to provide an opposing torque to the torque provided by the force due to gravity acting on the board.

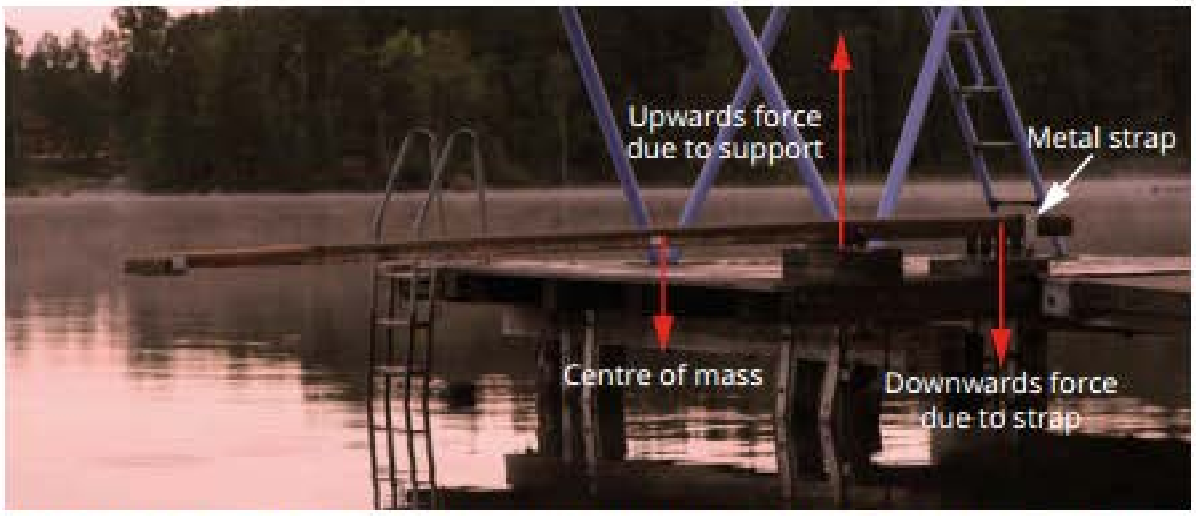
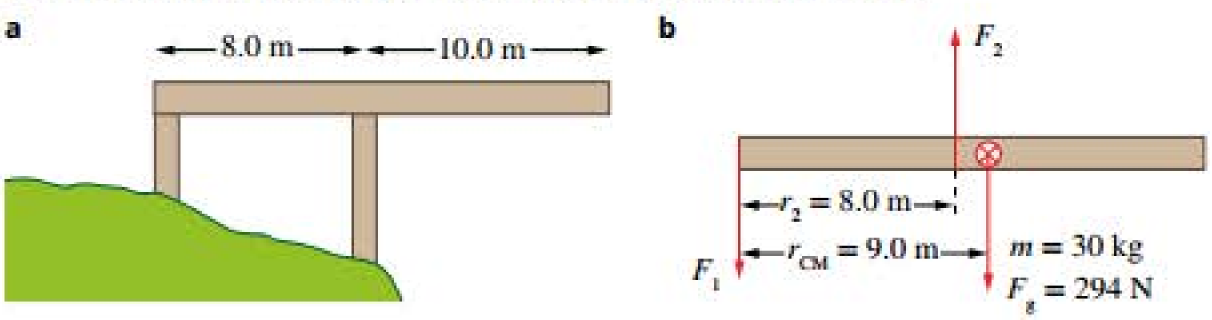


FIGURE 3.3.6 A diving board showing that a metal strap is needed to provide the downwards force on the fixed end of the board.

Worked example 3.3.4

CALCULATING THE STATIC EQUILIBRIUM OF A CANTILEVER

A uniform cantilever beam 18.0 m long is used as a viewing platform. It extends 10.0 m beyond two supports that are 8.00 m apart. The beam has a mass of 30.0 kg. Determine the magnitude and direction of the force that the right-hand support must supply so that the beam is in static equilibrium.

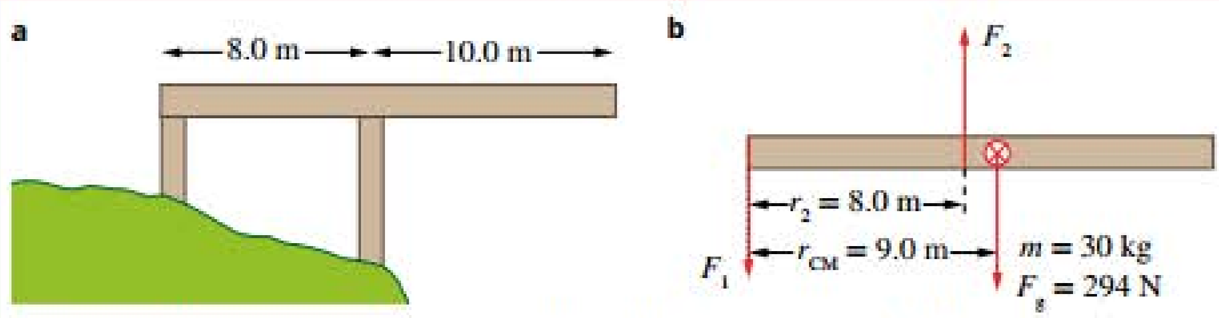


Thinking	Working
Identify the variables involved and state them in their standard form.	m_b = 30.0 kg $r_{\perp F_2-F_1}$ = 8.00 m $r_{\perp c-F_1}$ = 9.00 m g = 9.80 N kg^{-1}
Identify the object that is in rotational equilibrium. This is the object upon which all the torques are acting.	The object experiencing rotational equilibrium is the beam.
Decide the reference point about which the torques will be calculated. Choose the point on which the other unknown force acts to eliminate it as an unknown in the equation.	The reference point is the point at which the support providing the force F_1 is attached.

Decide which force causes the clockwise torque and which force causes the anticlockwise torque around the chosen reference point.	The force of the right-hand support on the beam provides the anticlockwise torque. The force of gravity on the beam provides the clockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the beam. The torque of the left-hand support is not included, as it acts through the reference point.	$F_b r_{\perp c-F_1} = F_2 r_{\perp F_2-F_1}$
Substitute the data into the equation and solve for the unknown.	$30.0 \times 9.80 \times 9.00 = F_2 \times 8.00$ $F_2 = \frac{30 \times 9.80 \times 9.00}{8.00}$ $= \frac{2646}{8.00}$ $= 331\text{ N}$
State the direction of the force acting on the object in equilibrium.	The force of 331 N is upwards on the beam.

Worked example: Try yourself 3.3.4

CALCULATING THE STATIC EQUILIBRIUM OF A CANTILEVER



Determine the magnitude and direction of the force that the left-hand support must supply so that the beam is in static equilibrium (F_1).

You can check these values using $\Sigma F = 0$.

The sum of the upwards forces (F_2) is 331 N.

This balances the sum of the two downwards forces, $(F_g + F_1) = 294 + 36.8 = 331$ N.

Struts and ties

As well as their main beams and pillars, many structures are strengthened in a number of ways. A structure such as a cantilever (Figure 3.3.7) may be supported by struts or ties. A strut is placed below the beam and is rigid, so it will be under compression. A tie, attached from above a beam, may be either rigid or flexible and will be under tension.

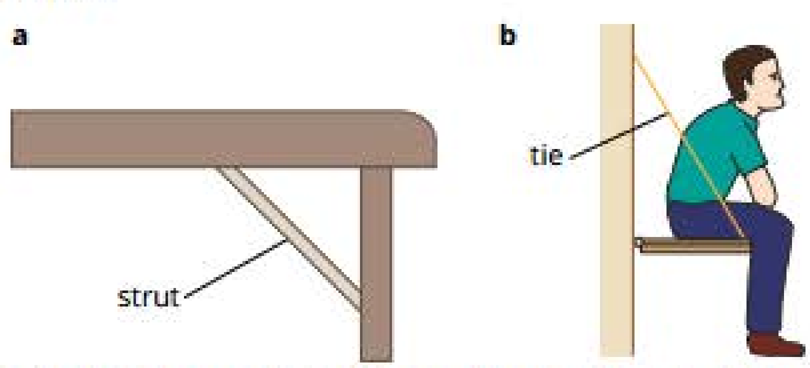
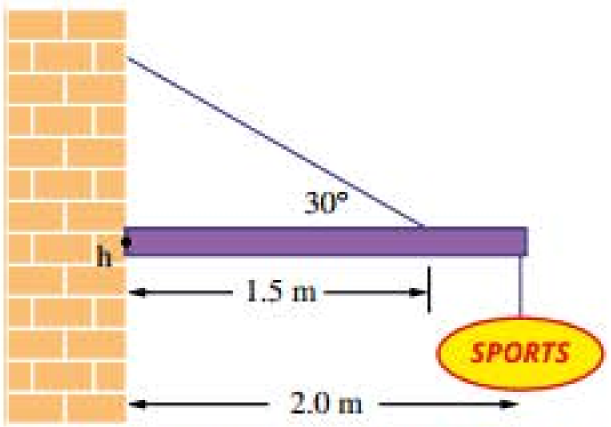


FIGURE 3.3.7 (a) A strut helps to support a cantilevered beam and is under compression. (b) A tie helps to support a fold-out bench and is under tension.

Worked example 3.3.5

CALCULATING THE STATIC EQUILIBRIUM OF A CANTILEVER SUPPORTED BY A TIE

A sign of mass 10 kg is suspended from the end of a uniform 2.0 m long cantilevered beam. The other end of the beam is attached to the wall by a hinge labelled h. The beam has a mass of 25 kg and is further supported by a wire tie that makes an angle of 30° to the beam. The wire is attached to the beam at a point 1.5 m from the wall. Use $g = 9.80\,\text{N kg}^{-1}$ and ignore the mass of the wire for these calculations.



Calculate the tension (F_t) in the wire that is supporting the beam.

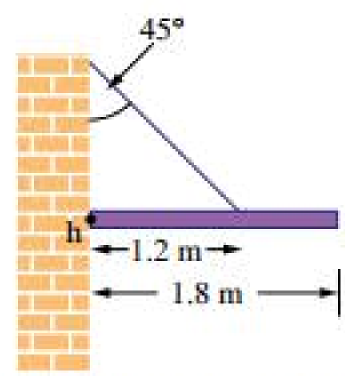
Thinking	Working
Identify the variables involved and state them in their standard form.	$m_b = 25\,\text{kg}$ $m_s = 10\,\text{kg}$ $r_{\perp c\text{-}h} = 1.0\,\text{m}$ $r_{\perp s\text{-}h} = 2.0\,\text{m}$ $r_{\perp w\text{-}h} = 1.5\,\text{m}$ $g = 9.80\,\text{N kg}^{-1}$
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the beam.
Decide the reference point about which the torques will be calculated. Choose the point at which the other unknown force acts to eliminate it from the equation.	The reference point is the hinge (h) at which the beam is connected to the wall.
Decide which force causes the anticlockwise torque, and which forces cause the clockwise torques around the chosen reference point.	The force of the wire tie on the beam provides the anticlockwise torque. The force of gravity on the beam provides one clockwise torque. The force of gravity on the sign provides another clockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the beam.	$F_b r_{\perp c\text{-}h} + F_s r_{\perp s\text{-}h} = F_w r_{\perp w\text{-}h}$
Substitute the data into the equation to solve for the perpendicular distances from the force arm to the line of action of the force.	$r_{\perp w\text{-}h} = r_{\perp w\text{-}h} \sin 30°$ $= 1.5 \times \sin 30°$ $= 0.75\,\text{m}$

Substitute the data into the equation and solve for the unknown force.	$(25 \times 9.80 \times 1.00) + (10 \times 9.80 \times 2.00) = F_t \times 0.75$ $F_t = \frac{245+196}{0.75}$ $= \frac{441}{0.75}$ $= 588\,\text{N}$
By finding the perpendicular distance using the angle the tension force is acting at, the tension force can be calculated directly by the equation. If the vertical component of the tension force was found, then the tension force could be calculated separately using the angle the force is acting at.	

Worked example: Try yourself 3.3.5

CALCULATING THE STATIC EQUILIBRIUM OF A CANTILEVER SUPPORTED BY A TIE

A uniform 5.00 kg beam, 1.80 m long, extends from the side of a building and is supported by a wire tie that is attached to the beam 1.20 m from a hinge (h) at an angle of 45°.



Calculate the tension (F_t) in the wire that is supporting the beam.

3.3 Review

SUMMARY

- Static equilibrium occurs when an object experiences translational equilibrium and rotational equilibrium.
- Static equilibrium can be represented mathematically as $\Sigma F = 0$ and $\Sigma \tau = 0$.
- Rotational equilibrium is best represented mathematically as $\Sigma \tau_{clockwise} = \Sigma \tau_{anticlockwise}$.
- In calculations of static equilibrium, the reference point is the point about which the torques act.
- In calculations of static equilibrium with two unknown forces, the reference point can be placed at the point at which one of the unknown forces acts. This eliminates any torque due to this force, as the distance from the force to the reference point is zero.

KEY QUESTIONS

1 Select the correct statement describing the requirements for an object to be in rotational equilibrium.

A A net torque acts about the reference point and rotation does not occur.

B A net torque acts about the reference point and rotation occurs.

C No net torque acts about the reference point and the rate of rotation does not increase.

D No net torque acts about the reference point and the rate of rotation increases.

2 An adult of mass 75.0 kg sits on a seesaw at a playground with a 25.0 kg child sitting 2.25 m from the pivot point on the other side from the adult. Calculate how far from the pivot the adult must sit for the seesaw to remain balanced and horizontal.

3 Select the statement that correctly describes an object in static equilibrium.

A The object experiences rotational equilibrium, but not translational equilibrium.

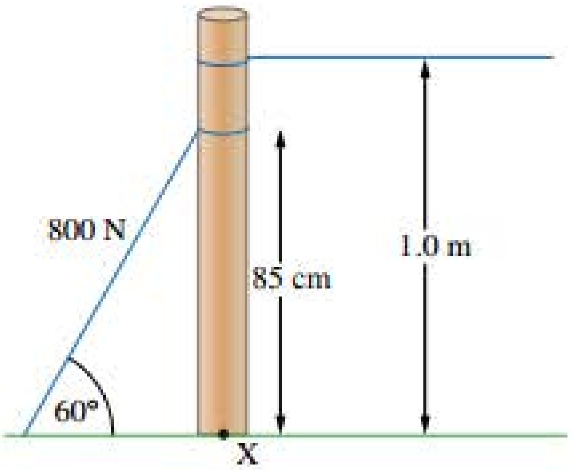
B The object experiences rotational equilibrium and translational equilibrium.

C The object experiences neither rotational equilibrium nor translational equilibrium.

D The object experiences translational equilibrium, but not rotational equilibrium.

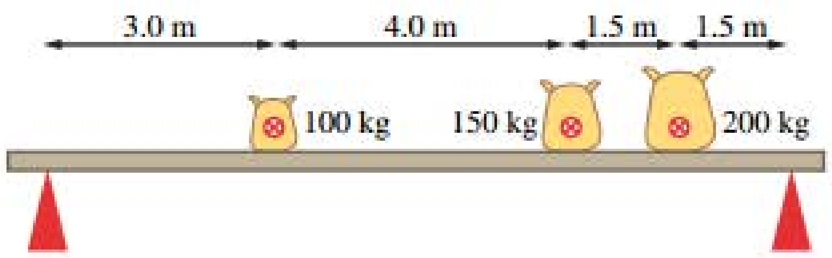
4 An adult of mass 90.0 kg sits on a seesaw at a playground with two 20.0 kg children. One child is sitting 1.50 m from the pivot point and the other is sitting 2.50 m from the pivot point, both on the other side from the adult. Calculate how far from the pivot the adult must sit for the seesaw to remain balanced and horizontal.

5 The end post of a vineyard trellis is held in position by a backstay that is under a tension of 800 N at an angle of 60° to the horizontal. The geometry of the situation is shown in the diagram below.



Using the base of the post, X, as the pivot point, determine the size of the tension in the vineyard trellis wire, F_t.

6 A makeshift shelf is used in a bakery to store sacks of flour. The shelf is constructed using a 10.0 m beam with a mass of 50.0 kg, with one support positioned at each end. The shelf holds sacks of mass 100 kg, 150 kg and 200 kg at the positions shown in the figure below. Calculate the forces on the beam due to the left and right supports.

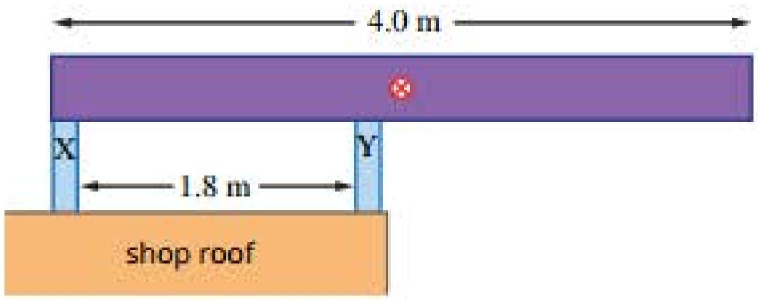


3.3 Review *continued*

The following information applies to questions 7–9.

The value for F_R can be confirmed using torques.

A 4.0 m cantilever-type awning is constructed on the roof of a shop so that it shades the front. The awning has a mass of 900 kg and is supported by two supporting columns, X and Y, which produce forces on the awning of F_X and F_Y respectively.



7 Determine the force supplied by column Y on the awning (F_Y).

8 Determine the direction in which F_X acts on the awning.

9 Determine the force supplied by column X on the awning (F_X).

Chapter review

03

KEY TERMS

axis of rotation
base
cantilever
centre of gravity
centre of mass
component
force arm
line of action
neutral equilibrium
pivot point
principle of moments
reference point
rotational equilibrium
stable equilibrium
static equilibrium
torque
translational equilibrium
unstable equilibrium
vector diagram

Use $g = 9.80\,\text{m s}^{-2}$ to answer the following questions.

1 Which of the options best completes the following statement about torque?
'When a force acts such that the line of action of the force is directed through the pivot point of the object then ...'
A a torque will result
B a torque will result only if the applied force is greater than the weight of the pivot point
C a torque will result only if the weight of the pivot point is greater than the applied force
D no torque will result

2 Tom is riding his skateboard and is in a state of translational equilibrium. Select the correct statement regarding Tom's motion.
A He must be stationary.
B He must be maintaining a constant velocity.
C He must be experiencing a translational acceleration.
D He must be experiencing a rotational acceleration.

3 A cyclist is coasting down a road at a decreasing velocity while standing on the pedals. Which one or more of the following objects is in static equilibrium?
A the rear wheel
B the front wheel
C the front cog connected to the pedals
D the cyclist standing on the pedals

4 A child pulls down on a lever-type door handle with a force of 30.0 N. Given that the length of the handle is 12.0 cm, calculate the maximum possible torque acting on the handle's pivot.

5 A railway signalman is responsible for pushing levers that move train tracks, which switches a train from one line to another. Calculate the torque on the axle at the bottom of a 1.35 m lever if a 74.0 N force acts at an angle of 60.0° to the lever.

6 Define 'stability'. Describe the different types of stability and the factors that affect stability.

The following information applies to questions 7 and 8.
A woman whose car has a flat tyre has two wheel spanners in the boot of her car. One wheel spanner is 15.0 cm long, the other is 75.0 cm long.

7 In order to undo the wheel nuts with a minimum amount of effort, which wheel spanner should the woman select? Explain.

8 If the maximum force that the woman can apply is 45.0 N, determine the maximum torque that can be delivered to a wheel nut.

9 In the following situations, a torque is acting. In each case, identify the axis of rotation or pivot point about which the torque acts and estimate the length of the lever arm.
a A garden tap is turned on.
b A wheelbarrow is lifted by the handles.
c An object is picked up with a pair of tweezers.
d A screwdriver is used to lever open a tin of paint

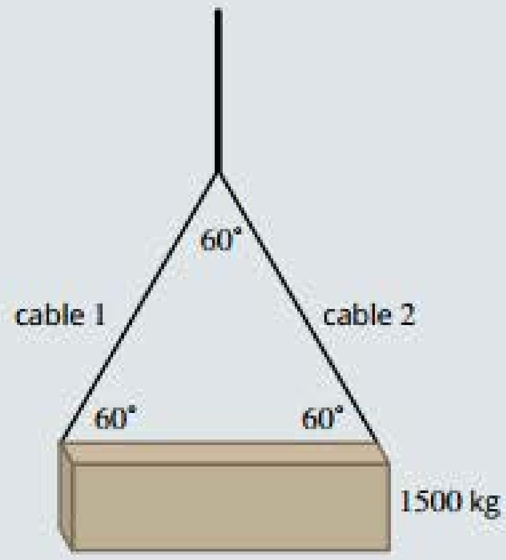
The following information applies to questions 10 and 11.
A crane with a horizontal lever arm is lifting a concrete wall of mass 2.50 tonnes. The load is 20.0 m from the axis of rotation.

10 Calculate the torque created by this load.

11 What stops the crane from toppling over as a result of this torque?

12 Which of the following are in translational equilibrium? More than one correct answer is possible.
A a stationary elevator
B an elevator going up with constant velocity
C an aeroplane during take-off
D a container ship sailing with constant velocity
E a car plummeting off a cliff

The following information applies to questions 13 and 14.
A concrete beam of mass 1500 kg is being lifted by steel cables, as shown in the diagram.



The beam is moving upwards with a constant velocity of 2.00 m s^{-1}. Ignore the mass of the cable.

13 Draw a force diagram showing all the forces acting on the beam.

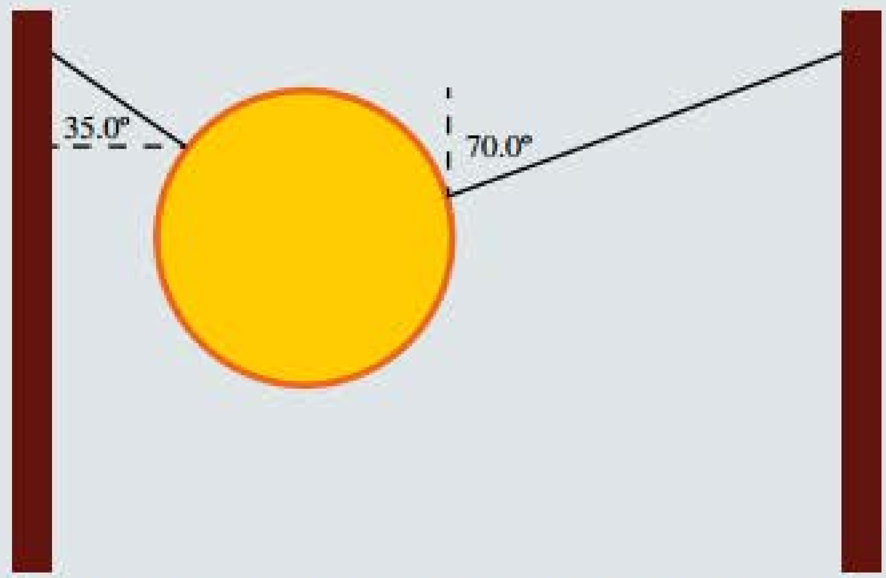
14 Calculate the tension in cable 1 and cable 2. Show your answer in kN.

15 Three cars are crossing a beam bridge of mass 500 kg in a single line. At one instant, the left pillar (X) is providing a force of 2.00 × 10^{4} N upwards. What is the size and direction of the force exerted by pillar Y at this instant?

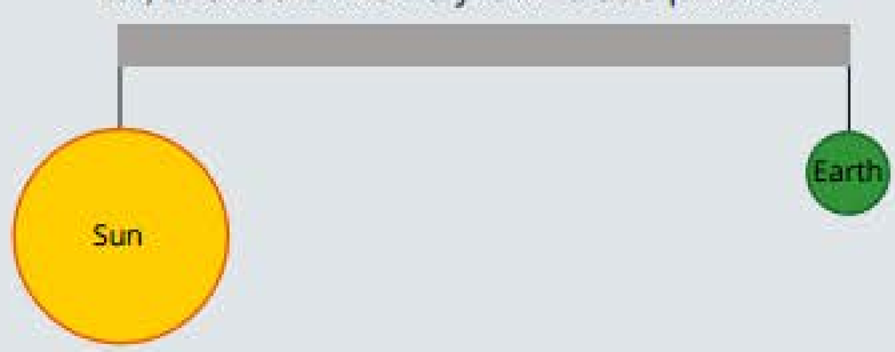
16 A dining table top, of mass 40.0 kg, is supported by four legs. On the table there is 4.50 kg of food. If the table is stationary, and the legs all provide the same support, calculate the force each leg would be exerting on the table.

17 A climber of mass of 86.5 kg is hanging from a rope half-way up a rock wall. Calculate the tension force provided by the rope in order for the climber to be in translational equilibrium.

18 A 12.5 kg sign is hung between two poles by two cables of negligible mass, as shown in the diagram below. The sign is in translational equilibrium. Calculate the tension in the shorter left-hand cable and the longer right-hand cable.



19 A child's toy is to be suspended above her bed from a string attached to the ceiling. The toy is made from a 1.00 m long aluminium rod of mass 10.0 g, with a 145 g model of the Sun hanging at one end of the rod and a 22.5 g model of the Earth hanging at the other end. Calculate how far from the Sun, in centimetres, the ceiling attachment string needs to be tied to the bar, so that the whole toy is in static equilibrium.



CHAPTER

04 Electric fields

In 1820, Hans Christian Oersted discovered that an electric current could produce a magnetic field. His work established the initial ideas behind electromagnetism. Since then, our understanding and application of electromagnetism has developed to the extent that much of our modern way of living relies upon it.

In this chapter you will investigate electric and magnetic fields, the concepts that apply to each, and some of the interactions between these closely related phenomena.

Science as a Human Endeavour

Electromagnetism is utilised in a range of technological applications, including:

- DC electric motor with commutator, and back emf
- AC and DC generators
- transformers
- regenerative braking
- induction hotplates
- large-scale AC power distribution systems

Science Understanding

- electrostatically charged objects exert a force upon one another; the magnitude of this force can be calculated using Coulomb's Law

  This includes applying the relationship

  $$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$$

- point charges and charged objects produce an electric field in the space that surrounds them; field theory attributes the electrostatic force on a point charge or charged body to the presence of an electric field
- a positively charged body placed in an electric field will experience a force in the direction of the field; the strength of the electric field is defined as the force per unit charge

  This includes applying the relationship

  $$E = \frac{F}{q} = \frac{V}{d}$$

- when a charged body moves or is moved from one point to another in an electric field and its potential energy changes, work is done on the charge by the field

  This includes applying the relationship

  $$V = \frac{W}{q}$$

4.1 Electric fields

FIGURE 4.1.1 Charged plasma follows lines of the electric field produced by a Van de Graaff generator.

FIGURE 4.1.2 The gravitational field of the Earth applies a force on the skydiver, while the gravitational field of the skydiver exerts a force on the Earth.

FIGURE 4.1.3 The girl's hair follows the lines of the electric field produced when she became charged while sliding down a plastic slide.

A field is a region of space where objects experience a force due to a physical property related to the field. Gravity, electricity and magnetism can all be described by fields. In Chapter 1, the direction, shape and strength of gravitational fields around a mass were described. In this section electric fields will be explained.

An **electric field** surrounds positive and negative charges, and exerts a force on other charges within the field. Just as a gravitational field can be represented by field lines, so can the electric field around a charged object. This is shown in Figure 4.1.1.

FIELD THEORY

There are four fundamental forces in nature that act at a distance; that is, these forces can exert a force on an object without making any physical contact with it. These are called non-contact forces, and include the strong nuclear force, the weak nuclear force, the electromagnetic force and the gravitational force.

In order to understand these forces, scientists use the idea of a field. A field is a region of space around an object that has certain physical properties such as electric charge or mass. Another object with that physical property in the field will experience a force without any contact between the two objects.

For example, as you saw in Chapter 1, there is a gravitational field around the Earth due to the mass of the Earth. Any object with mass that is located within this gravitational field experiences a force of attraction towards the Earth. According to Newton's third law, there is also an equal and opposite gravitational force acting on the Earth due to the gravitational field of the object. An example of this is shown in Figure 4.1.2.

Similarly, any charged object has a region of space around it (an electric field) in which another charged object will experience a force. This is one aspect of the electromagnetic force. Unlike gravity, which only exerts an attractive force, electric fields can exert forces of attraction or repulsion.

ELECTRIC FIELD LINES

An electric field is a vector quantity, which means it has both magnitude and direction.

In order to visualise electric fields around charged objects, you can use electric **field lines**. Some field lines are already visible—for example, the girl's hair in Figure 4.1.3 is tracing out the path of the field lines. Diagrams of field lines can also be constructed.

Field lines are drawn with arrowheads on them indicating the direction of the force that a small positive test charge would experience if placed in the electric field. Therefore, field lines point away from positively charged objects and towards negatively charged objects. Usually, only a few representative lines are drawn.

Remember: like charges repel and unlike (opposite) charges attract.

The density of field lines (how close they are together) is an indication of the relative strength of the electric field. This is explained in more detail later in this section.

Rules for drawing electric field lines

When drawing electric field lines (in two dimensions) around a charged object there are a few rules that need to be followed.

- Electric field lines go from positively charged objects to negatively charged objects.
- Electric field lines start and end at 90° to the surface, with no gap between the lines and the surface.

- Field lines can never cross; if they did it would indicate that the field is in two directions at that point, which can never happen.
- Around small charged spheres, called **point charges**, the field lines radiate like spokes on a wheel.
- Around point charges you should draw at least eight field lines: top, bottom, left, right and another field line in between each of these.
- Between two point charges, the direction of the field at any point is the resultant field vector determined by adding the field vectors due to each of the two point charges.
- Between two oppositely charged parallel plates, the field lines between the plates are evenly spaced and are drawn straight from the positive plate to the negative plate.
- Always remember that these drawings are two-dimensional representations of a three-dimensional field.

Figure 4.1.4 shows some examples of how to draw electric field lines.

Strength of the electric field

The distance between adjacent field lines indicates the strength of the field. Around a point charge, the field lines are closer together near the charge and get further apart as you move further away. You can see this in the field-line diagrams in Figure 4.1.4. Therefore, the **electric field strength** decreases as the distance from a point charge increases.

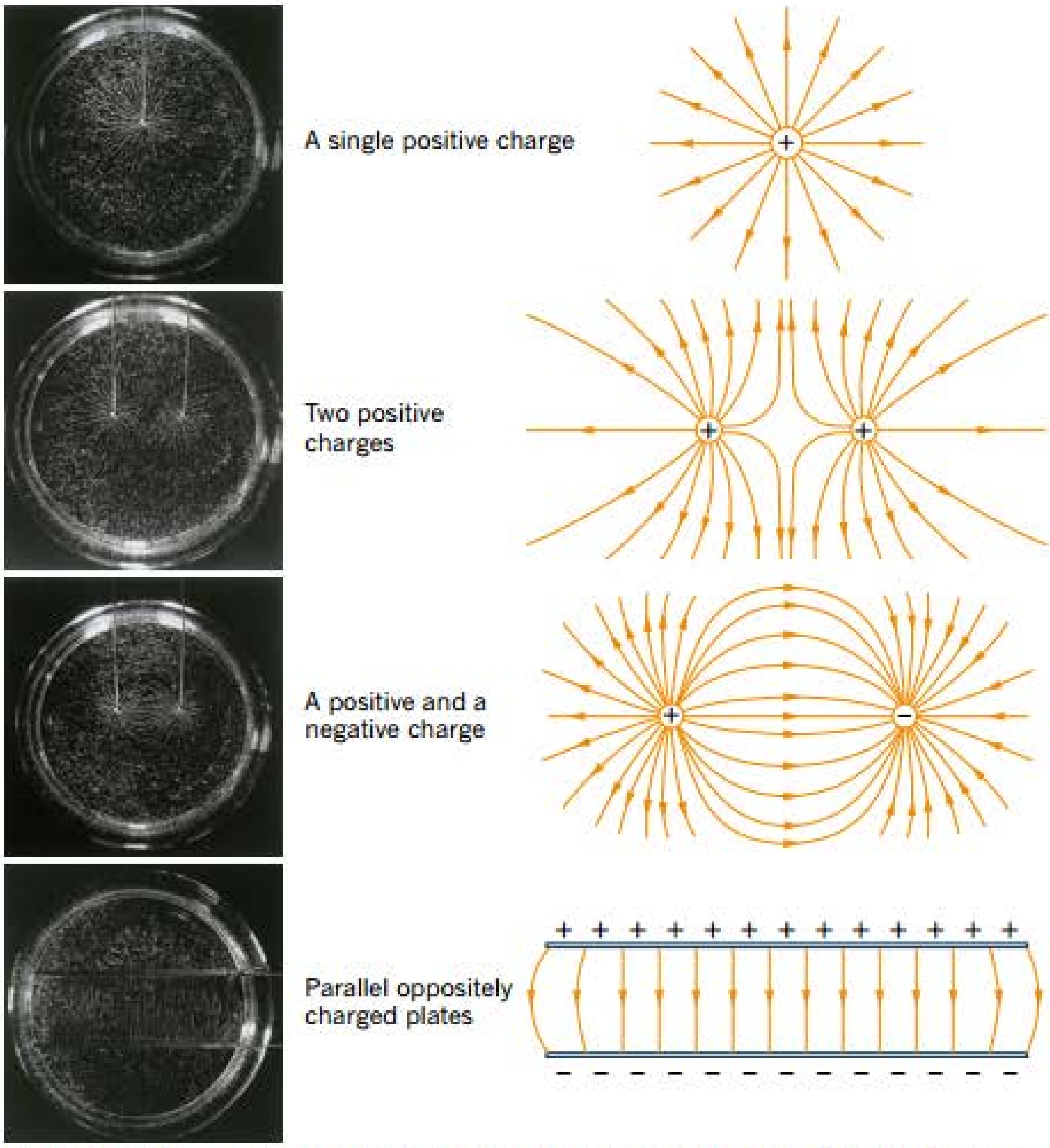


FIGURE 4.1.4 Grass seeds suspended in oil align themselves with the electric field. The diagram next to each photo shows lines representing the electric field.

A uniform electric field is established between two parallel metal plates that are oppositely charged. The field strength is constant at all points within a uniform electric field, so the field lines are parallel.

FORCES ON FREE CHARGES IN ELECTRIC FIELDS

If a charged particle, such as an electron, is placed within an electric field, it experiences a force. The direction of the field and the sign of the charge allow you to determine the direction of the force.

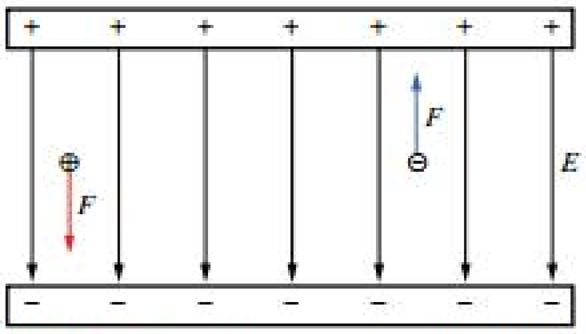


FIGURE 4.1.5 The direction of the electric field (E) indicates the direction in which a force acts on a positive charge. A negative charge experiences a force in the opposite direction to the field.

Figure 4.1.5 shows a positive test charge (proton) and a negative test charge (electron) within a uniform electric field. Recall that the direction of an electric field is defined as the direction of the force that a positive charge experiences within the electric field. So, an electron experiences a force in the opposite direction to the electric field, while a proton experiences a force in the same direction as the field.

The magnitude of the force experienced by a charged particle due to an electric field can be determined using the equation:

> $F = qE$
> where F is the force on the charged particle (N)
> q is the charge of the object experiencing the force (C)
> E is the strength of the electric field ($N\,C^{-1}$)

This equation illustrates that the force experienced by a charge is proportional to the strength of the electric field, E, and the size of the charge, q. The force on the charged particle will cause the charged particle to accelerate in the field. This means that the particle could increase or decrease its velocity, or change its direction while in the field.

To calculate the acceleration due to the force experienced, you can use the equation from Newton's second law:

$F = ma$

where m is the mass of the accelerating particle (kg)
a is the acceleration ($m\,s^{-2}$).

Worked example 4.1.1

USING $F = qE$

Calculate the magnitude of the uniform electric field that would cause a force of 5.00×10^{-21} N on an electron.
($q_e = -1.602 \times 10^{-19}$ C)

Thinking	Working
Rearrange the relevant equation to make electric field strength the subject.	$F = qE$ $E = \frac{F}{q}$
Substitute the values for F and q into the rearranged equation and calculate the answer. (As only magnitude is required, q can be kept positive.)	$E = \frac{F}{q}$ $= \frac{5.00 \times 10^{-21}}{1.602 \times 10^{-19}}$ $= 3.12 \times 10^{-2}\,N\,C^{-1}$

PHYSICSFILE

Bees

Bees are thought to use electric fields to communicate, find food and to avoid flowers that have been visited by another bee recently. Their antennae are bent (deflected) by electric fields and they can sense the amount of deflection.

The charge that builds up on their bodies helps them collect pollen grains and transport them to other flowers (Figure 4.1.6). The altered electric field around a flower that has recently been visited is a signal to other bees to find food elsewhere.


FIGURE 4.1.6 A bee can detect changes in the electric field in its environment.

Worked example: Try yourself 4.1.1

USING $F = qE$

Calculate the magnitude of the uniform electric field that creates a force of 9.00×10^{-23} N on a proton.
($q_p = +1.602 \times 10^{-19}$ C)

Electric field strength

Electric field strength can be thought of as the force applied per coulomb of charge, which is expressed in the equation:

$$E = \frac{F}{q}$$

An alternative measure of the electric field strength is volts per metre, which is calculated using the equation:

$$E = \frac{V}{d}$$

where d is distance (m).

You can equate both expressions and rearrange them to find an expression for the work done (J) to make a charged particle move a distance against a **potential difference**:

$$\frac{F}{q} = \frac{V}{d}$$

$Fd = qV$ and since $W = Fd$

$W = qV$

You will examine this expression in more detail in Section 4.3.

EXTENSION

Gravitational force and electric force

Oppositely charged parallel plates can be arranged one above the other, such that the electric field is vertical. The direction of the field can then be manipulated to create an upwards force on a charged particle in the field.

If this upwards electric force created by the field on the charged object is equal to the gravitational force on (or weight of) the object, then these two forces can add to provide a net force equal to zero. This means that the charged object will either be suspended between the plates, or (by Newton's first law of motion) will be falling or rising at constant velocity.

This phenomenon was used by Robert Millikan and his PhD student Harvey Fletcher in their oil drop experiment, performed in 1909, to determine the fundamental charge of an electron to within 1% of the currently accepted value.

4.1 Review

SUMMARY

- An electric field is a region of space around a charged object in which another charged object will experience a force.
- Electric fields are represented using field lines.
- Electric field lines point in the direction of the force that a positive charge within the field would experience.
- A positive charge experiences a force in the direction of the electric field and a negative charge experiences a force in the opposite direction to the field.
- The spacing between the field lines indicates the strength of the field. The closer together the lines are, the stronger the field.
- Electric field strength can be expressed as $E = \frac{F}{q}$ and as $E = \frac{V}{d}$.
- Around point charges the electric field radiates in all directions (three-dimensionally).
- Between two oppositely charged parallel plates, the field lines are parallel and therefore the field has a uniform strength.
- When charges are in an electric field, they accelerate in the direction of the force acting on them.
- The force on a charged particle can be determined using the equation $F = qE$.
- Force can be related to the acceleration of a particle using the equation $F = ma$.

KEY QUESTIONS

1 Which of the following options correctly describes an electric field?

A a region around a charged object that causes a charge on other objects within that region
B a region around a charged object that causes a force on other objects within that region
C a region around a charged object that causes a force on other charged objects in that region
D a region around an object that causes a force on other objects within that region

2 Which of the following options correctly defines the direction of an electric field?

A away from a negatively charged object
B away from a positively charged object
C away from a neutrally charged object
D towards a positively charged object

3 Identify whether the rules below for drawing electric field lines are true or false.

a Electric field lines start and end at 90° to the surface, with no gap between the lines and the surface.
b Field lines can cross; this indicates that the field is in two directions at that point.
c Electric fields go from negatively charged objects to positively charged objects.
d Around small charged spheres called point charges you should draw at least eight field lines: top, bottom, left, right and in between each of these.
e Around point charges the field lines radiate like spokes on a wheel.
f Between two point charges, the direction of the field at any point is the field due to the closest of the two point charges.
g The field between two oppositely charged parallel plates is evenly spaced and is drawn straight from the negative plate to the positive plate.

4 Draw diagrams of the field lines for each of the following.

a a single negative point charge
b two positive point charges
c a positive point charge and an earthed plate

5 Calculate the force applied to a balloon carrying a charge of 5.00 mC in a uniform electric field of 2.50 $N C^{-1}$.

6 Calculate the charge on a plastic disc if it experiences a force of 0.025 N in a uniform electric field of 18 $N C^{-1}$.

7 Calculate the acceleration an electron undergoes in a uniform electric field of 3.25 $N C^{-1}$.

The mass of an electron is 9.11×10^{-31} kg and its charge is -1.602×10^{-19} C.

4.2 Coulomb's law

The electromagnetic force is one of nature's fundamental forces. It was Charles Coulomb, in 1785, who first published the quantitative details of the force that acts between two electric charges. The force between any combination of electrical charges can be understood in terms of the force between two 'point charges' separated by a certain distance, as seen in Figure 4.2.1. The effect of distance on the electric field strength from a single charge and the force created by that field between charges is explored in this section.

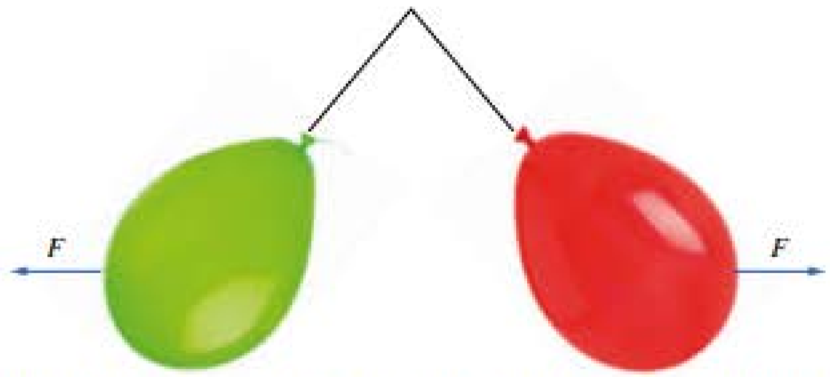


FIGURE 4.2.1 Two similarly charged balloons will repel each other by applying a force on each other.

THE FORCE BETWEEN CHARGED PARTICLES

Coulomb found that the force between two point charges (q_1) and (q_2) separated by a distance (r) was proportional to the product of the two charges, and inversely proportional to the square of the distance between them.

This is another example of the inverse square law discussed in Chapter 1.

> Coulomb's law can be expressed by the following equation:
>
> $$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$$
>
> where F is the force on each charged object (N)
> q_1 is the charge on one point (C)
> q_2 is the charge on the other point (C)
> r is distance between each charged point (m)
> ε_0 is the permittivity of free space, $\varepsilon_0 = 8.8542 \times 10^{-12}\,C^2\,N^{-1}\,m^{-2}$

By including the sign of the charges in the calculation, a positive force value indicates repulsion and a negative force value indicates attraction.

The permittivity of free space (ε_0) has a value of $8.8542 \times 10^{-12}\,C^2\,N^{-1}\,m^{-2}$ in air or a vacuum. As this value is constant for air or a vacuum, the expression at the front of Coulomb's law can be calculated. The result of the calculation is given the name Coulomb's constant (k) and is equal to $8.9875\,m \times 10^9\,N\,m^2\,C^{-2}$. For ease of calculations this is usually rounded to two significant figures as $9.0 \times 10^9\,N\,m^2\,C^{-2}$. So, in Coulomb's law, if:

$$k = \frac{1}{4\pi\varepsilon_0}$$

then the equation becomes:

> $$F = k\frac{q_1 q_2}{r^2}$$
>
> where $k = 9.0 \times 10^9\,N\,m^2\,C^{-2}$

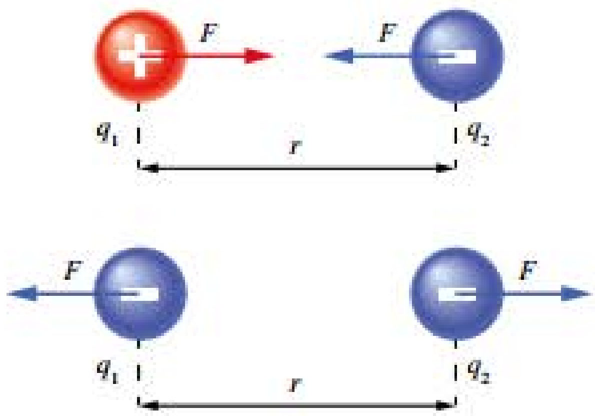


FIGURE 4.2.2 Forces acting between two point charges.

FIGURE 4.2.3 Two 1 C charges 1 m apart would produce a force of 10^{10} N, which is almost twice the weight of the Sydney Harbour Bridge.

Factors affecting the electric force

The force between two charged points is proportional to the product of the two charges (Figure 4.2.2). If there was a force of 10 N between two charged points and either charge was doubled, then the force between the two points would increase to 20 N. It is interesting to note that regardless of the charge on each point, the forces on each point in a pair will be the same. For example, if q_1 is +10 μC and q_2 is +10 μC, then the repulsive force on each of these points would be equal in magnitude. The forces would also be equal on both points if q_1 is +100 μC and q_2 is +1 μC.

The force is also inversely proportional to the square of the distance between the two charged points. This means that if the distance between q_1 and q_2 is doubled, the force on each point charge will decrease to one-quarter of the previous value.

One coulomb in perspective

Coulomb's law can be used to calculate the force between two charges of 1 C each, placed 1 m apart. The force would be 9.0×10^9 N, or approximately 10^{10} N. This is equivalent to the weight provided by a mass of 918 000 tonnes (Figure 4.2.3).

This demonstrates that a 1 C charge is a huge amount of charge. In reality, the amount of charge that can be placed on ordinary objects is a tiny fraction of a coulomb. Even a highly charged Van de Graaff generator will have only a few microcoulombs ($1\,\mu C = 10^{-6}$ C) of excess charge.

Another way to get a feel for the magnitude of electrical forces is to realise that all matter is held together by the electrical forces between atoms. For example, the mass of Mount Everest is supported by the electrostatic repulsion between the electrons around neighbouring atoms in the rock underneath it. The strength of the hardest steel is due to the electrical forces of attraction between its metal ions and the delocalised electrons between them. In comparison to the Earth's gravitational force on an atom, the electrical forces between atoms are about a billion billion (1×10^{18}) times stronger. In fact, only in the last stages of the gravitational collapse of a giant star can the gravitational forces overwhelm the electrical forces between its atoms and cause the star to collapse into a super-dense neutron star.

Worked example 4.2.1

USING COULOMB'S LAW TO CALCULATE FORCE

Two small spheres A and B act as point charges separated by 10.0 cm in air. Calculate the force on each point charge if A has a charge of 3.00 μC and B has a charge of −45.0 nC.
(Use $\varepsilon_0 = 8.8542 \times 10^{-12}\,C^2 N^{-1} m^{-2}$.)

Thinking	**Working**
Convert all values to SI units.	$q_A = 3.00 \times 10^{-6}$ C $q_B = 45.0 \times 10^{-9} = -4.50 \times 10^{-8}$ C $r = 0.100$ m
State Coulomb's law.	$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$
Substitute the values for q_A, q_B, r and ε_0 into the equation and calculate the answer.	$F = \frac{1}{4\pi \times 8.8542 \times 10^{-12}} \times \frac{3.00 \times 10^{-6} \times -4.50 \times 10^{-8}}{0.100^2}$ $= -0.121$ N
Assign a direction based on a negative force being attraction and a positive force being repulsion.	$F = 0.121$ N attraction

Worked example: Try yourself 4.2.1

USING COULOMB'S LAW TO CALCULATE FORCE

Two small spheres A and B act as point charges separated by 75.0 mm in air. Calculate the force on each point charge if A has a charge of 475 nC and B has a charge of 833 pC.

(Use $\varepsilon_0 = 8.8542 \times 10^{-12}\,C^2 N^{-1} m^{-2}$.)

Worked example 4.2.2

USING COULOMB'S LAW TO CALCULATE CHARGE

Two small positive point charges with equal charge are separated by 1.25 cm in air. Calculate the charge on each point charge if there is a repulsive force of 6.48 mN between them.

(Use $k = 9.0 \times 10^9\,N\,m^2 C^{-2}$.)

Thinking	Working
Convert all values to SI units.	$F = 6.48 \times 10^{-3}\,N$ $r = 1.25 \times 10^{-2}\,m$
State Coulomb's law.	$F = k\frac{q_1 q_2}{r^2}$
Substitute the values for F, r and k into the equation and calculate the answer.	$q_1 q_2 = \frac{Fr^2}{k}$ $= \frac{6.48 \times 10^{-3} \times (1.25 \times 10^{-2})^2}{9.0 \times 10^9}$ $= 1.125 \times 10^{-16}$ As $q_1 = q_2$: $q_1^2 = 1.125 \times 10^{-16}$ $q_1 = \sqrt{1.125 \times 10^{-16}}$ $= +1.06 \times 10^{-8}\,C$ $q_1 = q_2 = +1.06 \times 10^{-8}\,C$

Worked example: Try yourself 4.2.2

USING COULOMB'S LAW TO CALCULATE CHARGE

Two small point charges are produced by transferring a number of electrons from q_1 to q_2, and are separated by 12.7 mm in air. The charges on the two points are equal and opposite. Calculate the charge on q_1 and q_2 if there is an attractive force of 22.5 μN between them.

(Use $k = 9.0 \times 10^9\,N\,m^2 C^{-2}$.)

THE ELECTRIC FIELD AT A DISTANCE FROM A CHARGE

In the previous section, the electric field, E, is defined as being proportional to the force exerted on a positive test charge and inversely proportional to the magnitude of that charge; it is measured in $N\,C^{-1}$. Defining the electric field in this way means that it is independent of the size of the charge and describes only the effect of the charge creating the field at a particular point.

It is useful also to be able to determine the electric field at a distance from a single point charge.

The magnitude of the electric field at a distance from a single point charge is given by:

$$E = k\frac{q}{r^2}$$

where E is the strength of the electric field around a point ($N\,C^{-1}$)
q is the charge on the point creating the field (C)
r is the distance from the charge (m)
$k = 9.0 \times 10^9\,N\,m^2\,C^{-2}$

The magnitude of E that is determined is independent of the value of the test charge and depends only on the charge, q, producing the field. This formula can also be referred to as Coulomb's law, in this case for determining the magnitude of the electric field produced by a point charge.

Worked example 4.2.3

ELECTRIC FIELD OF A SINGLE POINT CHARGE

Calculate the magnitude and direction of the electric field at a point P at a distance of 20 cm below a negative point charge, q, of 2.0×10^{-6} C.

Thinking	Working
Convert units to SI units as required.	$q = -2.0 \times 10^{-6}$ C $r = 20\,cm = 0.20\,m$
Substitute the known values to find the magnitude of E using $E = k\frac{q}{r^2}$.	$E = k\frac{q}{r^2}$ $= 9.0 \times 10^9 \times \frac{2.0 \times 10^{-6}}{0.20^2}$ $= 4.5 \times 10^5\,N\,C^{-1}$
The direction of the field is defined as that acting on a positive test charge (see previous section). Point P is below the charge.	Since the charge is negative, the direction will be toward the charge q, or upwards.

Worked example: Try yourself 4.2.3

ELECTRIC FIELD OF A SINGLE POINT CHARGE

Calculate the magnitude and direction of the electric field at point P at a distance of 15 cm to the right of a positive point charge, q, of 2.0×10^{-6} C.

4.2 Review

SUMMARY

- Coulomb's law for the force between two charges q_1 and q_2 separated by a distance of r is:

  $$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$$

- The permittivity constant, ε_0, in Coulomb's law has a value of $8.8542 \times 10^{-12}\,C^2 N^{-1} m^{-2}$.
- For air or a vacuum, the expression $\frac{1}{4\pi\varepsilon_0}$ at the front of Coulomb's law can be simplified to the value of k, called Coulomb's constant, which has a value of approximately $9.0 \times 10^9\,N\,m^2 C^{-2}$. So:

  $$F = k\frac{q_1 q_2}{r^2}$$

- The magnitude of the electric field, E, at a distance r from a single point charge q is given by:

  $$E = k\frac{q}{r^2}$$

  where $k = 9.0 \times 10^9\,N\,m^2 C^{-2}$ (Coulomb's constant).

KEY QUESTIONS

1 Choose a response from each shaded box to correctly complete the following table summarising forces, charges and actions.

	Force	q_1 charge	q_2 charge	Action
a	positive	positive	positive / negative	attraction / repulsion
b	negative	positive / negative	positive	attraction / repulsion
c	positive	negative	positive / negative	attraction / repulsion
d	negative	positive / negative	negative	attraction / repulsion

2 A point charge, q, is moved from a position 30 cm away from a test charge to a position 15 cm from the same test charge. If the magnitude of the original electric field, E, was $6.0 \times 10^3\,N C^{-1}$, what is the magnitude of the electric field at the new position?

A $3.0 \times 10^3\,N C^{-1}$
B $6.0 \times 10^3\,N C^{-1}$
C $12.0 \times 10^3\,N C^{-1}$
D $24.0 \times 10^3\,N C^{-1}$

3 A hydrogen atom consists of a proton and an electron separated by a distance of 53 pm (picometres). Calculate the magnitude and sign of the force applied to a proton carrying a charge of $+1.602 \times 10^{-19}$ C by an electron carrying a charge of -1.602×10^{-19} C. ($1\,pm = 1 \times 10^{-12}$ m and $\varepsilon_0 = 8.8542 \times 10^{-12}\,C^2 N^{-1} m^{-2}$)

4 The electric field is being measured at point P, a distance of 5.0 cm from a positive point charge, q, of 3.0×10^{-6} C. What is the magnitude of the field at P to two significant figures? (Use $k = 9 \times 10^9\,N\,m^2 C^{-2}$.)

5 Calculate the magnitude of the force that would exist between two point charges of 1.00 C each, separated by 1.00 km. (Use $k = 9 \times 10^9\,N\,m^2 C^{-2}$.)

6 A point charge of 6.50 mC is suspended from a ceiling by an insulated rod. At what distance vertically below the point charge will a small sphere of mass 10.0 g with a charge of −3.45 nC be located if it is suspended in air? (Use $k = 9 \times 10^9\,N\,m^2 C^{-2}$.)

7 A charge of $+q$ is placed a distance r from another charge also of $+q$. A repulsive force of magnitude F is found to exist between them. Choose the correct answer from the options in bold to describe the changes, if any, that will occur to the force in the following.

a If one of the charges is doubled to $+2q$, the force will **halve/double/quadruple/quarter** and **repel/attract**.
b If both charges are doubled to $+2q$, the force will **halve/double/quadruple/quarter** and **repel/attract**.
c If one of the charges is changed to $-2q$, the force will **halve/double/quadruple/quarter** and **repel/attract**.
d If the distance between the charges is halved to 0.5r, the force will **halve/double/quadruple/quarter** and **repel/attract**.

8 Calculate the repulsive force on each proton in a helium nucleus separated in a vacuum by a distance of 2.50 fm. (Use $k = 9 \times 10^9\,N\,m^2 C^{-2}$, $1\,fm = 1 \times 10^{-15}$ m and $q_p = +1.602 \times 10^{-19}$ C.)

9 Two point charges 30.0 cm apart in air are charged by transferring electrons from one point to another. Calculate how many electrons must be transferred so that an attractive force of 1.0 N exists. Consider only the magnitude of q_e in your calculations. (Use $k = 9 \times 10^9\,N\,m^2 C^{-2}$ and $q_e = -1.602 \times 10^{-19}$ C.)

4.3 Work done in an electric field

WORK DONE IN UNIFORM ELECTRIC FIELDS

Electrical potential energy is a form of energy that is stored in a field. Work is done on the field when a charged particle is forced to move in the electric field. Conversely, when energy is stored in the electric field then work can be done by the field on the charged particle.

Electrical potential (V) is defined as the work required per unit charge to move a positive point charge from infinity to a place in the electric field. The electrical potential at infinity is defined as zero. This definition leads to the equation:

$$V = \frac{W}{q}$$

$$W = qV$$

where W is the work done on a positive point charge or on the field (J)
q is the charge of the point charge (C)
V is the electrical potential ($J\,C^{-1}$) or volts (V)

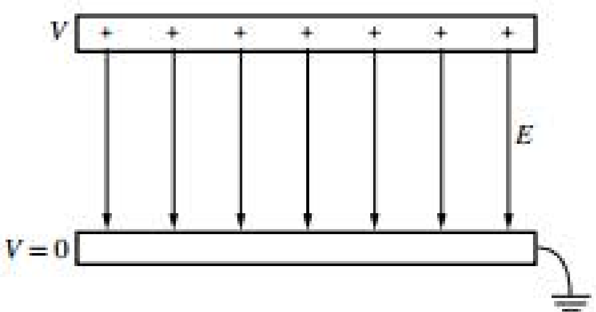


FIGURE 4.3.1 The potential of two plates when one has a positive potential and the other is earthed.

Consider two parallel plates, as shown in Figure 4.3.1, in which the positive plate is at a potential (V) and the other plate is earthed, which is defined as zero potential. The difference in potential between these two plates is called the electrical potential difference (V).

Between any two points in an electric field (E) separated by a distance (d) that is parallel to the field, the potential difference V is defined as the change in the electrical potential between these two points (Figure 4.3.2).

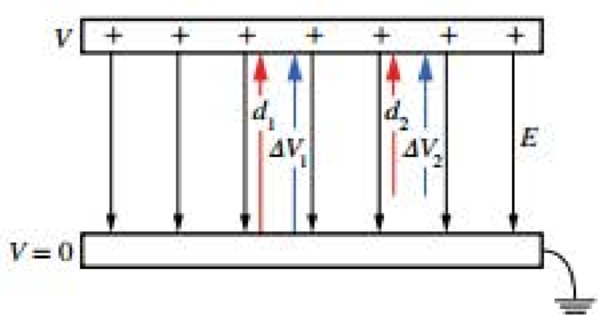


FIGURE 4.3.2 The potential difference between two points in a uniform electric field.

$$E = \frac{V}{d}$$

$$V = Ed$$

where V is the difference in electrical potential (V)
E is the electrical field strength ($V\,m^{-1}$)
d is the distance between points, parallel to the field (m)

CALCULATING WORK DONE

By combining the two equations mentioned so far, you can derive an equation for calculating the work done on a point test charge to move it a distance across a potential difference.

$W = qV$ and $V = Ed$

so $W = qEd$

where W is the work done on the point charge or on the field (J)
q is the charge of the point charge (C)
E is the electrical field strength ($V\,m^{-1}$ or $N\,C^{-1}$)
d is the distance between points, parallel to the field (m)

EXTENSION

Dimensional analysis of the units for field strength

To calculate the value of E in the above equation, you can use either unit for electrical field strength, since they are equivalent. The following dimensional analysis shows why they are equivalent.

$$E = \frac{V}{d} \text{ and } E = \frac{F}{q}$$

$$\text{so } \frac{V}{d} = \frac{F}{q}$$

Looking at the units for each side of the equation, V m^{-1} must equal N C^{-1}.

To prove this, you can break down each unit:

$$\text{V} = \text{J C}^{-1} = \text{kg m}^2\text{s}^{-2}\text{C}^{-1}$$

$$\text{so V m}^{-1} = (\text{kg m}^2\text{s}^{-2}\text{C}^{-1})\text{m}^{-1} = \text{kg m s}^{-2}\text{C}^{-1}$$

Since $\text{N} = \text{kg m s}^{-2}$:

$$\text{V m}^{-1} = \text{kg m s}^{-2}\text{C}^{-1} = \text{N C}^{-1}$$

Work done by or on an electric field

When calculating work done, which changes the electrical potential energy, remember that work can be done either:

- by the electric field on a charged object, or
- on the electric field by forcing the object to move.

You need to examine what is happening in a particular situation to know how the work is being done.

For example, if a charged object is moving in the direction it would naturally tend to go within an electric field, then work is done by the field. So when a positive point charge is moved in the direction of the electric field, the electric field has done work on the point charge. (Refer to q_2 in Figure 4.3.3.)

When work is done by a charged object on an electric field, the object is forced to move against the direction it would naturally go. Work has been done on the field by forcing the object to move. For example, if a force causes a positive charge to move towards the positive plate within a uniform electric field, work has been done on the electric field by forcing the object to move. (See q_1 in Figure 4.3.3.)

If a charge doesn't move any distance parallel to the field then no work is done on or by the field. (See q_3 in Figure 4.3.3.)

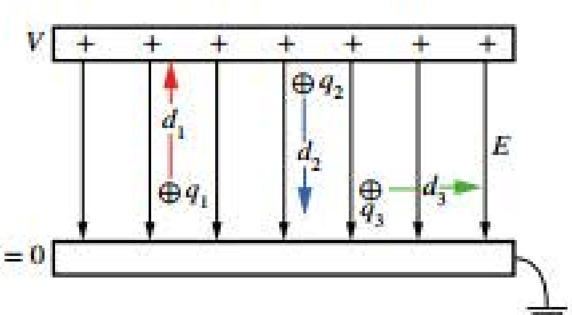


FIGURE 4.3.3 Work is being done on the field by moving q_1 and work is being done by the field on q_2. No work is done on q_3 since it is moving perpendicular to the field.

Worked example 4.3.1

WORK DONE ON A CHARGE IN A UNIFORM ELECTRIC FIELD

A student sets up a parallel plate arrangement so that one plate is at a potential of 12.0 V and the other earthed plate is positioned 0.50 m away. Calculate the work done to move a proton a distance of 10.0 cm towards the negative plate. ($q_p = +1.602 \times 10^{-19}$ C)

In your answer identify what does the work and what the work is done on.

Thinking	Working
Identify the variables presented in the problem to calculate the electric field strength *E*.	$V_2 = 12.0$ V $V_1 = 0$ V $d = 0.50$ m
Use the equation $E = \frac{V}{d}$ to determine the electric field strength.	$E = \frac{V}{d}$ $= \frac{12.0 - 0}{0.50}$ $= 24.0\ \text{V m}^{-1}$
Use the equation $W = qEd$ to determine the work done. Note that *d* here is the distance the proton moves.	$W = qEd$ $= 1.602 \times 10^{-19} \times 24.0 \times 0.100$ $= 3.84 \times 10^{-19}$ J
Determine if work is done on the charge by the field or if work is done on the field.	As the positively charged proton is moving naturally towards the negative plate, work is done on the proton by the field.

Worked example: Try yourself 4.3.1

WORK DONE ON A CHARGE IN A UNIFORM ELECTRIC FIELD

A student sets up a parallel plate arrangement so that one plate is at a potential of 36.0 V and the other earthed plate is positioned 2.00 m away. Calculate the work done to move an electron a distance of 75.0 cm towards the negative plate.

($q_e = -1.602 \times 10^{-19}$ C)

In your answer identify what does the work and what the work is done on.

4.3 Review

SUMMARY

- Electrical potential energy is stored in an electric field.
- When a charged object is moved against the direction it would naturally move in an electric field, then work is done on the field.
- When a charged object is moved in the direction it would naturally tend to move in an electric field, then the field does work on the particle.
- The work done on or by an electric field can be calculated using the equations $W = qV$ or $W = qEd$.

KEY QUESTIONS

1. Calculate the potential difference between two metal plates separated by 30.0 cm, parallel to the field lines, in an electric field of strength $4000\,V\,m^{-1}$.
2. A researcher sees an oil drop of mass 1.161×10^{-14} kg stationary between two horizontal parallel plates. An electric field of strength $3.55 \times 10^{4}\,N\,C^{-1}$ exists between the plates. The field is pointing vertically downwards. Calculate how many extra electrons are present on the oil drop.
3. For each of the following charged objects in a uniform electric field, determine if work was done on the field or by the field or if no work is done.
   - **a** An electron moves towards a positive plate.
   - **b** A positively charged point remains stationary.
   - **c** A proton moves towards a positive plate.
   - **d** A lithium ion (Li^{+}) moves parallel to the plates.
   - **e** An alpha particle moves away from a negative plate.
   - **f** A positron moves away from a positive plate.
4. An alpha particle is located in a parallel plate arrangement that has a uniform electric field of $34.0\,V\,m^{-1}$.
   - **a** Calculate the work done to move the alpha particle a distance of 1.00 cm from the earthed plate to the plate with a positive potential.
   - **b** For the situation in part (a) decide whether work was done on the field, by the field or if no work was done.
5. A charged particle is situated between two parallel plates. The electric field between the plates is $5.05 \times 10^{7}\,N\,C^{-1}$ and it does 1.24 J of work on the particle in moving it a distance of 11.0 mm perpendicular to the plates. Calculate the magnitude of the charge on the particle.

Chapter review

04

KEY TERMS

electric field
electric field strength
electrical potential
field lines
point charge
potential difference

1 Calculate the force applied to an oil drop carrying a charge of 3.00 mC in a uniform electric field of 7.50 $N C^{-1}$.

2 A test charge is placed at a point, P, 30 cm directly above a charge, q, of $+30 \times 10^{-6}$ C. What is the magnitude and direction of the electric field at point P?
   A 300 $N C^{-1}$ downwards
   B 300 $N C^{-1}$ upwards
   C 3×10^6 $N C^{-1}$ downwards
   D 3×10^6 $N C^{-1}$ upwards

3 Explain the difference between electrical potential and potential difference.

4 Calculate the potential difference that exists between two points separated by 25.0 mm, parallel to the field lines, in an electric field of strength 1000 $V m^{-1}$.

5 Between two plates forming a uniform electric field, where will the electrical field strength be at a maximum?
   A close to the positive plate
   B close to the earthed plate
   C at all points between the plates
   D at the mid-point between the plates

6 Choose the correct terms from the ones in bold to complete the statement about the relationship between work done and potential difference.
When a positively charged particle moves across a potential difference from a positive plate towards an earthed plate, work is done by the **field/charged particle** on the **field/charged particle**.

7 Calculate the work done to move a positively charged particle of 2.5×10^{-18} C a distance of 3.0 mm towards a positive plate in a uniform electric field of 556 $N C^{-1}$.

8 A particular electron gun accelerates an electron across a potential difference of 15 kV, a distance of 12 cm between a pair of charged plates. What is the magnitude of the force acting on the electron? (Use $q_e = 1.6 \times 10^{-19}$ C.)

9 A charge of $+q$ is placed a distance r from another charge also of $+q$. A repulsive force of magnitude F is found to exist between them. Choose the correct options from the ones in bold to describe the changes, if any, that will occur to the force in the following situations.
   a The distance between the charges is doubled to $2r$, so the force will **halve/double/quadruple/quarter** and **repel/attract**.
   b The distance between the charges is halved to $0.5r$, so the force will **halve/double/quadruple/quarter** and **repel/attract**.
   c The distance between the charges is doubled and one of the charges is changed to $-2q$, so the force will **halve/double/quadruple/quarter** and **repel/attract**.

10 A gold(III) ion is accelerated by the electric field created between two parallel plates separated by 0.020 m. The ion carries a charge of $+3e$ and has a mass of 3.27×10^{-25} kg. A potential difference of 1000 V is applied across the plates. The work done to move the ion from one plate to the other results in an increase in the kinetic energy of the gold(III) ion. If the ion starts from rest, calculate its final velocity. (Use $q_e = -1.602 \times 10^{-19}$ C.)

11 Calculate the magnitude of the force that would exist between two point charges of 5.00 mC and 4.00 nC separated by 2.00 m. (Use $k = 9 \times 10^9$ $N m^2 C^{-2}$.)

12 A point charge of 2.25 mC is positioned on top of an insulated rod on a table. At what distance above the point charge should a sphere of mass 3.00 kg containing a charge of 3.05 mC be located, so that it is suspended in the air? (Use $k = 9 \times 10^9$ $N m^2 C^{-2}$.)

13 A charged plastic ball of mass 5.00 g is placed in a uniform electric field pointing vertically upwards with a strength of 300.0 $N C^{-1}$. Calculate the magnitude and sign of the charge required on the ball in order to create a force upwards that exactly equals the weight force of the ball.

14 Two electrons approach each other at a distance of 5.4×10^{-12} m. The charge on each electron is -1.6×10^{-19} C. What is the electrical force of repulsion between the two electrons? (Use $k = 9 \times 10^9$ $N m^2 C^{-2}$.)

15 Two large horizontal metal plates are 4.50 cm apart and connected to a direct current (DC) power supply such that the upper plate is positive relative to the lower plate, which is earthed. The electric field between the plates is
$5.50 \times 10^5 \text{ V m}^{-1}$. An electron at rest is liberated from the lower plate.

- **a** Calculate the force exerted by the field on this electron.
- **b** What is the potential difference between the plates?
- **c** How much work is done by the field on the electron in moving it from the negative to the positive plate.

16 Two metal spheres, one carrying a charge of −16.0 μC and the other a charge of +5.00 μC, are attracted to each other and touch. A force of repulsion then exists between them and they move apart.

- **a** Calculate the force of attraction between them when they are 2.00 mm apart and before they touch.
- **b** What is the charge on each sphere after they touch?
- **c** What is the force between them after they move apart and are 6.75 mm apart?

17 A beam of electrons, initially at rest, is accelerated in the electron gun of a linear accelerator to a velocity of $5.30 \times 10^7 \text{ m s}^{-1}$.

- **a** How much work is done on each electron to achieve this velocity?
- **b** Through what voltage were they accelerated?

18 Two small charged spheres, each with a mass of 5.65 μg, are each attached to very thin nylon threads of length 50.0 cm and hung vertically from the same point. The same quantity of positive charge is then placed on each sphere, causing them to repel each other. The angle between the two threads is measured and found to be 14.0°. Calculate the force acting on each sphere.

CHAPTER

05 Magnetic field and force

Using magnetic fields to apply forces on charged particles and electric conductors has made possible innovations such as the electric motor and television. With the development of more sustainable ways to create and store electrical energy, the use of electric motors to drive our transport will become more important in the near future.

In this chapter you will investigate the concepts that apply to magnetic fields and its application to the development of the electric motor.

Science as a Human Endeavour

Electromagnetism is utilised in a range of technological applications, including:

- DC electric motor with commutator

Science Understanding

- the direction of conventional current is that in which the flow of positive charges takes place, while the electron flow is in the opposite direction
- current-carrying wires are surrounded by magnetic fields; these fields are utilised in solenoids and electromagnets
- the strength of the magnetic field produced by a current is a measure of the magnetic flux density

  This includes applying the relationship

  $$B = \frac{\mu_0}{2\pi}\frac{I}{r}$$
- magnets, magnetic materials, moving charges and current-carrying wires experience a force in a magnetic field when they cut flux lines; this force is utilised in DC electric motors and particle accelerators

  This includes applying the relationships

  $F = qvB$ where $v \perp B$

  $F = I\ell B$ where $\ell \perp B$
- the force due to a current in a magnetic field in a DC electric motor produces a torque on the coil in the motor

  This includes applying the relationship

  $\tau = r_{\perp} F$

5.1 The magnetic field

FIGURE 5.1.1 In 1820, Hans Christian Oersted discovered the magnetic effect created by an electric current. Oersted is honoured by this statue in Oersted's Park, Copenhagen.

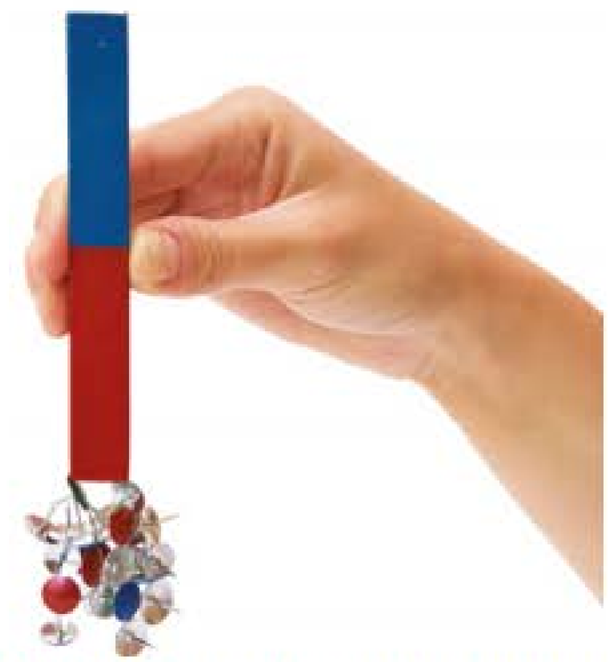


FIGURE 5.1.2 A bar magnet attracting drawing pins.

Although naturally occurring magnets had been known for many centuries, by the early 19th century there was still no scientifically proven way of creating an artificial magnet. In 1820, the Danish physicist Hans Christian Oersted (whose statue is pictured in Figure 5.1.1) developed a scientific explanation for the magnetic effect created by an electric current.

Oersted was a keen believer in the 'unity of nature', the concept that everything in the universe is somehow connected. He noticed that when he switched on a current from a **voltaic pile** (a simple early battery), a magnetic compass nearby moved. Intrigued by this observation, he carried out further experimentation, which demonstrated that it was the current from the voltaic pile that was affecting the movement of the compass. His experiments showed that the stronger the current, and the closer the compass was to it, the greater the observed effect. These observations led him to conclude that the electric current was creating a magnetic field. This connection between electric and magnetic fields is fundamental to society today.

MAGNETISM

Before looking further into the connection between electric current and **magnetic fields**, it is necessary to review some fundamentals of magnetism.

The magnetic effect most people are familiar with is the attraction of iron or other **magnetic** materials to a magnet (Figure 5.1.2).

But, if you experiment with a magnet yourself, you will find that each end of the magnet behaves differently, particularly when interacting with another magnet. One end will be attracted while the other is repelled. Each end of a magnet is referred to as a **magnetic pole**.

Like magnetic poles repel each other; unlike magnetic poles attract each other.

PHYSICSFILE

Dipoles

Try breaking a (cheap) magnet in half. All you get is two smaller magnets, each with its own north and south poles. No matter how many times you break the magnet and how small the pieces are, each will be a separate little magnet with two poles. Because magnets always have two poles, they are said to be *dipolar*.

Magnets are dipolar and a magnetic field is said to be a **dipole** field (Figure 5.1.3). This is similar to electric charges where a positive and negative charge in close proximity to each other are said to form a dipole. A key difference is that you cannot have a single magnetic pole (monopole), whether it be a south pole or a north pole; however, a charge *can* exist on its own as either a positive or negative charge.

FIGURE 5.1.3 Magnets are always dipolar.

A suspended magnet that is free to move will always orientate itself in a north–south direction. That's basically what the needle of a compass is—a freely suspended, small magnet. If allowed to swing in the vertical plane as well, then the magnet will tend to tilt up or down. The vertical direction (upwards/downwards) and the magnitude of the angle depend upon the distance of the magnet from either of the Earth's poles.

As you can see in Figure 5.1.4, the Earth itself can be shown to have a giant magnetic field around it.

The names for the poles of a magnet derive from early observations of magnets orientating themselves with the Earth's geographic poles.

Initially, the end of the magnet pointing toward the Earth's geographic north was denoted the north **pole**, and compasses are thus marked with this end as north. However, it is now known that the North Geographic Pole and the North Magnetic Pole are some distance apart, and the same applies to the South Geographic Pole and the South Magnetic Pole. By definition then, the Earth acts like a giant magnet with a south pole at what is called the North Magnetic Pole. The north end of a compass aligns itself along the field lines pointing to the south pole of the magnet within the Earth (Figure 5.1.4).

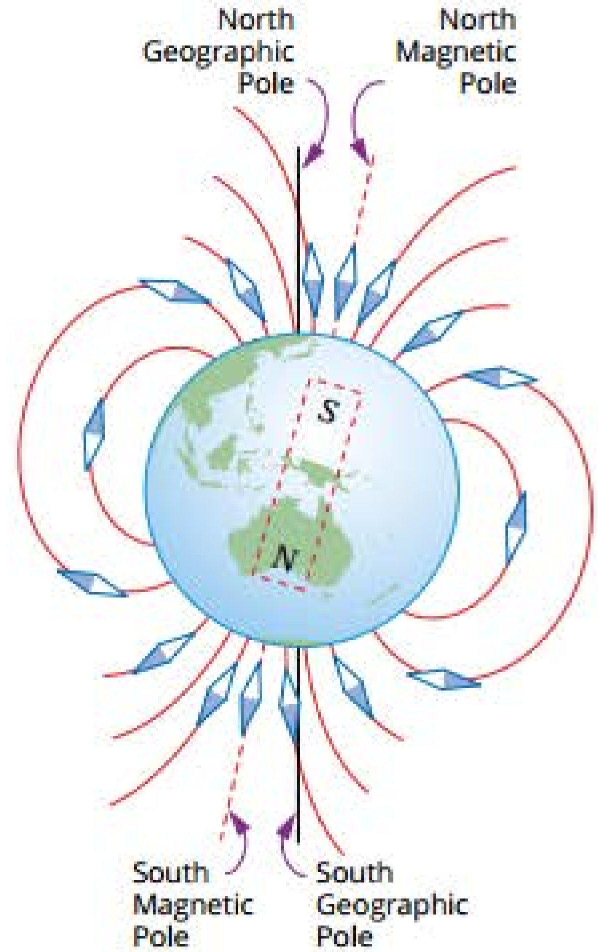


FIGURE 5.1.4 The Earth acts somewhat like a huge bar magnet. The south pole of this imaginary magnet is near the North Geographic Pole and is the point to which the north pole of a compass appears to point.

PHYSICSFILE

'Flipping' poles

The Earth's magnetic poles are not static like their geographic counterparts. For many years, the North Magnetic Pole had been measured as moving at about 9 km per year (Figure 5.1.5). In recent years that has accelerated to an average of 52 km per year. Once every few hundred thousand years the magnetic poles actually flip, in a phenomenon called 'geomagnetic reversal', so that a compass would point south instead of north. The Earth is well overdue for the next flip, and recent measurements have shown that the Earth's magnetic field is starting to weaken faster than in the past, so the magnetic poles may be getting close to a 'flip'. Although past studies have suggested that such a flip is not instantaneous—it would take many hundreds if not a few thousands of years—some more recent studies have suggested that it could happen over a significantly shorter time period.

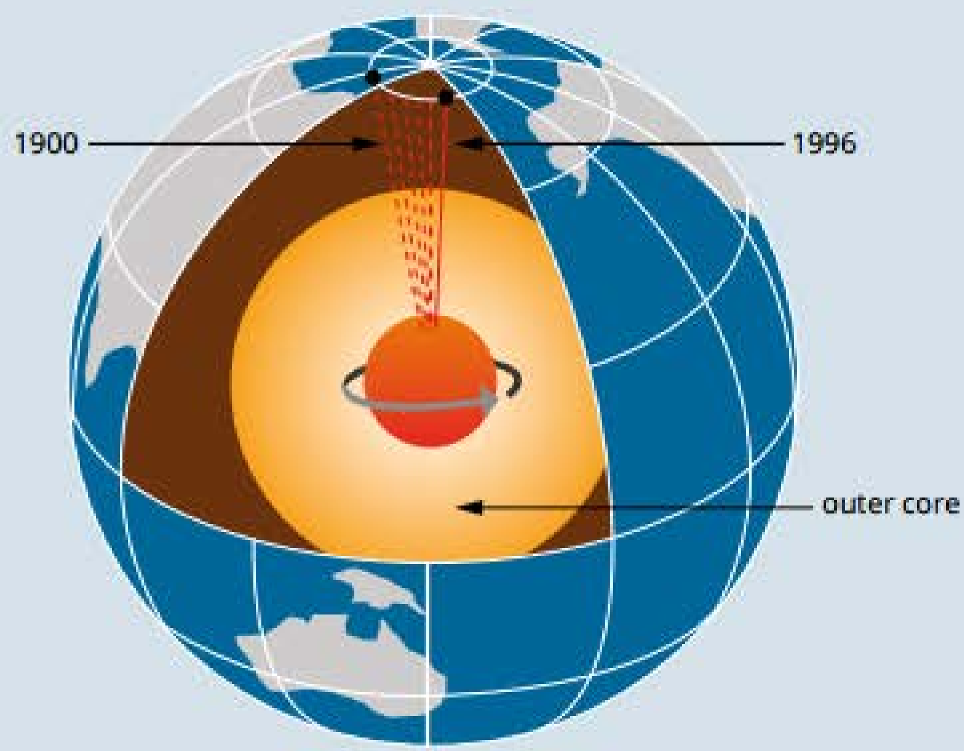


FIGURE 5.1.5 Diagram of Earth's interior and the movement of magnetic north from 1900 to 1996. The Earth's outer core is believed to be the source of the geomagnetic field.

MAGNETIC FIELD

In the previous chapter, you saw that point charges and charged objects produce an electric field in the space that surrounds them. For this reason, charged bodies within the field will experience a force. The direction of the electric field is determined by the direction of that force.

Magnets also create fields. If you do a simple test such as placing a pin near a magnet, you will observe that the pin will be pulled towards the magnet. This shows that the space around the magnet must therefore be affected by the magnet.

If you sprinkle iron filings on a piece of clear acetate that is held over a magnet, you will observe that the magnetic field will be clearly defined (Figure 5.1.6). The iron filings will line up with the field, showing clear field lines running from one end of the magnet to the other.

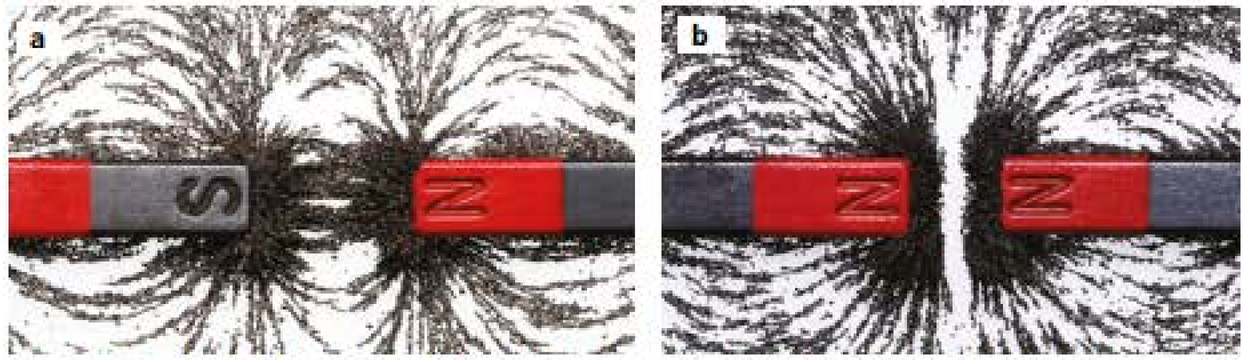


FIGURE 5.1.6 Iron fillings sprinkled around magnets (a) with unlike poles close together and (b) with like poles close together. The patterns in the fields show the attraction and repulsion between poles, respectively.

Vector field model for magnetic fields

The diagram in Figure 5.1.7 shows the magnetic field associated with a simple bar magnet. The magnetic field around the bar magnet can be defined in vector terms by specifying both direction and magnitude.

The direction of the magnetic field at any point is the direction in which a compass would point if placed at that point—that is, towards the south pole of the magnet and away from the north pole end. This is also the direction of the force the magnetic field would exert on the north pole end of a magnet. Note that magnetic field lines run north to south external to the magnet, and south to north inside the magnet. The magnetic field lines are continuous.

Denser (closer) lines indicate a relatively stronger magnetic field. As the distance from the magnet increases, the magnetic field is spread over a greater area and its strength, or magnetic flux density, at any point decreases. The strength and direction associated with the magnetic field at any point signifies that it is a vector quantity. The flux density, or vector magnitude, of the magnetic field at a particular point is denoted by B and has units of tesla (T).

The fields between magnets are dependent on whether like or unlike poles are close together, the distance between the poles, and the relative strength of the magnetic field of each magnet. Iron filings or small plotting compasses can be used to visualise the field between and around the magnets, as shown in Figure 5.1.8.

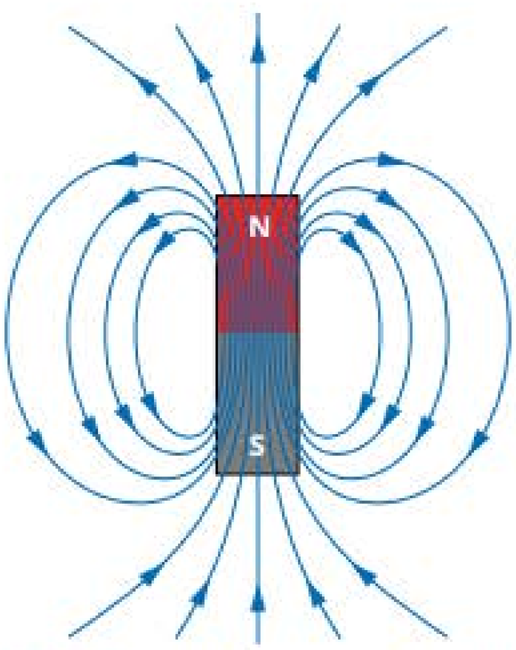


FIGURE 5.1.7 The field lines around and inside a bar magnet. The lines show the direction of the force on the north pole end of a magnet.

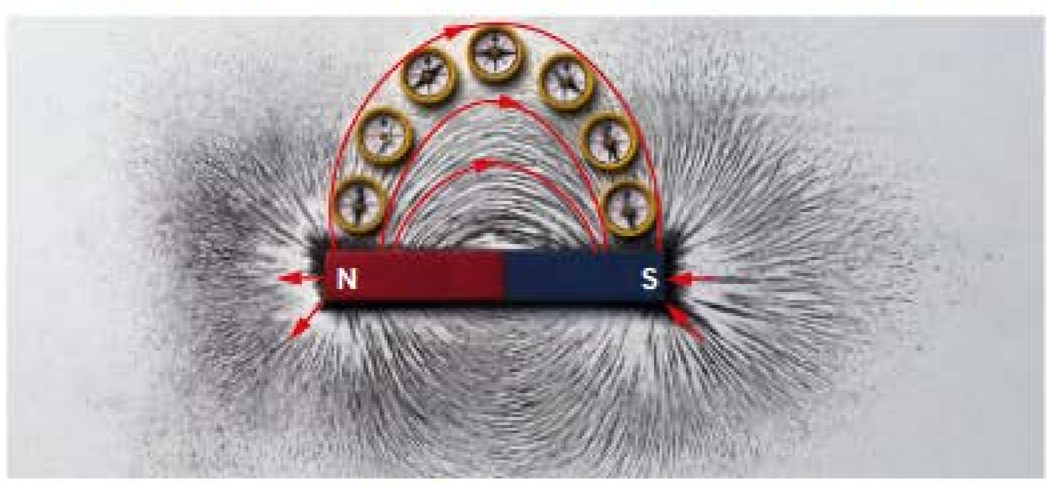


FIGURE 5.1.8 Plotting field lines around a bar magnet. Small plotting compasses are placed around the magnet. Field lines are drawn linking the direction each compass points in, creating field lines that run from the north pole to the south pole of the magnet.

Because the Earth has a giant magnetic field around it, compasses will orientate themselves along the magnetic field lines. In Figure 5.1.8, note the direction of the magnetic field close to either pole, where the magnetic field lines run almost vertically. Compasses placed near the Earth's magnetic poles will behave in the same way.

Different shaped magnets produce different shaped fields. The diagram in Figure 5.1.9 shows the magnetic field plotted for a horseshoe magnet.

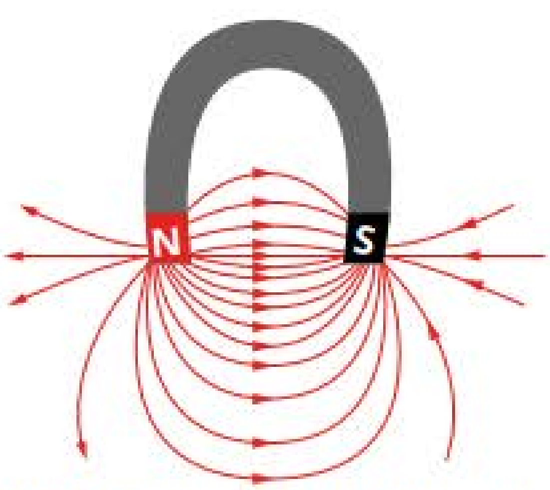


FIGURE 5.1.9 The horseshoe magnet has two unlike poles close to each other. This creates a very strong magnetic field.

The resultant direction of the magnetic field at a particular point will be the vector addition of each individual magnetic field acting at that point.

When two magnets are placed close together, two situations may arise. If the poles are unlike, as per Figure 5.1.10(b), then attraction will occur between them and a magnetic field will be created that extends between the two poles. However, if like poles are very near each other as per Figure 5.1.10(c), repulsion will occur. In this situation, there will be a neutral point between the two poles at which there is no magnetic field.

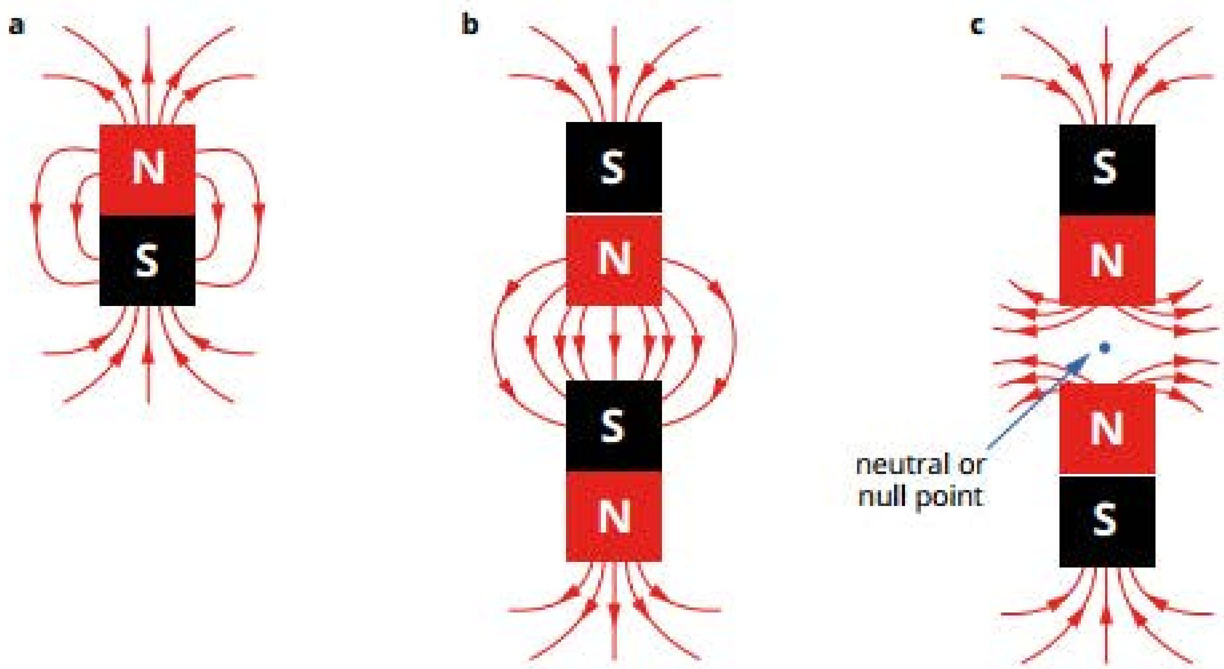


FIGURE 5.1.10 Magnetic field lines plotted for (a) a bar magnet, (b) opposite poles of magnets in close proximity and (c) like poles of magnets in close proximity.

As the bar magnets in Figure 5.1.10 have a fixed strength and position, the associated magnetic fields will be static. Varying the magnetic flux density, by changing the magnets or varying the relative position of the magnets, would produce a changing magnetic field.

MAGNETIC FIELDS AND CURRENT-CARRYING WIRES

In the introduction to this section the connection between electric current and magnetic fields was noted. Oersted found that when he switched on the current from a voltaic pile, a nearby magnetic compass would move. It is believed that the Earth's magnetic field is created by a similar effect—circulating electric currents in the Earth's molten metallic core.

A circular magnetic field is created around a current-carrying wire. This can be seen in Figure 5.1.11. A compass aligns itself at a tangent to the concentric circles around the wire (i.e. the magnetic field). The stronger the current and the closer the compass is to the wire, the greater the effect.

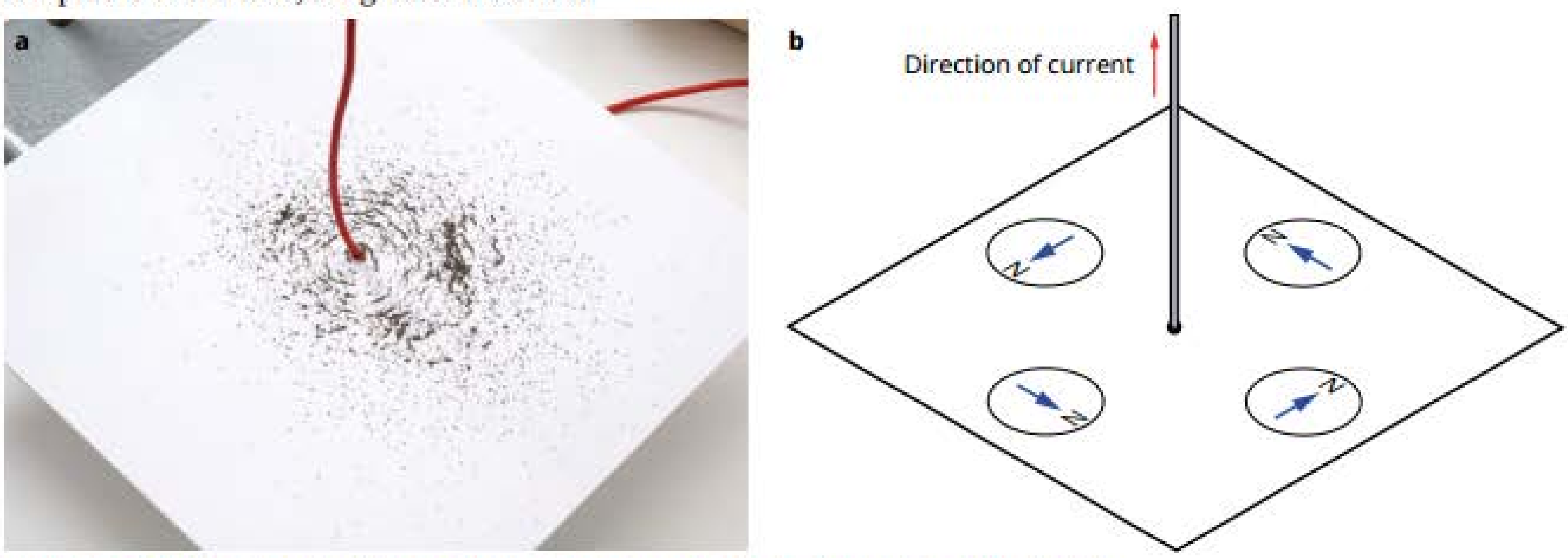


FIGURE 5.1.11 (a) The magnetic field around a current-carrying wire. The iron filings align with the field to show the circular nature of the magnetic field. (b) Small compasses will indicate the direction of the field.

The magnetic field is perpendicular to the current-carrying wire and the direction of the field will depend upon the current direction. There's a simple and easy way to determine the direction of the magnetic field, which is commonly referred to as the **right-hand grip rule**.

Grasp the conducting wire with your right hand with your thumb pointing in the direction of the conventional electric current, I (positive flow). Curl your fingers around the wire. The magnetic field will be perpendicular to the wire and in the direction your fingers are pointing, as shown in Figure 5.1.12.

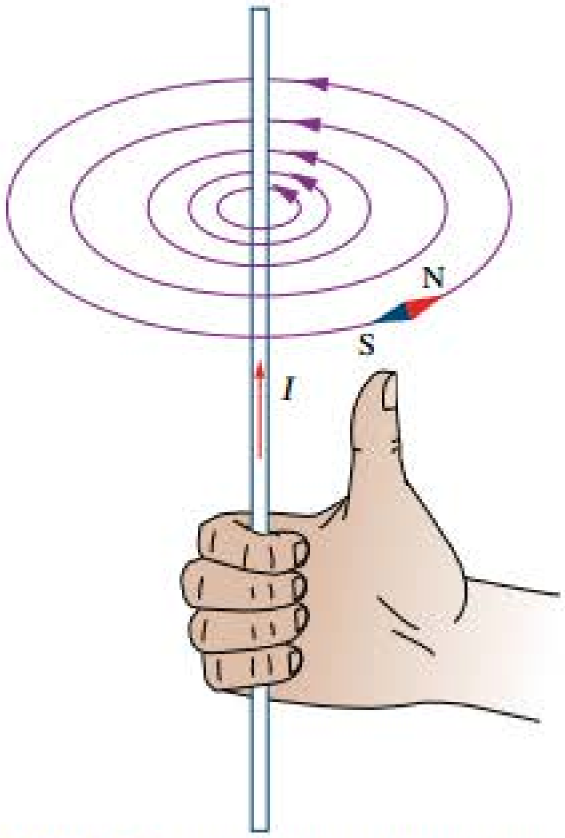


FIGURE 5.1.12 The right-hand grip rule can be used to find the direction of the magnetic field around a current-carrying wire, when the direction of the conventional current, I, is known.

Worked example 5.1.1

DIRECTION OF THE MAGNETIC FIELD

A current-carrying wire runs horizontally across a table. The conventional current direction, I, is running from left to right. What is the direction of the magnetic field created by the current?	
Thinking	**Working**
Recall that the right-hand grip rule indicates the direction of the magnetic field.	Hold your hand with your fingers aligned as if gripping the wire. Point your thumb to the right in the direction of the current flow. 
Describe the direction of the field in relation to the reference object or wire in simple terms, so that the description can be readily understood.	The magnetic field direction is perpendicular to the wire and hence it is circular. It runs from up the back of the wire, over the top towards the front of the wire.

Worked example: Try yourself 5.1.1

DIRECTION OF THE MAGNETIC FIELD

A current-carrying wire runs along the length of a table. The conventional current direction, I, is running towards an observer standing at the near end. What is the direction of the magnetic field created by the current as seen by the observer?

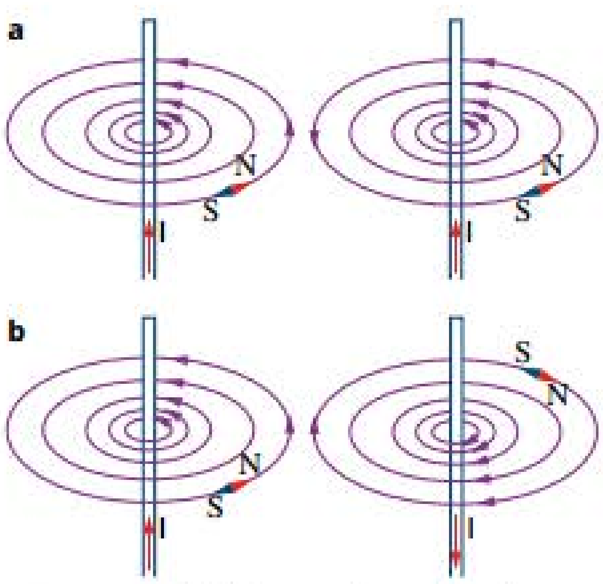


FIGURE 5.1.13 (a) Two current-carrying wires attract when current runs through them in the same direction. This is because the magnetic fields between the wires are in opposite directions. (b) Two current-carrying wires repel when the current passes through them in opposite directions. This is because the magnetic fields are in the same direction.

Magnetic fields between parallel wires

Two current-carrying wires arranged parallel to each other will each have its own magnetic field. The direction of the magnetic field around each wire is given by the right-hand grip rule. If the two wires are brought closer together, their associated magnetic fields will interact, just as any two regular magnets would interact. The interaction could result in either an attraction or repulsion of the wires, depending on the direction of the magnetic fields between them (Figure 5.1.13). When the magnetic fields are in the opposite directions, this represents unlike poles, and so the wires attract. When the magnetic fields are in the same direction, the wires repel.

THREE-DIMENSIONAL FIELDS

Field lines can also be drawn for more-complex, three-dimensional fields such as that around the Earth or those around current-carrying loops and coils. Even in more-complex fields, the right-hand grip rule is still applicable, as you can see in Figures 5.1.14 and 5.1.15.

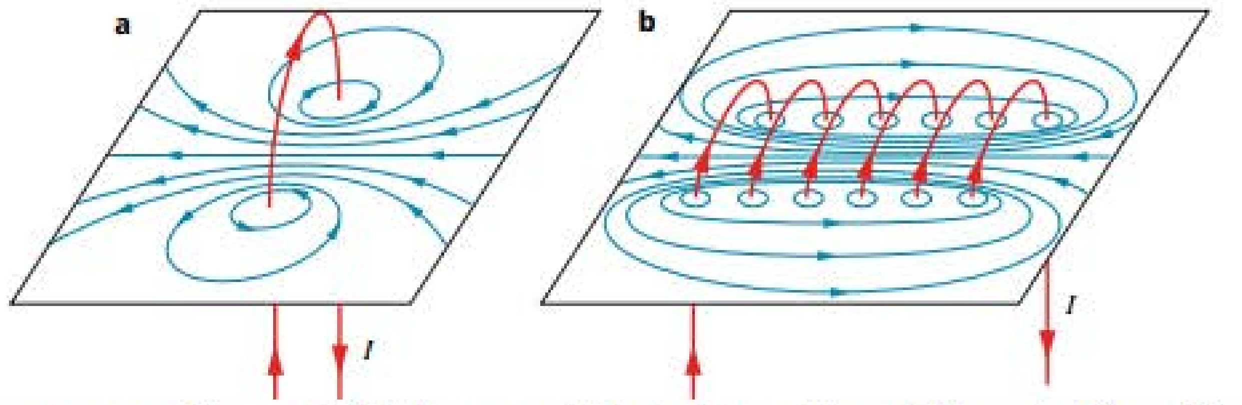


FIGURE 5.1.14 The magnetic field lines around (a) a single current loop and (b) a series of loops. The blue arrows indicate the direction of the magnetic field. The more concentrated the lines are inside the loops, the stronger the magnetic field is in this region.

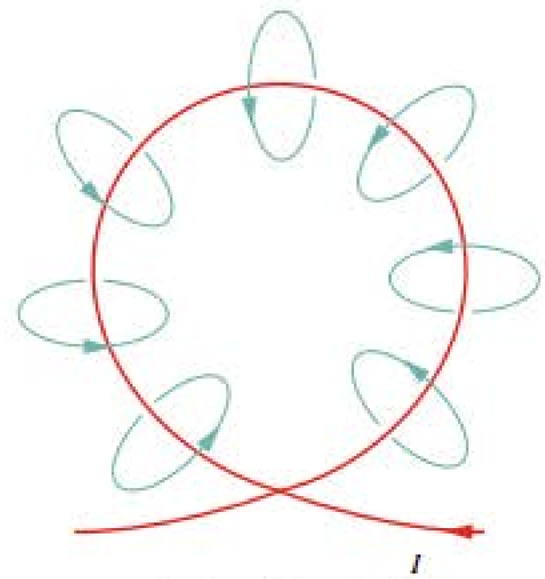


FIGURE 5.1.15 A three-dimensional representation of the magnetic field around a loop of wire in the plane of the page. The blue arrows show the direction of the magnetic field. Notice that the magnetic field is a circular shape, with no field lines crossing.

The direction of a magnetic field can be shown with a simple arrow on a field line when the field is travelling within the plane of a page (as shown in Figure 5.1.16), or a simple three-dimensional depiction can be used as in Figures 5.1.14 and 5.1.15.

When a field is running directly into or out of the plane of a page, dots are used to show a field coming out of the page and crosses are used to depict a field running directly into the page. This convention was adopted from the idea of viewing an arrow. The dot is the point of the arrow coming towards you, and the cross represents the tail feathers as the arrow travels away.

Figure 5.1.16 shows the two-dimensional representation of the magnetic field around the simple loop of wire shown in Figure 5.1.15.

The strength of a field is depicted by varying the density of the lines or dots and crosses. Showing lines coming closer together indicates a strengthening of a field, lower density indicates a weaker field. More densely placed dots or crosses can also show a stronger area of the field. Figure 5.1.16 shows a non-uniform magnetic field with a weaker magnetic field in the centre of the coil and a stronger magnetic field near the wire.

As the magnetic fields associated with current-carrying coils are dependent upon the size of the current, a changing current will produce a changing magnetic field.

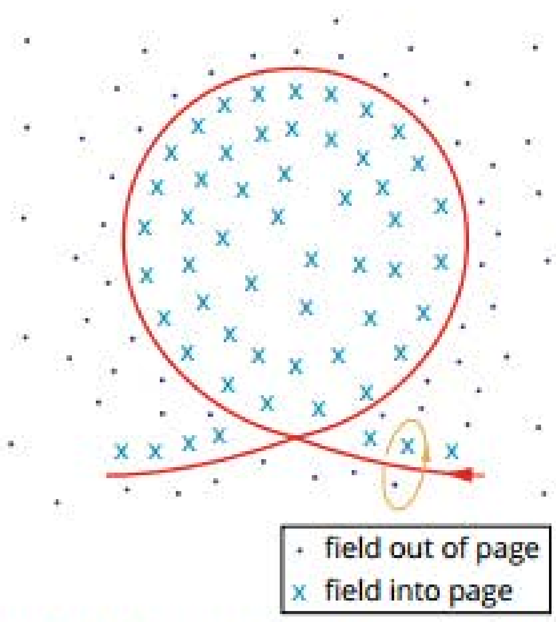


FIGURE 5.1.16 A two-dimensional representation of the same current-carrying loop depicted in Figure 5.1.15. Areas where the magnetic field is stronger are shown with a greater density of dots and crosses.

The magnetic field around a solenoid

If many loops are placed side by side, their fields all add together and there is a much stronger effect. This can easily be achieved by winding many turns of wire into a coil, termed a **solenoid**, as shown in Figure 5.1.14(b). The field around the solenoid is like the field around a normal bar magnet. The direction of the overall magnetic field can be determined by considering the direction of the magnetic field around each current-carrying loop. Using the right-hand grip rule for one of the wires above the plane in Figure 5.1.14(b) indicates that the magnetic field inside the loop points to the left. This is the north end of the solenoid. Each coil will reinforce the field.

This is explained in the diagram in Figure 5.1.17.

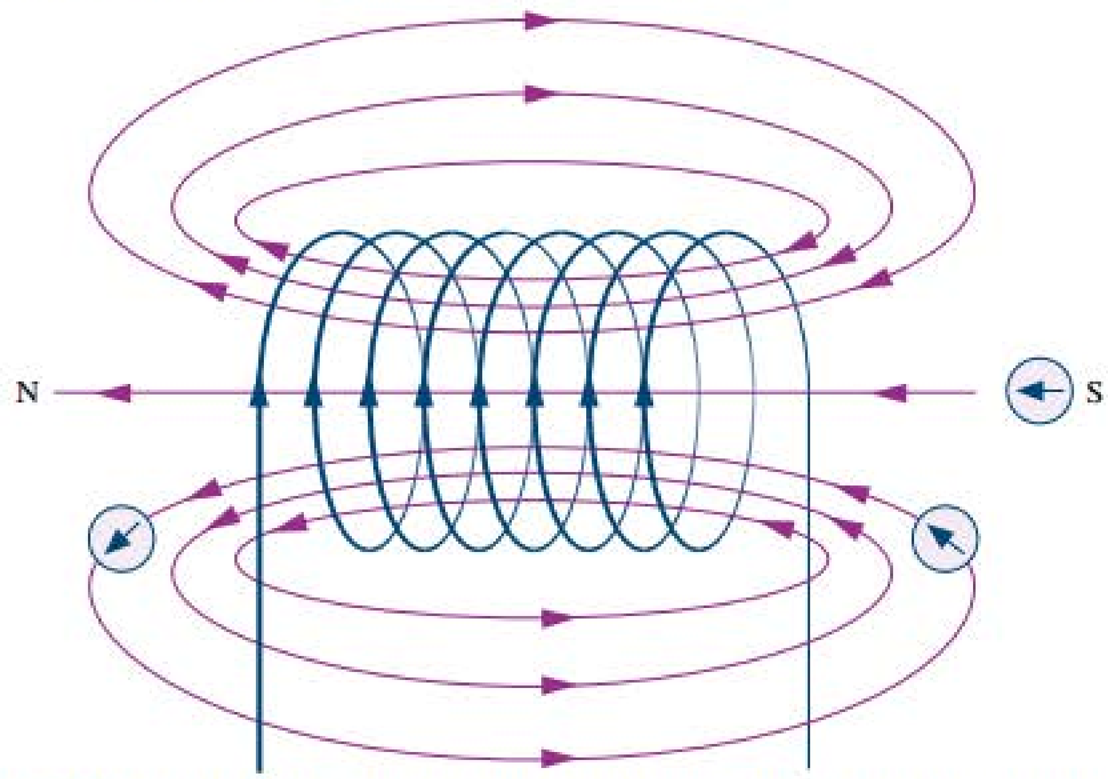


FIGURE 5.1.17 This solenoid has an effective 'north' end at the left and a 'south' end at the right. The compass points in the direction of the field lines.

The direction of north in a solenoid can also be determined using the right-hand grip rule in reverse. Curl your fingers in the direction of the conventional current in the wires; your thumb will point to north pole of the solenoid.

PHYSICS IN ACTION

Creating an electromagnet

The earliest magnets were all naturally occurring. If you wanted a magnet, you needed to find one. They were regarded largely as curiosities. Hans Christian Oersted's discoveries made it possible to manufacture magnets, making the widescale use of magnets possible.

An **electromagnet**, as the name infers, runs on electricity. It works because an electric current produces a magnetic field around a current-carrying wire. If the conductor is looped into a series of coils to make a solenoid, then the magnetic field can be concentrated within the coils. The more coils, the stronger the magnetic field and, therefore, the stronger the electromagnet.

The magnetic field can be strengthened further by wrapping the coils around a core. Normally, the atoms in materials like iron are aligned in random directions and the individual magnetic fields tend to cancel each other out. However, the magnetic field produced by coils wrapped around an iron core can force the atoms within the core to align in one direction. Their individual magnetic fields add together, creating a stronger magnetic field.

The magnetic flux density of an electromagnet can also be changed by varying the amount of electric current that flows through it.

The direction of the current creates poles in the electromagnet. The poles of an electromagnet can be reversed by reversing the direction of the electric current.

Today, electromagnets are used directly to lift heavy objects (Figure 5.1.18), as switches and relays, and as a way of creating new permanent magnets by aligning the atoms within magnetic materials.

FIGURE 5.1.18 A large electromagnet is used to lift waste iron and steel at a scrapyard. Valuable metals such as these are separated and then recycled.

CALCULATING THE STRENGTH OF THE MAGNETIC FIELD AROUND A WIRE

Wires running between transmission towers, such as those shown in Figure 5.1.19, will generate a magnetic field around them as long as current is flowing

FIGURE 5.1.19 Each one of the current-carrying wires running between the transmission towers will generate a magnetic field around it, as long as a current is flowing.

Like most fields, the closer an object is to a current-carrying wire, the stronger the magnetic field. It is also observed that the amount of current in a wire affects the strength of the magnetic field. The magnetic flux density (B) is the term used to describe the strength of a magnetic field and is measured in the unit tesla (T).

In fact, the flux density of the magnetic field, B, around a wire is directly proportional to the current, I, in a long straight conductor and inversely proportional to the distance from the conductor, r.

$$B = \frac{\mu_0 I}{2\pi r}$$

where B is the flux density of the magnetic field in tesla (T)

I is the current in the conductor in ampere (A)

r is the distance from the conductor in metre (m)

μ_0 is a constant referred to as the 'permeability of free space', $\mu_0 = 4\pi \times 10^{-7}\ \text{T m A}^{-1}$

B is a vector quantity, the direction being that of the magnetic field around the conductor

Worked example 5.1.2

FLUX DENSITY OF A MAGNETIC FIELD

A current of 5.00 A is running through a long straight wire. Determine the flux density of the magnetic field at a point 10.0 cm from the wire.

Thinking	Working
Identify the quantities supplied in the question and the quantity required. Convert units to SI.	$B = ?$ $I = 5.00\,\text{A}$ $r = 10.0\,\text{cm} = 0.100\,\text{m}$ $\mu_0 = 4\pi \times 10^{-7}\,\text{T m A}^{-1}$
Substitute quantities into the equation for the flux density of a magnetic field around a wire.	$B = \frac{\mu_0 I}{2\pi r}$ $= \frac{4\pi \times 10^{-7} \times 5.00}{2\pi \times 0.100}$
Simplify to find *B*.	$B = \frac{2.00 \times 10^{-7} \times 5.00}{0.100}$ $= 1.00 \times 10^{-5}\,\text{T}$
Note that only the magnitude is required. State your answer as your calculated result only. Do not include the direction.	$B = 1.00 \times 10^{-5}\,\text{T}$

Worked example: Try yourself 5.1.2

FLUX DENSITY OF A MAGNETIC FIELD

A current of 4.00 A is running through a long straight wire. Determine the flux density of the magnetic field at a point 3.00 cm from the wire.

PHYSICSFILE

The tesla

The unit for the flux density of a magnetic field, *B*, was given the name tesla (T) in honour of Nikola Tesla. Nikola Tesla (1856–1943) was the first person to advocate the use of alternating current (AC) generators for use in town power-supply systems. He was also a prolific inventor of electrical machines of all sorts, including the Tesla coil, a source of high-frequency, high-voltage electricity.

A magnetic flux density of 1 T is a very strong field. For this reason, a number of smaller units, especially the millitesla (mT), 10^{-3} T, and microtesla (μT), 10^{-6} T, are in common use. The table below shows the strength of some magnets for comparison.

TABLE 5.1.1 Comparison of magnet strengths

Type of magnet	Magnetic flux density (*B*)
very strong electromagnets and 'supermagnets'	1 to 20 T
Alnico and ferrite magnets	10^{-2} to 1 T
Earth's surface	5×10^{-5} T

5.1 Review

SUMMARY

- Like magnetic poles repel, and unlike magnetic poles attract.
- Magnetic poles exist only as dipoles, having both north and south poles. A single magnetic pole (monopole) is not known to exist.
- Magnetic field lines run from the north pole to the south pole outside a magnet or solenoid, and from south to north inside the magnet or solenoid.
- The direction of a magnetic field at a particular point is the same as that of the force on the north pole end of a magnet.
- Magnetic field lines are continuous, enter and exit surfaces at 90°, and do not overlap.
- The resultant direction of interacting magnetic fields at any particular point will be the vector addition of each individual magnetic field acting at that point.
- The Earth has a dipolar magnetic field that acts as a huge bar magnet, with the south end of the Earth's magnet at the North Magnetic Pole and the north end of the Earth's magnet at the South Magnetic Pole.
- A magnetic field associated with a constant current in a wire is static. Where the current is changing, such as that associated with an alternating current direction, the magnetic field will also be changing.
- A uniform distribution of parallel field lines represents a uniform magnetic field. A non-uniform field, such as the radiating field lines around the poles of a magnet, is shown by variations in the separation of the field lines.
- An electrical current produces a magnetic field that is circular around a current-carrying conductor. The direction of the field is given by using the *right-hand grip rule* when considering the direction of the conventional current.
- More complex fields can be determined by applying the right-hand grip rule to the loops or coils making up the current-carrying conductor in a solenoid.
- The strength of the magnetic field, called the magnetic flux density B, is directly proportional to the current, I, in a long straight conductor and inversely proportional to the distance from the conductor, r. That is, $B = \frac{\mu_0 I}{2\pi r}$ where:

  B is the flux density of the magnetic field in tesla (T)

  I is the current in the conductor in ampere (A)

  r is the distance from the conductor in metres (m)

  μ_0 is a constant referred to as the 'permeability of free space', $\mu_0 = 4\pi \times 10^{-7}\,\text{T m A}^{-1}$.
- The unit for magnetic flux density, B, is the tesla (T). A field of 1.0 T is a very strong field.

KEY QUESTIONS

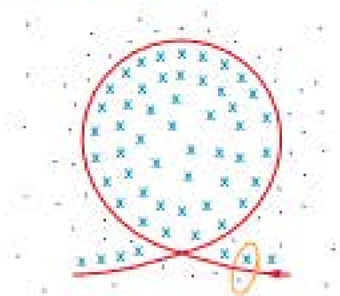
1 Repeatedly cutting a magnet in half always produces magnets with two opposite poles. From this information, which of the following can be deduced in relation to the poles of a magnet?

A Magnets are easily sliced in half.
B All magnets are dipolar.
C When the magnets are cut, the poles are split in half.
D All split magnets are monopolar.

2 A magnet is suspended on a thin wire at its midpoint so that it is free to swing. In which direction, approximately, will the north pole of the magnet point?

A along the field lines towards Earth's North Magnetic Pole
B along the field lines towards Earth's South Magnetic Pole
C along the field lines towards the equator
D to the sky, due to Earth's gravity

3 The magnetic field around a particular current-carrying loop shows a variation in magnetic flux density, as depicted below. The current in the loop (shown in red) is being switched on and off but is constant in direction and size. The magnetic field directions are shown in blue.

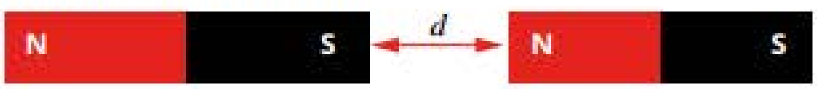


Which of the following best describes the resulting magnetic field of the loop?

A a static uniform magnetic field
B a static non-uniform magnetic field
C a variable non-uniform magnetic field
D a variable uniform magnetic field

5.1 Review *continued*

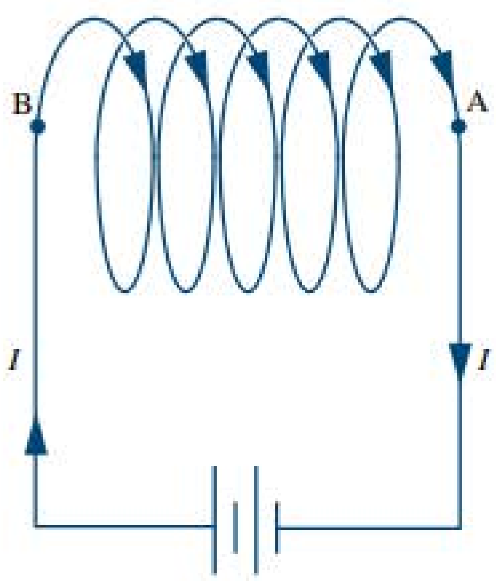
4 The following diagram shows two bar magnets separated by a distance *d*. At this separation, the magnitude of the magnetic force between the poles is equal to *F*. Which of the following is true if the distance, *d*, is increased?



A An attractive force greater than *F* will exist between the poles.
B A repulsive force greater than *F* will exist between the poles.
C An attractive force less than *F* will exist between the poles.
D A repulsive force less than *F* will exist between the poles.

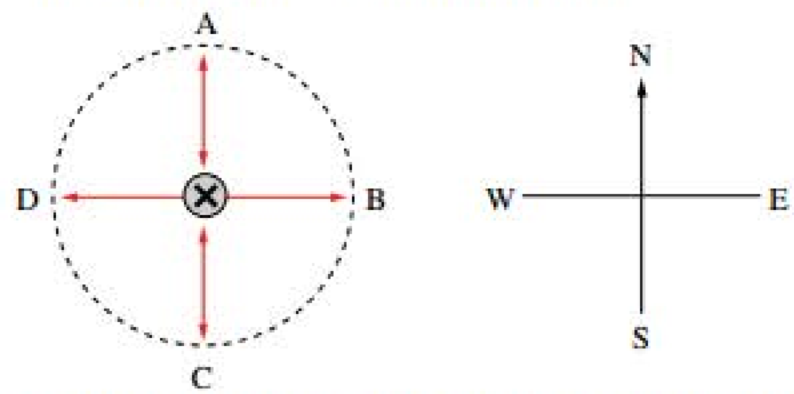
5 A current-carrying wire runs horizontally across a table. The conventional current direction, *I*, is running from right to left. Draw a diagram showing the direction of the magnetic field around the wire.

6 The following diagram shows a current-carrying solenoid.



Which end (A or B) represents the north pole of this solenoid?

7 The figure below shows a cross-sectional view of a long, straight current-carrying conductor, with its axis perpendicular to the plane of the page. The conductor carries an electric current into the page.



a What is the direction of the magnetic field produced by this conductor at each of the points A, B, C and D?

b The direction of the current in the conductor is now changed so that it is carried out of the page. What is the direction of the magnetic field produced by this conductor at the four points A, B, C and D?

8 A magnetic flux density, *B*, of 1.50×10^{-5} T is measured at a distance of 3.00 cm away from a long, straight wire. Calculate the current flowing through the wire that would create this magnetic field.

9 A magnetic flux density, *B*, of 0.0500 mT is measured at a quarter of a metre distance from a long, straight wire. Calculate the current flowing through the wire that would create this magnetic field. Give your answer to three significant figures.

10 A magnetic flux density, *B*, of 1.00 μT is measured at a distance *r* away from a long, straight wire. If the current, *I*, is measured as being 95.7 mA, what is the distance from the wire in centimetres?

5.2 Forces on charged objects

An electric current is a flow of electric charges. These may be electrons in a metal wire, electrons and mercury ions in a fluorescent tube, or cations and anions in an electrolytic cell. The nature of the flowing charge that makes up the current does not matter. A magnetic field is produced around the flow of charge, and a force is experienced within this field (Figure 5.2.1). In each case, it is the total rate of flow of charge (i.e. the current) that determines the field produced or the magnitude of the force.

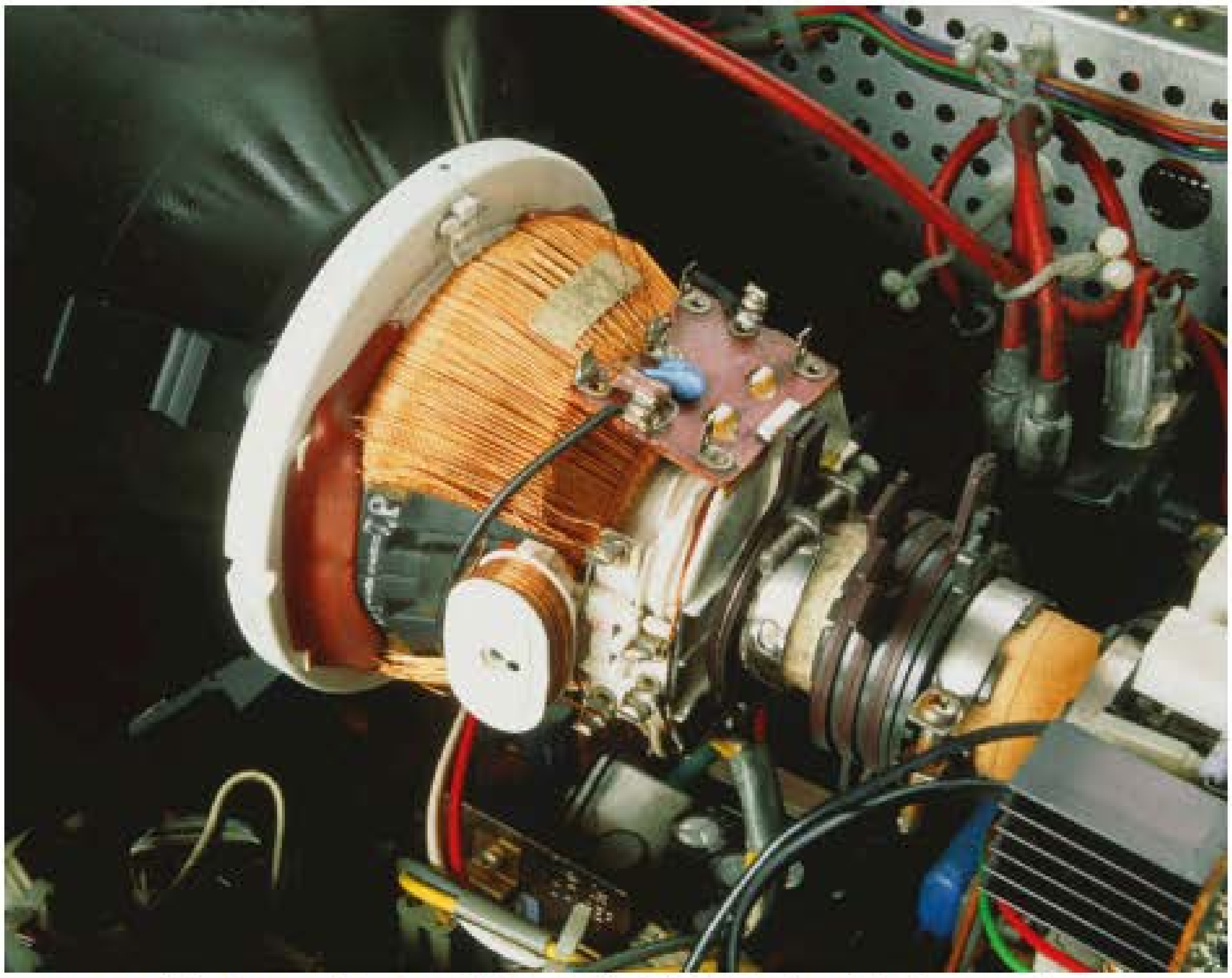


FIGURE 5.2.1 Electrons rushing down the length of a CRT (cathode ray tube) were the basis upon which old-style television sets worked. The electrons were deflected by the magnetic force they experienced as they passed through the 'yoke'—coils of copper wire at the back of the tube creating a strong variable magnetic field.

MAGNETIC FORCE ON CHARGED PARTICLES

The principle behind a **cathode ray tube** (CRT) is that a charged particle moving within a magnetic field will experience a force. In Figure 5.2.2, a beam of electrons in a CRT is experiencing a force due to a magnetic field. The force causes the beam of electrons to bend. The magnitude of the force is proportional to the flux density of the magnetic field, B, the component of the velocity of the charge that is perpendicular (at right angles) to the magnetic field and the charge on the particle.

FIGURE 5.2.2 The electron beam of a cathode ray tube is deflected by a magnet in the foreground.

When v and B are perpendicular:

$F = qvB$

where F is the force in newtons (N)

q is the electric charge on the particle in coulombs (C)

v is the component of the instantaneous velocity of the particle that is perpendicular to the magnetic field ($m\,s^{-1}$)

B is the flux density of the magnetic field (T)

This force is referred to as the **Lorentz force**. The force is at a maximum when the charged particle is moving at right angles to the field. There is no force acting when the charged particle is travelling parallel to the magnetic field.

Determining the direction of the force

The simple **mnemonic** shown in Figure 5.2.3, called the **right-hand palm rule**, can be used to determine the direction of the force on a positively charged particle moving in a magnetic field. Using your right hand, with fingers outstretched and flat, point the thumb toward the direction that a positive charge is moving and the outstretched fingers in the direction of the external magnetic field. The direction of the resulting force on the charge is the direction in which your palm is pointing. The force on a negatively charged particle will therefore be in the *opposite* direction to that on a positively charged particle.

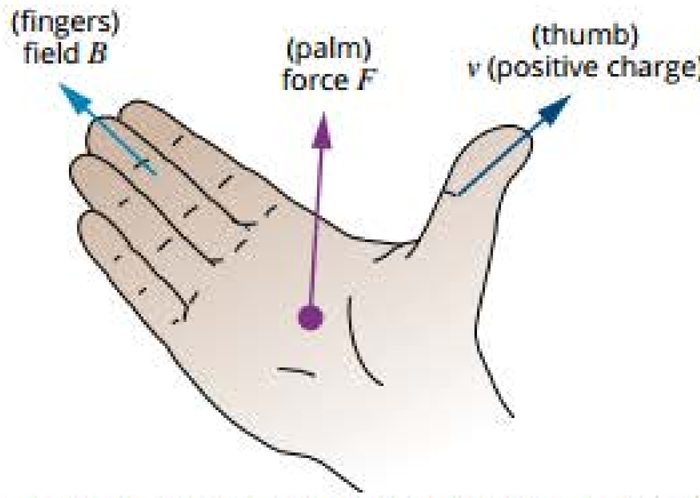


FIGURE 5.2.3 The right-hand palm rule: Point the thumb of the right hand in the direction of the movement of *a positive charge* (conventional current direction) and the fingers in the direction of the magnetic field. The force on the charge will point out from the palm.

The external magnetic field will exert a force on the charged particle that is perpendicular to the direction of its velocity. These are the conditions for centripetal force, which results in the particle moving in a circular path.

Therefore the magnetic force ($F_B = qvB$) provides the centripetal force $F_c = \frac{mv^2}{r}$ on the motion of the charged particle.

This is particularly relevant in accelerators, discussed more fully in Chapter 9.

Worked example 5.2.1

MAGNITUDE OF FORCE ON A POSITIVELY CHARGED PARTICLE

A single, positively charged particle with a charge of $+1.6 \times 10^{-19}$ C travels at a velocity of $10\,\text{m s}^{-1}$ perpendicular to a magnetic field, B, of flux density 4.0×10^{-5} T.

What is the magnitude of the force the particle will experience from the magnetic field?

Thinking	Working
Check the direction of the velocity and determine whether a force will apply. Forces only apply on the component of the velocity perpendicular to the magnetic field.	The particle is moving perpendicular to the field, so a force will apply. $F = qvB$
Establish which quantities are known and which quantities are required.	$F = ?$ $q = +1.6 \times 10^{-19}$ C $v = 10\,\text{m s}^{-1}$ $B = 4.0 \times 10^{-5}$ T
Substitute values into the force equation.	$F = qvB$ $= 1.6 \times 10^{-19} \times 10 \times 4.0 \times 10^{-5}$
Express the final answer in an appropriate form. Note that only magnitude has been requested, so do not include direction.	$F = 6.4 \times 10^{-23}$ N

Worked example: Try yourself 5.2.1

MAGNITUDE OF FORCE ON A POSITIVELY CHARGED PARTICLE

A single, positively charged particle with a charge of $+1.6 \times 10^{-19}$ C travels at a velocity of $50\,m\,s^{-1}$ perpendicular to a magnetic field, B, of flux density 6.0×10^{-5} T.

What is the magnitude of the force the particle will experience from the magnetic field?

EXTENSION

Objects moving at an angle to the magnetic field

The force experienced by a charge moving in a magnetic field is a vector quantity. The original expression noted above applies only to that component of the velocity of the charge perpendicular to the magnetic field. To find the force acting on an object moving at an angle θ to the magnetic field, use:

$$F = qvB\sin\theta$$

A charged particle travelling at a steady speed in a magnetic field experiences this force at an angle to its path and will be diverted. This is the theory behind CRT screens. As the direction of the charged particle changes, so does the angle of the force acting on it. In a very large magnetic field, the charged particles will move in a circular path. Mass spectrometers and particle accelerators both work on this principle.

The Sun ejects a continuous stream of charged particles consisting mainly of protons and electrons. This stream is referred to as the solar wind. The particles have enough energy to escape the Sun's gravitational field, at speeds ranging from about 300 to $800\,km\,s^{-1}$, allowing the solar wind to reach the Earth in about 3.9 days. When these high-energy charged particles meet the Earth's magnetic field, they experience a magnetic force and are deflected in such a way that they spiral towards the poles (Figure 5.2.4(a)), losing much of their energy and creating the auroras (the southern aurora, or aurora australis, and the northern aurora, or aurora borealis, as shown in Figure 5.2.4(b)).

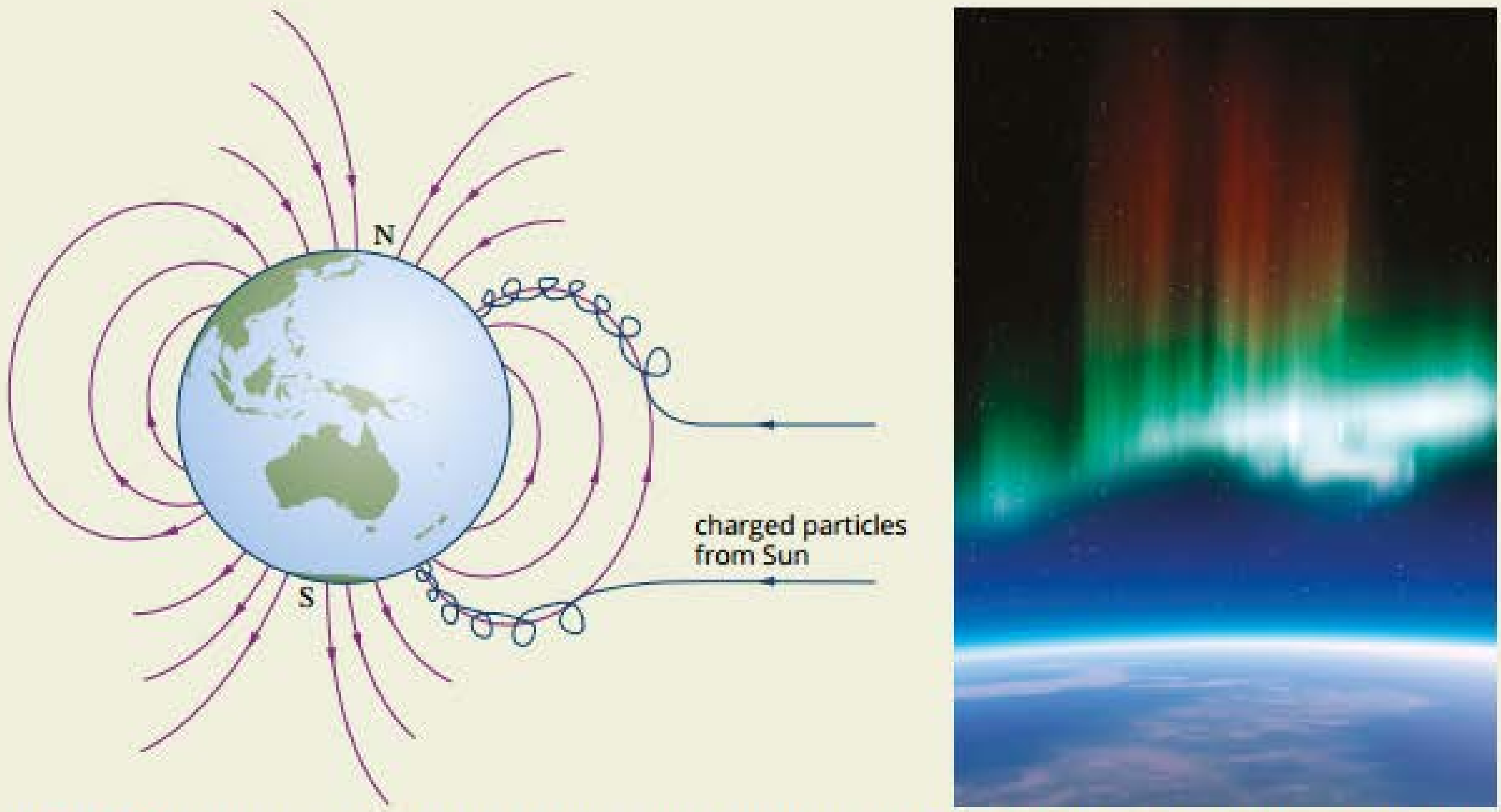
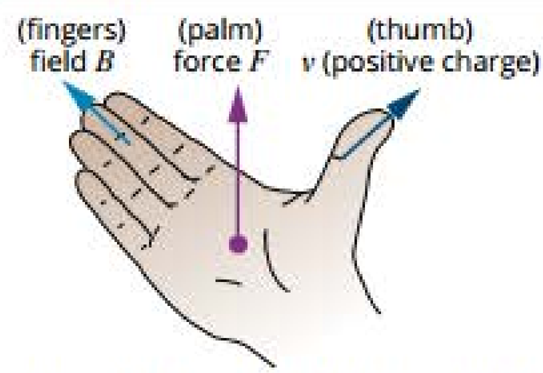


FIGURE 5.2.4 (a) Charged particles from the Sun or deep space are trapped by the Earth's magnetic field, causing them to spiral towards the poles. (b) As they do this, the particles lose energy and create the auroras.

Worked example 5.2.2

DIRECTION OF FORCE ON A NEGATIVELY CHARGED PARTICLE

A single, negatively charged particle with a charge of -1.6×10^{-19} C is travelling horizontally out of a computer screen and perpendicular to a magnetic field, B, that runs horizontally from left to right across the screen. In what direction will the force experienced by the charge act?	
Thinking	**Working**
 The right-hand palm rule is used to determine the direction of the force on a positively charged particle.	Align your hand so that your fingers are pointing in the direction of the magnetic field, i.e. left to right and horizontal. If the negatively charged particle is travelling out of the screen, a positively charged particle would be moving in the opposite direction. Align your thumb so it is pointing into the screen, in the direction in which a positive charge would travel. Your palm should be facing downwards. That is the direction of the force applied by the magnetic field on the negative charge out of the screen.

Worked example: Try yourself 5.2.2

DIRECTION OF FORCE ON A NEGATIVELY CHARGED PARTICLE

A single, negatively charged particle with a charge of -1.6×10^{-19} C is travelling horizontally from left to right across a computer screen and perpendicular to a magnetic field, B, that runs vertically down the screen. In what direction will the force experienced by the charge act?

PHYSICS IN ACTION

Particle accelerators

To study the basic constituents of matter, particles such as electrons and protons are accelerated to very high speeds before crashing them into atoms. The by-products from these collisions reveal a vast array of sub-particles, which have led to a better understanding of the fundamental properties of particles.

The charged particles are accelerated by electromagnetic fields, but very long paths are required for the particles to reach the extremely high speeds needed (very close to the speed of light). To achieve this without the need for tunnels hundreds of kilometres long, particles travel through very strong magnetic fields, causing them to move in circles.

The Australian Synchrotron (Figure 5.2.5) near Monash University in Melbourne, is 70 m in diameter. It accelerates electrons through an equivalent of 3000 million volts (3 gV). At this energy, they travel at 99.999 99% of the speed of light. The relativistic effects that occur at these near-light speeds increases the effective mass of the electrons to about 6000 times their mass at rest. As they are being accelerated, the electrons give off electromagnetic radiation. It is this light, ranging from infrared through to X-ray wavelengths, that is used for the research projects being conducted at the synchrotron. More details about particle accelerators can be found in Chapter 9.

FIGURE 5.2.5 An inside view of the Australian Synchrotron, which uses very strong magnets to accelerate electrons to near-light speeds.

5.2 Review

SUMMARY

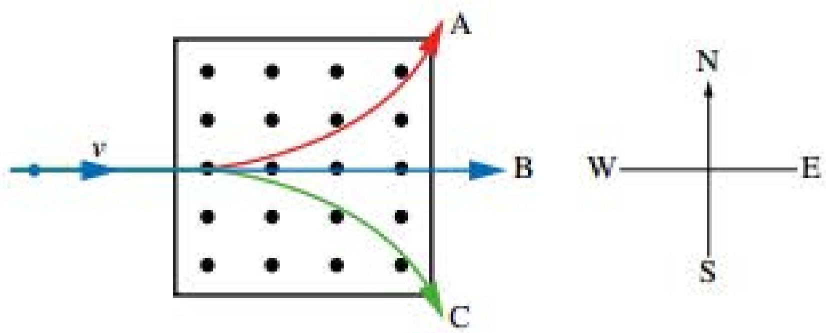
- The magnitude of the force on a charged object within a magnetic field is proportional to the flux density of the magnetic field B, the component of the velocity of the charge that is perpendicular (at right angles) to the magnetic field, and the charge on the particle: $F = qvB$.
- This force is referred to as the Lorentz force.
- The force is at a maximum when the charged particle is moving at right angles to the magnetic field.
- The force is zero when the charged particle is travelling parallel to the magnetic field.
- The right-hand palm rule is used to determine the direction of the force on a positive charge moving in a magnetic field, B. The direction of the force on a negatively charged particle is in the opposite direction.
- The force experienced by electrons in a magnetic field is utilised in particle accelerators.

KEY QUESTIONS

1 A single, positively charged particle with a charge of $+1.6 \times 10^{-19}$ C is travelling towards a computer screen and perpendicular to a magnetic field, B, that runs horizontally from left to right across the screen. In what direction will the force experienced by the charge act?

A left to right
B right to left
C vertically up
D vertically down

2 The following diagram shows a particle, with initial velocity v, about to enter a uniform magnetic field, B, directed out of the page.



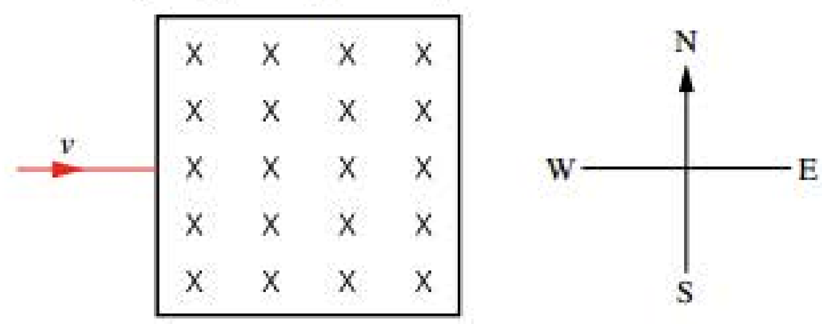
a If the charge on this particle is positive, what is the direction of the force on this particle just as it enters the field?
b Which path will this particle follow, A, B or C?
c Does the kinetic energy of the particle increase, decrease or remain constant?
d If this particle were negatively charged, what path would it follow?
e What kind of particle could follow path B?

3 A single, positively charged particle with a charge of $+1.6 \times 10^{-19}$ C travels at a velocity of $0.5\,\text{m s}^{-1}$ from left to right perpendicular to a magnetic field, B, of flux density 2.0×10^{-5} T, running vertically downwards. What is the magnitude of the force that the particle will experience from the magnetic field?

A 1.6×10^{-5} N
B 3.2×10^{-5} N
C 1.6×10^{-19} N
D 1.6×10^{-24} N

4 A single, negatively charged particle with a charge of -1.6×10^{-19} C travels at a velocity of $1.0\,\text{m s}^{-1}$ from right to left parallel to a magnetic field, B, of flux density 3.0×10^{-5} T.
What is the magnitude of the force the particle will experience from the magnetic field?

The following diagram applies to questions 5 and 6.



5 An electron with a charge of -1.6×10^{-19} C is moving eastwards into a magnetic field of flux density 1.5×10^{-5} T acting into the page, as shown in the diagram. If the magnitude of the initial velocity is $2\,\text{m s}^{-1}$, what is the magnitude and direction of the force the electron initially experiences as it enters the magnetic field?

6 An alpha particle with a charge of $+3.2 \times 10^{-19}$ C is moving eastwards into a magnetic field acting into the page, as shown in the diagram. The force it experiences is F. If the velocity, v, of the particle is doubled, what will be the magnitude and direction of the magnetic force it would experience in terms of F?

5.2 Review *continued*

7 How are particle accelerators able to provide the centripetal acceleration to change the direction of a charged particle using electromagnetic fields?

8 An ion with a charge of -3.20×10^{-19} C is moving into a perpendicular magnetic field of flux density 3.00×10^{-5} T and experiences a centripetal force of 9.60×10^{-23} N. Calculate the speed of the ion.

9 A mass spectrometer uses a magnetic force to separate ions with the same charge and speed but different masses (e.g. isotopes of an element). Calculate the magnetic flux density needed to provide a force of 3.20×10^{-16} N on a Cl^- ion travelling at 4.00×10^4 m s^{-1}.

10 If the speed of the ion in Question 9 is doubled, what change would be needed to the magnetic flux density to produce the same deflection?

5.3 The force on a conductor

Since a conducting wire is essentially a stream of charged particles flowing in one direction, it is not hard to imagine that a conductor carrying a stream of charges within a magnetic field will also experience a force. This is the theory behind the operation of electric motors, such as the one that powers an electric car (Figure 5.3.1).

FIGURE 5.3.1 From simple toys to electric trains to the latest electric cars, the electric motors that power them rely on the force applied to a current-carrying conductor in a magnetic field.

FORCE ON A CURRENT-CARRYING CONDUCTOR

The current in a conductor is dependent on the rate at which charges are moving through the conductor; that is:

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

where I is the current (A)

q is the total charge (C)

t is the time taken (s).

For a length of conductor ℓ, the velocity of the charges through the conductor is:

$$v = \frac{\ell}{t}, \text{ so } t = \frac{\ell}{v}$$

and hence

$$I = \frac{q}{t} = q \times \frac{v}{\ell}$$

$$I\ell = qv$$

As $F = qvB$ for a single charge, q, moving perpendicular to a magnetic field, then by substituting $I\ell$ for qv:

for a conductor of length ℓ, the force on the conductor is calculated using $F = I\ell B$

and for a conductor made up of n loops or conductors of length ℓ:

$F = nI\ell B$

where F is the force on the conductor perpendicular to the magnetic field in newtons (N)

n is the number of loops or conductors

I is the current in the conductor in amperes (A)

ℓ is the length of the conductor in metres (m)

B is the magnetic flux density in tesla (T)

Just as for a single charge moving in a magnetic field, the force on the conductor is at a maximum when the conductor is at right angles to the field. The force is zero when the conductor is parallel to the magnetic field. The right-hand palm rule is used to determine the direction of the force.

EXTENSION

Conductors at an angle to a magnetic field

The force experienced by a current-carrying conductor is a vector quantity. The expression noted above applies only to that component of the conductor perpendicular to the magnetic field. To find the force acting on any conductor, or part of a conductor, moving at an angle θ to the magnetic field, use the equation:

$$F = nI\ell B\sin\theta$$

This is particularly relevant when applied to practical electric motors.

PHYSICS IN ACTION

The current balance

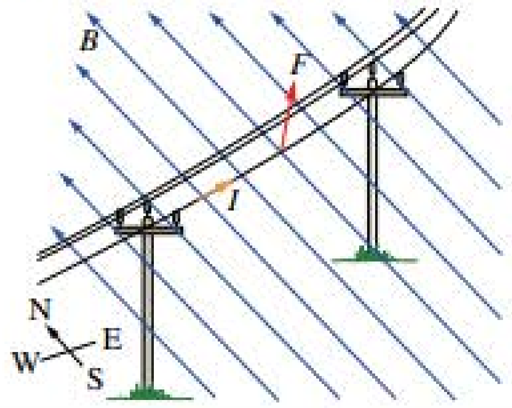
A current balance can be used to determine the force on a conductor in a magnetic field, as shown in Figure 5.3.2. If a set of data is collected by experiment, then a calibration curve of force versus current can be plotted. The gradient of the graph can then be used to measure the magnitude of unknown magnetic flux density. These instruments can be highly sensitive and introduce random errors. The direction of the magnetic field can be found from the direction of the force (whether the mass reading increases or decreases), remembering that it is the reaction force of the wire on the magnet that is recorded, and so the magnetic force on the wire is in the opposite direction.

FIGURE 5.3.2 A current balance is used to measure the interaction between an electric conductor and a magnetic field. The relationship between force, current and conductor length can be shown.

Worked example 5.3.1

MAGNITUDE OF THE FORCE ON A CURRENT-CARRYING WIRE

Determine the magnitude of the force per metre due to the Earth's magnetic field that acts on a suspended power line running east–west near the equator at the moment it carries a current of 100 A from west to east. Assume that the flux density of the Earth's magnetic field at this point is 5.0×10^{-5} T.

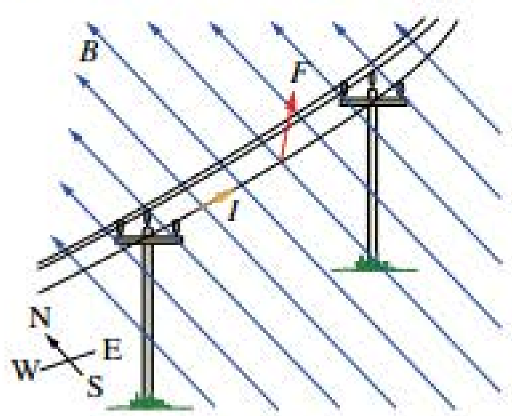


Thinking	Working
Check the direction of the conductor and determine whether a force will apply. Forces only apply to the component of the wire perpendicular to the magnetic field.	As the current is running west–east and the Earth's magnetic field runs south–north, the current and the field are at right angles and a force will exist.
Establish what quantities are known and what are required. Since the force per metre is being considered, use a length of 1 m.	$F = ?$ $n = 1$ $I = 100$ A $\ell = 1.0$ m $B = 5.0 \times 10^{-5}$ T
Substitute values into the force equation and simplify.	$F = nI\ell B$ $= 1 \times 100 \times 1.0 \times 5.0 \times 10^{-5}$ N $= 5.0 \times 10^{-3}$ N
Express the final answer in an appropriate form with a suitable number of significant figures. Note that only magnitude has been requested, so do not include direction.	$F = 5.0 \times 10^{-3}$ N per metre of power line

Worked example: Try yourself 5.3.1

MAGNITUDE OF THE FORCE ON A CURRENT-CARRYING WIRE

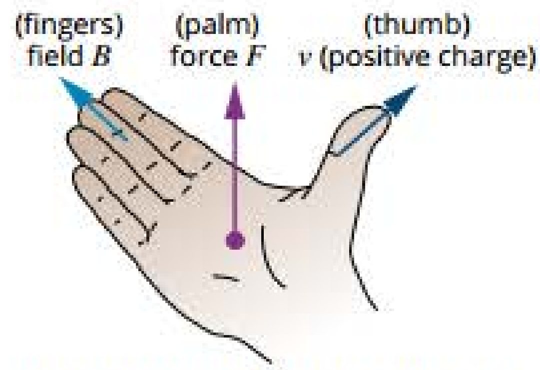
Determine the magnitude of the force per metre due to the Earth's magnetic field that acts on a suspended power line running east–west near the equator at the moment it carries a current of 50 A from west to east. Assume that the flux density of the Earth's magnetic field at this point is 5.0×10^{-5} T.



Worked example 5.3.2

DIRECTION OF THE FORCE ON A CURRENT-CARRYING WIRE

A current balance is used to measure the force from a magnetic field on a wire of length 5.0 cm running perpendicular to the magnetic field. The conventional current direction in the wire is from left to right. The magnetic field can be considered to be running into the page. What is the direction of the force on the wire?

Thinking	Working
 The right-hand palm rule is used to determine the direction of the force.	Align your hand so that your fingers are pointing in the direction of the magnetic field, i.e. into the page. Align your thumb so it is pointing right, in the direction of the current. Your palm should be facing up the page. That is the direction of the force applied by the magnetic field on the wire.
State the direction in terms of the other directions included in the question. Make the answer as clear as possible to avoid any misunderstanding.	The force on the wire is acting up the page.

Worked example: Try yourself 5.3.2

DIRECTION OF THE FORCE ON A CURRENT-CARRYING WIRE

A current balance is used to measure the force from a magnetic field on a wire of length 5.0 cm running perpendicular to the magnetic field. The conventional current direction in the wire is from left to right. The magnetic field can be considered to be running out of the page. What is the direction of the force on the wire?

Worked example 5.3.3

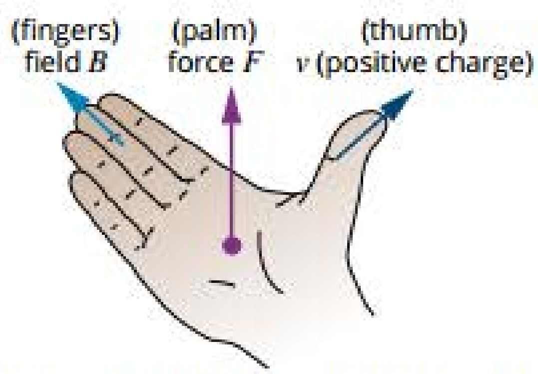
FORCE AND DIRECTION ON A CURRENT-CARRYING WIRE

The Amundsen–Scott South Pole Station sits at a point that can be considered to be at the Earth's South Magnetic Pole (which behaves like the north pole of a magnet).

Assume the flux density of the Earth's magnetic field at this point is 5.0×10^{-5} T.

a Determine the magnitude and direction of the magnetic force on a 2.0 m length of wire carrying a conventional current of 10.0 A vertically up the exterior wall of one of the buildings.

Thinking	Working
Forces only apply to the components of the wire running perpendicular to the magnetic field. The direction of the magnetic field at the South Magnetic Pole will be almost vertically upwards.	The section of the wire running up the wall of the building will be parallel to the magnetic field, *B*. Hence, no force will apply.
State your answer. A numeric value is required. As there is no force, it is not necessary to state a direction.	$F = 0$ N

b Determine the magnitude and direction of the magnetic force on a 2.0 m length of wire carrying a conventional current of 10.0 A running horizontally right to left across the exterior of one of the buildings.	
Forces only apply to the components of the wire running perpendicular to the magnetic field. The direction of the magnetic field at the South Magnetic Pole will be almost vertically upwards (that is, out of the ground).	The section of the wire running horizontally along the building will be perpendicular to the magnetic field, B. A force F equivalent to $nI\ell B$ will apply.
Identify the known quantities.	$F = ?$ $n = 1$ $I = 10.0\ \text{A}$ $\ell = 2.0\ \text{m}$ $B = 5.0 \times 10^{-5}\ \text{T}$
Substitute into the appropriate equation and simplify.	$F = nI\ell B$ $= 1 \times 10.0 \times 2.0 \times 5.0 \times 10^{-5}$ $= 1.00 \times 10^{-3}\ \text{N}$
 The direction of the magnetic force is also required to fully specify the vector quantity. Determine the direction of the magnetic force using the right-hand palm rule.	Align your hand so that your fingers are pointing in the direction of the magnetic field, i.e. vertically up. Align your thumb so it is pointing left, in the direction of the current. Your palm should be facing inwards (towards the building). That is the direction of the force applied by the magnetic field on the wire.
State the magnetic force in an appropriate form with a suitable number of significant figures. Include the direction to fully specify the vector quantity.	$F = 1.0 \times 10^{-3}\ \text{N}$ into the wall

Worked example: Try yourself 5.3.3

FORCE AND DIRECTION ON A CURRENT-CARRYING WIRE

Santa's house sits at a point that can be considered the Earth's North Magnetic Pole (which behaves like the south pole of a magnet).

Assume the flux density of the Earth's magnetic field at this point is 5.0×10^{-5} T.

a Calculate the magnitude and direction of the magnetic force on a 2.0 m length of wire carrying a conventional current of 10.0 A vertically up the outside wall of Santa's house.

b Calculate the magnitude and direction of the magnetic force on a 2.0 m length of wire carrying a conventional current of 10.0 A running horizontally right to left across the outside of Santa's house.

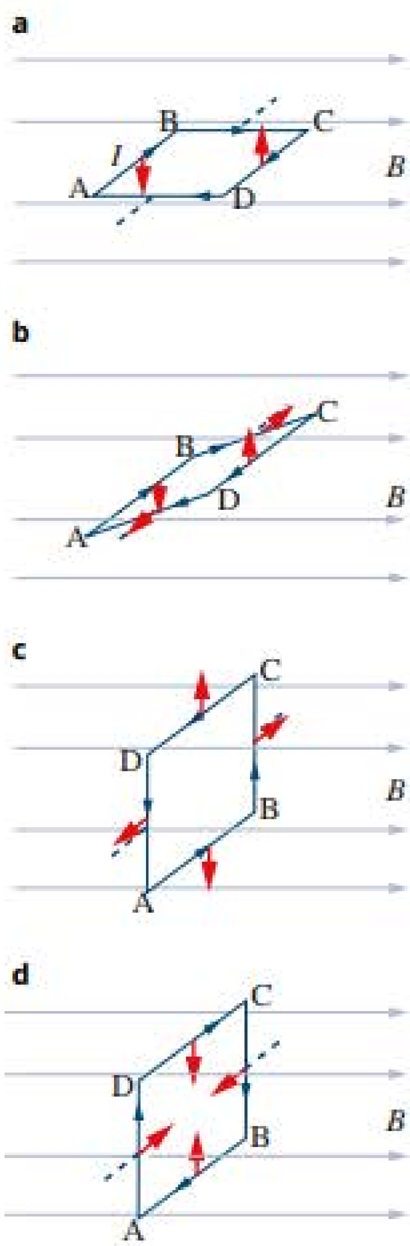


FIGURE 5.3.3 The magnetic force acting on each side of a current-carrying square wire loop in a magnetic field.

DC MOTORS

The main components and the principles have been the same for all DC motors since Michael Faraday built the first one in 1821.

A current-carrying loop of wire in an electric field experiences a magnetic force $F = I\ell B$ on two or more of its sides. In practice, many turns of wire are used and the magnetic field is provided by more than one permanent magnet or electromagnet.

Consider a single square loop of wire with vertexes ABCD and carrying a current, I, in a magnetic field, B, as shown in Figure 5.3.3.

Initially the wire loop is aligned horizontally in a magnetic field, B, as in Figure 5.3.3(a). Sides AD and BC are parallel to the magnetic field so no magnetic force will act on them. Sides AB and CD are perpendicular to the field so both of these sides will experience a magnetic force. Using the right-hand palm rule, we can show that there is a downwards force on AB and an upwards force on CD. These two forces will act together on the coil and cause it to rotate anticlockwise. If the coil is free to turn, it will move towards the position shown in Figure 5.3.3(b).

In Figure 5.3.3(b), there will be a magnetic force acting on every side. However, the forces acting on sides AD and BC will be equal and opposite in direction. They will tend to stretch the coil outwards but won't affect its rotation. The forces on sides AB and CD will remain and the coil will continue to rotate anticlockwise.

As the coil rotates to the position shown in Figure 5.3.3(c), the forces acting on each side are such that they will tend to keep the coil in this position. The force on each side will act outwards from the coil. There are no turning forces at this point, but any further rotation will cause a force in the opposite direction that will cause the coil to rotate clockwise, back to this perpendicular position.

For the coil to continue to rotate anticlockwise at this point, the current direction needs to be reversed. This is shown in Figure 5.3.3(d). With the current reversed, all of the forces are reversed, and provided the coil has a little momentum to get it past the perpendicular position, the coil will continue to rotate anticlockwise. This ability to reverse the current direction at the point where the coil is perpendicular to the magnetic field is a key design feature in DC motors. It is a **commutator** that allows the current to be reversed. (This will be discussed later in this chapter.)

Torque

The turning force that the coil experiences is referred to as the torque on the coil. A torque is the turning effect of any force; for example, pushing on a swinging door. To achieve the maximum effect, the force should be applied at right angles to the door and at the largest distance possible from the point where the door is hinged, as shown in Figure 5.3.4. Refer back to Chapter 3 for a fuller discussion of torque.

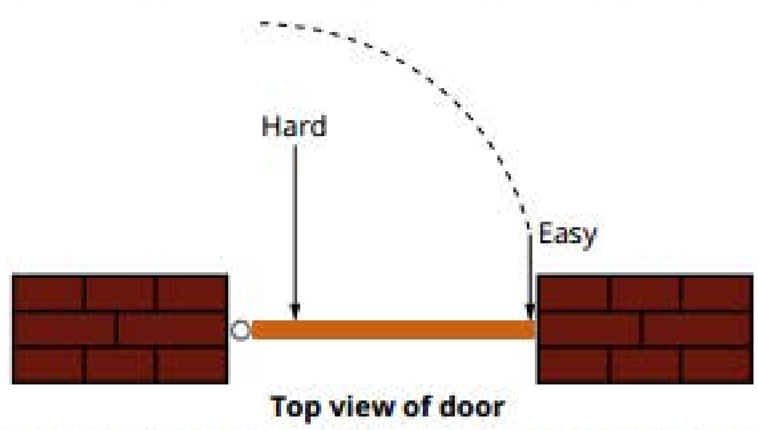


FIGURE 5.3.4 The force required to open a swinging door decreases as the perpendicular distance from the point of rotation increases. The torque, or turning force, is maximised.

Torque is defined as:

$\tau = r_\perp F$

where τ is the torque in newton metre (N m)

$r_\perp$ is the perpendicular distance between the axis of rotation and the point of application of the force in metres (m)

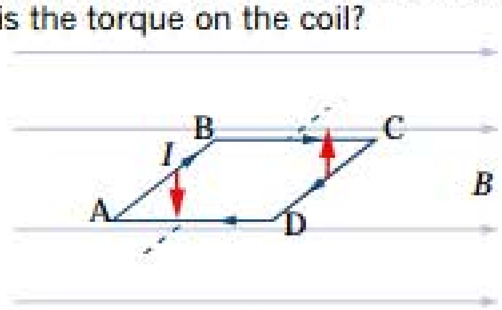
F is the component of the force perpendicular to the axis of rotation in newtons (N)

In the case of a single square or rectangular coil, the total torque applied to the coil will be twice that acting on one side, since each of the two sides perpendicular to the magnetic field will experience a force contributing to the total torque.

Worked example 5.3.4

TORQUE ON A COIL

A single wire coil, ABCD, of side length 5.00 cm and free to rotate, sits within a magnetic field, B, of flux density 1.00×10^{-4} T. A current of 1.00 A is flowing through the coil. What is the torque on the coil?



Thinking	Working
Confirm that the coil will experience a force based on the magnetic field and current directions supplied.	Using the right-hand palm rule confirms that a downwards force applies on side AB. An upwards force applies to side CD. The coil will turn anticlockwise. Sides AD and BC lie parallel to the magnetic field and no force will apply.
Calculate the magnetic force on one side.	$F = I\ell B$ $F = 1.00 \times 0.0500 \times 1.00 \times 10^{-4}$ $F = 0.0500 \times 10^{-4}$ N
Determine the distance, r, from the point of rotation at which the magnetic force is applied.	Length of side = 5.00 cm Length of rotation = $\frac{1}{2}$ × side length r = 2.50 cm = 0.0250 m
Calculate the torque applied by the magnetic force on one side of the coil.	$\tau = r_\perp F$ $= 0.0250 \times 0.0500 \times 10^{-4}$ $= 0.00125 \times 10^{-4}$ $\tau = 1.25 \times 10^{-7}$ N m
Since two sides, AB and CD, both experience a magnetic force and hence a torque, the torque on one side should be multiplied by 2 to find the total torque. State also the direction of rotation.	Total torque = $2 \times 1.25 \times 10^{-7}$ N m $\tau = 2.50 \times 10^{-7}$ N m The direction is anticlockwise.

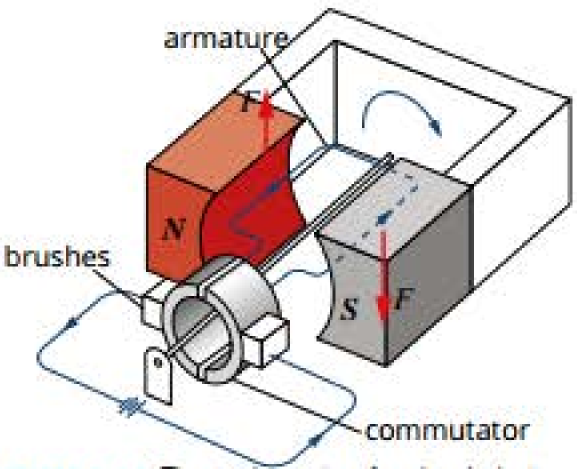
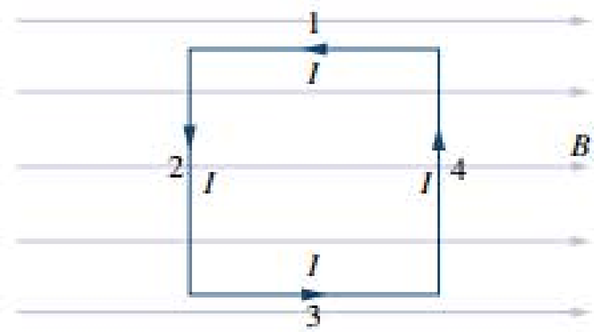


FIGURE 5.3.5 The main parts of a simple but practical electric motor.

Worked example: Try yourself 5.3.4

TORQUE ON A COIL

A single wire coil with a side length 4.0 cm, and free to rotate, sits within a magnetic field, B, of flux density 1.0×10^{-4} T. A current of 1.0 A is flowing through the coil. What is the torque on the coil? Please give your answer to two significant figures.



PRACTICAL DC MOTORS

A basic single coil electric motor with a simple arrangement to reverse the current direction will work, but it won't turn very smoothly, since maximum torque will only apply each half turn or twice for every full rotation. A number of enhancements have been developed over time to make DC motors the highly practical devices they are today.

The commutator is usually made from a split ring of copper or another good conductor on which conducting brushes (usually carbon blocks) rub. Each half is connected to one end of the coil of wire. This arrangement of brushes prevents the wire from becoming tangled as the coils rotate. The commutator reverses the current at the point where the coil is perpendicular to the magnetic field and the current direction reverses to keep the coil rotating (Figure 5.3.5).

Practical motors will have many coils of many turns, rather than just one, spaced at an angle to each other, as shown in Figure 5.3.6. The commutator is arranged to feed current to the particular coil that is in the best position to provide maximum torque. The total torque will be the sum of the torques on each individual coil.

The coils are wound around a soft iron core to increase the magnetic field through them. The whole arrangement of core and coils is called an **armature**. Permanent magnets are generally used to provide the magnetic field in small motors, but electromagnets are used in larger motors, as they can produce larger and stronger fields. These magnets are usually stationary, as distinct from the rotating rotor or armature, and are often referred to as the **stator**.

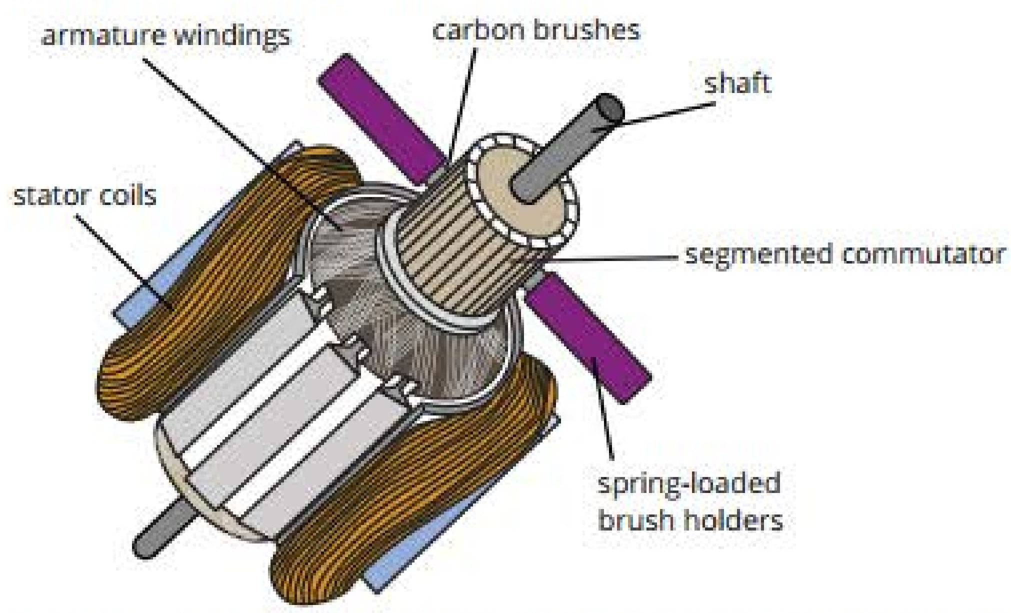


FIGURE 5.3.6 A typical universal electric motor, showing the main components. Some motors would have additional stator coils. The commutator feeds current to the armature coils in the position where most torque will be experienced. This type of motor will operate on either DC or AC power.

Generally speaking, the larger the torque in an electric motor the better. This is achieved by the use of a strong magnetic field, a large number of turns of wire in the coils, a high current and a large area of coil. All this adds to the cost, so when designing an electric motor, each aspect may be compromised to some extent in light of its potential use.

PHYSICSFILE

Michael Faraday (1791–1867), depicted in Figure 5.3.7, was an English scientist who worked in the areas of chemistry and physics. He had little formal education. At 14 he became the apprentice to a London bookbinder. During his apprenticeship he read many of the books that came his way. At 21 he became a laboratory assistant to Humphry Davy, one of the most prominent scientists of the day. Faraday was a gifted experimenter and after returning from a scientific tour through Europe with Davy, he began to be recognised in his own right for his scientific work. He was admitted to the Royal Society at age 32. He is credited with the discoveries of benzene, electromagnetic induction and the basis of the modern electric motor. He died in 1867 at Hampton Court. His contributions to science, and in particular his work in the area of electromagnetism, are recognised through the unit of measurement of capacitance known as the farad. In Chapter 6, you will study more of Faraday's work on electromagnetic induction.

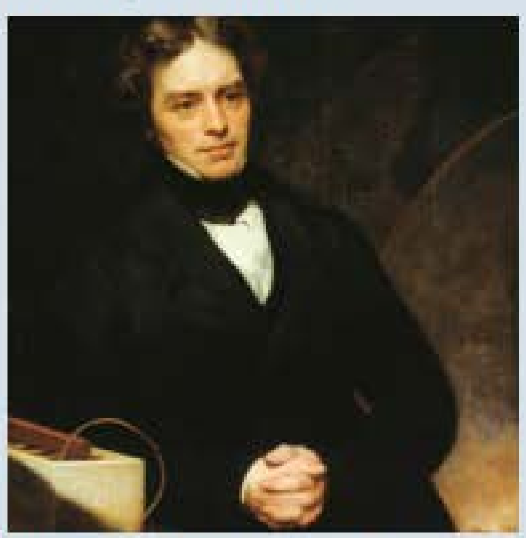


FIGURE 5.3.7 Michael Faraday.

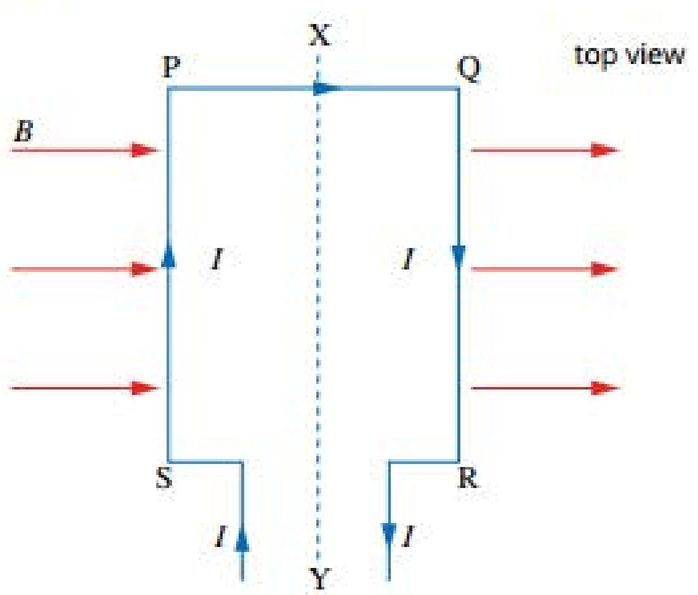
5.3 Review

SUMMARY

- The magnetic force on a current-carrying wire within a magnetic field is:
  $F = nI\ell B$
  where F is the force on the conductor perpendicular to the magnetic field in newtons (N)
  n is the number of loops or conductors
  I is the current in the conductor in amperes (A)
  ℓ is the length of the conductor in metres (m)
  B is the flux density of the magnetic field in tesla (T).
- The direction of the force is given by the right-hand palm rule by which the force travels out of the palm of the hand, once the thumb and fingers are orientated in the direction of the (conventional) current and magnetic field respectively.
- There is a torque on a loop of wire carrying a current whenever the current is not parallel to the field. Torque is defined as:
  $\tau = r_{\perp}F$
  where τ is the torque in newton metre (N m)
  $r_{\perp}$ is the perpendicular distance, in metres (m), between the axis of rotation and the point of application of the force
  F is the component of the force, in newtons (N), perpendicular to the axis of rotation.
- The wire loop of a simple DC motor keeps rotating because the direction of current, and hence the torque, is reversed each half turn by the commutator.
- The armature of a practical motor consists of many loops that are fed current by the commutator when they are in the position of maximum torque.
- The total torque will be the sum of the torques on each individual coil.
- In the case of a single square or rectangular coil, the total torque applied to the coil will be twice that acting on the one side, since each of the two sides with current perpendicular to the magnetic field will experience a force contributing to the total torque.

KEY QUESTIONS

1 A rectangular loop of wire is carrying a current, I, in a magnetic field, B, as shown below. What is the direction of the force on the length of wire marked PQ?



2 An east–west power line of length 100 m is suspended between two towers. Assume that the flux density of the magnetic field of the Earth in this region is 5.0×10^{-5} T. Calculate the magnetic force (including direction) on this power line at the moment it carries a current of 80 A from west to east.

3 An east–west power line of length 80 m is suspended between two towers. Assume that the flux density of the magnetic field of the Earth in this region equals 4.5×10^{-5} T.

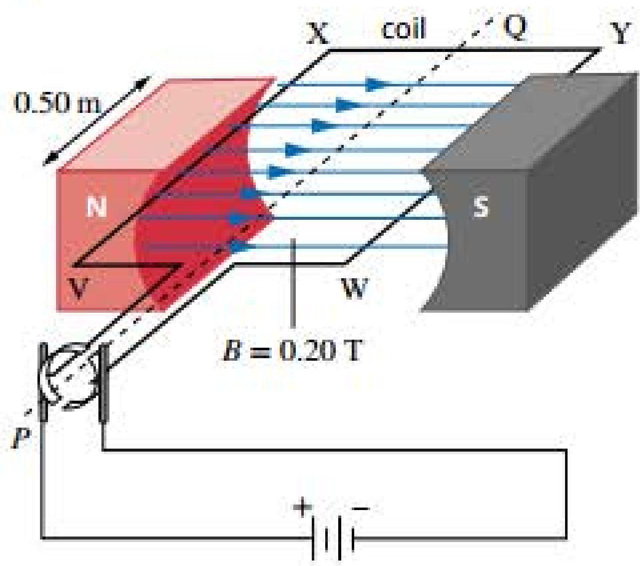
a Calculate the magnitude and direction of the magnetic force on this power line at the moment it carries a current of 50 A from east to west.

b Over time, the ground underneath the eastern tower subsides, so that the power line is lower at that tower. Assuming that all other factors are the same, is the magnitude of the magnetic force on the power line greater than before, less than before or the same as before?

5.3 Review *continued*

The following information applies to questions 4–6.

The diagram shows a simplified version of a direct current motor.



4 For the position of the coil shown, calculate the magnitude of the force on segment WY when a current of 1.0 A flows through the coil.

5 In which direction will the coil begin to rotate? Give your reasoning.

6 Which of the following actions would cause the coil to rotate faster?

A increasing the current

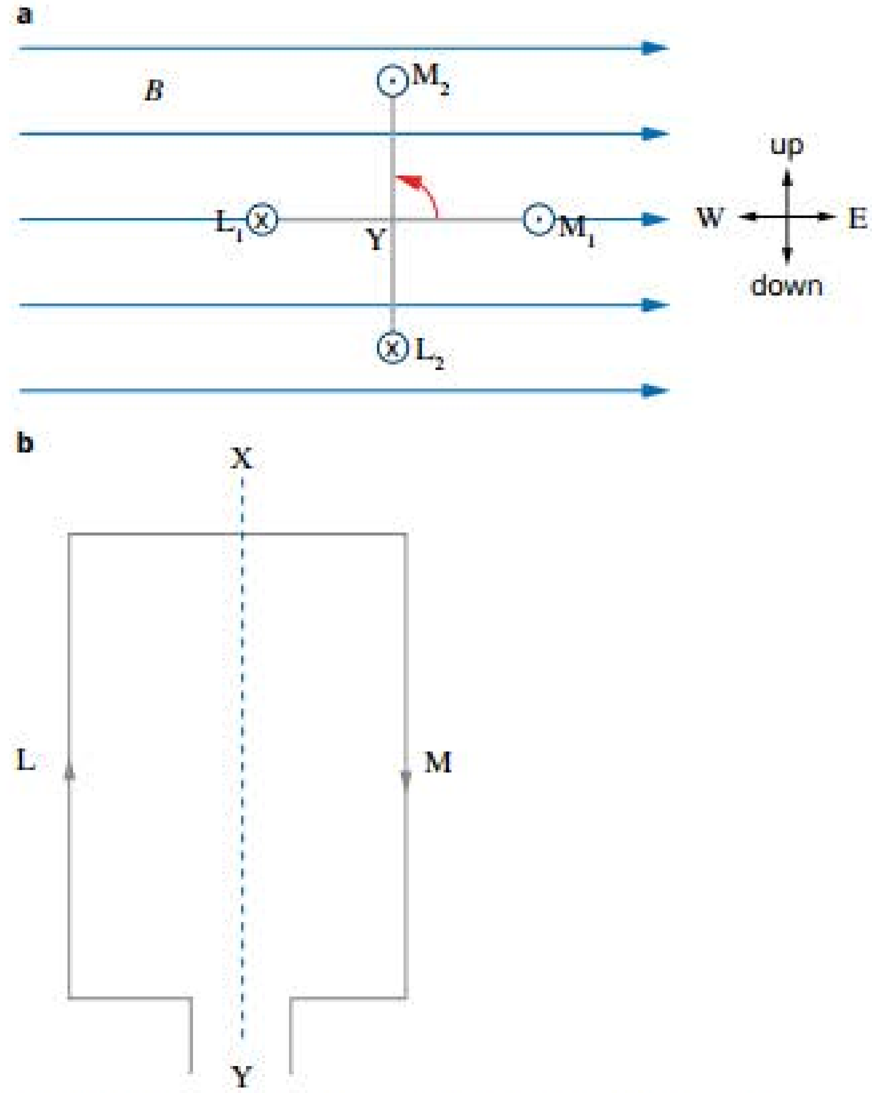
B increasing the magnetic flux density

C increasing the cross-sectional area of the coil

D all of the above

The following information applies to questions 7–10.

Diagram (a) shows an end-on view of a current-carrying loop, LM. The loop is free to rotate about a horizontal axis XY. You are looking at the loop from the Y end of the axis. The same loop is seen from the top in figure (b). Initially, arms L and M are horizontal (L_1–M_1). Later they are rotated so that they are vertical (L_2–M_2). The loop is located in an external magnetic field of magnitude *B* directed east (at right angles to the axis of the loop). Note the current directions in (a): out of the page in M and into the page in L. Give your answers with reference to the up–down, W–E cross arrows in (a).



7 When LM is aligned horizontally (L_1–M_1), what is the direction of the magnetic force on:

a side L?

b side M?

8 In what direction, as seen from Y, will the loop rotate?

9 When LM is aligned vertically (L_2–M_2) what is the:

a direction of the magnetic force on side L?

b direction of the magnetic force on side M?

c magnitude of the torque acting on the loop? Give a reason for your answer.

10 When LM is aligned vertically, which one of the following actions will result in a torque acting on the coil that will keep it rotating in an anticlockwise direction? (Assume it still has some momentum when it reaches the vertical position.)

A decrease the current through the loop

B increase the magnetic flux density

C reverse the direction of the current through the coil

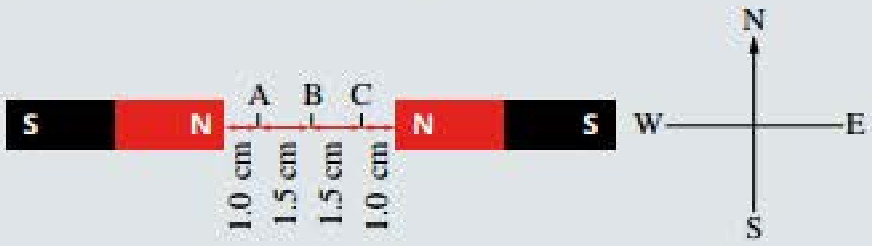
Chapter review

KEY TERMS

armature
cathode ray tube
commutator
dipole
electromagnet
Lorentz force
magnetic
magnetic field
magnetic pole
mnemonic
pole
right-hand grip rule
right-hand palm rule
solenoid
stator
voltaic pile

The following information applies to questions 1–3.

Two strong bar magnets, which produce magnetic fields of equal flux density, are arranged as shown.



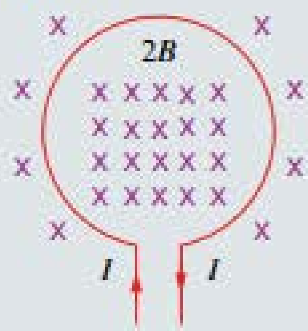
1 Ignoring the magnetic field of the Earth, what is the approximate direction of the resulting magnetic field at point A?

2 Ignoring the magnetic field of the Earth, what is the approximate direction of the resulting magnetic field at point C?

3 Ignoring the magnetic field of the Earth, what is the magnitude of the resulting magnetic field at point B?

The following information relates to questions 4–6.

The diagram shows a loop carrying a current *I* that produces a magnetic field of magnitude *B* in the centre of the loop. It is in a region where there is already a steady field of magnitude *B* (the same magnitude as that due to *I*) directed into the page. The resultant magnetic field has a magnitude of 2*B*.



4 What would be the magnitude and direction of the resultant field at the centre of the loop if the current in the loop is switched off?

5 What would be the magnitude and direction of the resultant field at the centre of the loop if the current in the loop were doubled?

6 What would be the magnitude and direction of the resultant field at the centre of the loop if the current in the loop were reversed but maintained at the same magnitude as *I*?

7 A magnetic flux density, *B*, of 4.00×10^{-5} T is measured at a distance *r* away from a long, straight wire. If the current, *I*, is measured as being 3.00 A, what is the distance from the wire in centimetres?

8 Jumper cables are used to start a car when the battery is flat. When starting the car the voltage is limited to about 12 V but the current can sometimes be as much as 15 A. What is the magnitude of the magnetic field in μT created by the jumper cables at a point 15 cm from the cables?

9 If a magnetic field no larger than that of the Earth, 5.00×10^{-5} T, is the maximum allowed 30.0 cm from an electrical wire, what is the maximum current that the wire can carry?

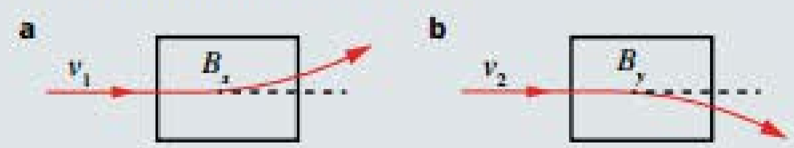
10 Complete the following sentence by selecting the best option.
The magnitude of the magnetic force on a conductor aligned so that the current is running parallel to a magnetic field is:

A dependent on the size of the current
B dependent on the size of the magnetic field
C dependent on the length of the conductor
D zero
E a maximum

11 The right-hand palm rule is used to determine the force on a current-carrying conductor perpendicular to a magnetic field. Identify what part of the hand corresponds to the following physical quantities:

a magnetic force
b magnetic field
c current in the conductor

12 The following diagrams show two different electron beams being bent as they pass through two different regions of a uniform magnetic field of equal magnitudes B_x and B_y. The initial velocities of the electrons in the respective beams are v_1 and v_2. Complete the following sentence by choosing the correct term from those in **bold**.



For the electron beams to behave as shown in (a), v_1 is **equal to/less than** v_2 and the region of the magnetic field, B_y, must be acting **out of/into** the page.

13 How much current, I, must be flowing in a wire 3.2 m long if the maximum force on it is 0.800 N and it is placed in a uniform magnetic field of 0.0900 T?

14 Calculate the magnitude and direction of the magnetic force on conductors with the following sets of data:

a $B = 1.0$ mT left, $\ell = 5.0$ mm, $I = 1.0$ mA up

b $B = 0.10$ T left, $\ell = 1.0$ cm, $I = 2.0$ A up

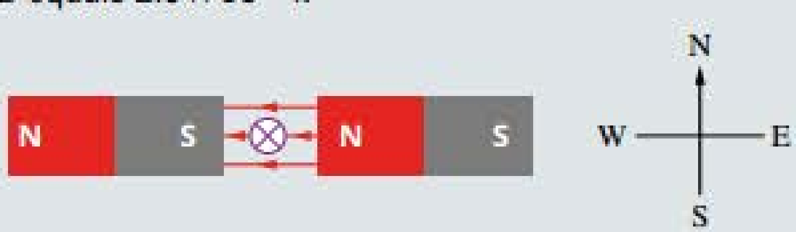
15 Calculate the magnitude of the force exerted on an electron ($q = 1.6 \times 10^{-19}$ C) travelling at a speed of 7.0×10^{6} m s^{-1} at right angles to a uniform magnetic field of flux density 8.6×10^{-3} T.

16 A horseshoe magnet is held vertically with the north pole of the magnet on the left and the south pole of the magnet on the right.

What is the direction of the magnetic force acting on a wire between the poles, directing conventional current into the page?

17 Power lines carry an electric current in the Earth's magnetic field. Which would experience the greater magnetic force: a north–south power line or an east–west power line? Explain your answer.

18 The diagram below depicts a cross-sectional view of a long, straight, current-carrying conductor, located between the poles of a permanent magnet. The magnetic field, B, of the magnet, and the current, I, are perpendicular. Calculate the magnitude and direction of the magnetic force on a 5.0 cm section of the conductor when the current is 2.0 A into the page and B equals 2.0×10^{-3} T.



19 For which of the following situations is torque at a maximum?

A The force is applied perpendicular to the axis of rotation.

B The force is applied parallel to the axis of rotation.

C The force is applied at a maximum regardless of direction.

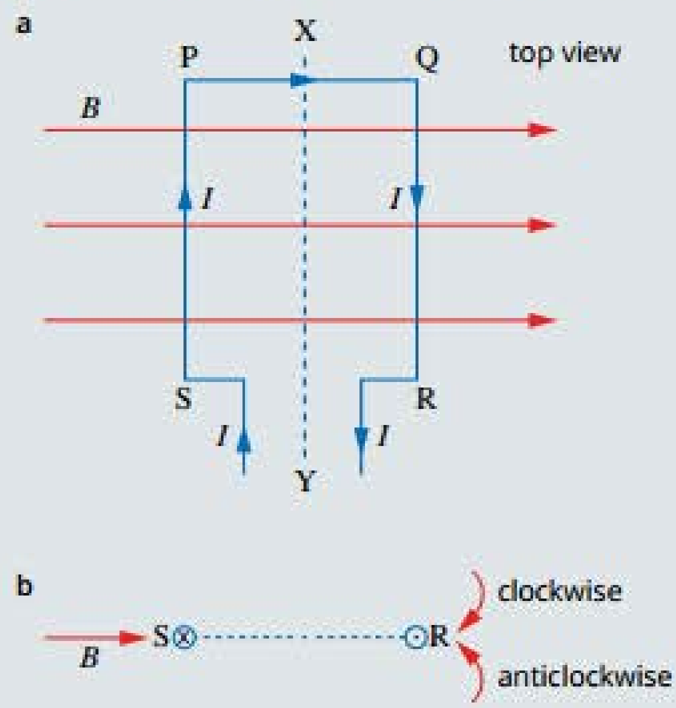
D The force applied is zero.

The information below applies to questions 20–25.

Part (a) of the diagram below depicts a top view of a single current-carrying coil in an external magnetic field B.

Part (b) of the diagram is the corresponding cross-sectional view as seen from point Y. The following data apply:

$B = 0.10$ T, PQ = 2.0 cm, PS = QR = 5.0 cm, $I = 2.0$ A



20 What is the magnitude and direction of the magnetic force acting on side PS?

21 What is the magnitude and direction of the magnetic force acting on side QR?

22 What is the magnitude of the force on side PQ?

23 The coil is free to rotate about an axis through XY. In what direction, as seen from Y, would the coil rotate?

24 Which of the following does not affect the magnitude of the torque acting on this coil?

A the dimensions of the coil

B the magnetic flux density

C the magnitude of the current through the coil

D the direction of the current through the coil

25 What is the total torque acting on the coil?

26 Briefly explain the function of the commutator in an electric motor.

CHAPTER

06 Magnetic field and emf

In this chapter, electromagnetic induction—the creation of an electric current from a changing magnetic flux—is explored.

In 1831, Englishman Michael Faraday and American Joseph Henry independently discovered that a changing magnetic flux could induce an electric current in a conductor. This discovery made possible the production of vast quantities of electricity. Today, whether the primary energy source is burning coal, wind, nuclear fission or falling water, the vast bulk of the world's electrical energy production is the result of electromagnetic induction.

Science as a Human Endeavour

Electromagnetism is utilised in a range of technological applications, including:

- DC electric motor with commutator, and back emf
- AC and DC generators
- transformers
- regenerative braking
- induction hotplates
- large scale AC power distribution systems

Science Understanding

- an induced emf is produced by the relative motion of a straight conductor in a magnetic field when the conductor cuts flux lines
  This includes applying the relationship
  induced emf = ℓvB where $v \perp B$
- magnetic flux is defined in terms of magnetic flux density and area
  This includes applying the relationship
  $\Phi = BA_{\perp}$
- a changing magnetic flux induces a potential difference; this process of electromagnetic induction is used in step-up and step-down transformers, DC and AC generators
  This includes applying the relationships
  $$\text{induced emf} = -N\frac{(\Phi_2 - \Phi_1)}{t} = -N\frac{\Delta\Phi}{t} = -N\frac{\Delta(BA_{\perp})}{t}$$
  $$\text{AC generator emf}_{max} = 2N\ell vB = 2\pi NBA_{\perp}f,\ \text{emf}_{rms} = \frac{\text{emf}_{max}}{\sqrt{2}}$$
  $$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$
  $$P = VI = I^2R = \frac{V^2}{R}$$
- conservation of energy, expressed as Lenz's Law of electromagnetic induction, is used to determine the direction of induced current

6.1 Induced emf in a conductor moving in a magnetic field

FIGURE 6.1.1 Michael Faraday's original induction coil. Passing a current through one coil induces a voltage in the second coil by a process called mutual inductance. This coil is now on display at the Royal Institution in London.

After Oersted's discovery that an electric current produces a magnetic field (see Section 5.1, page 152), Michael Faraday, an English scientist, was convinced that the reverse should also be true—a magnetic field should be able to produce an electric current.

Faraday wound two coils of wire onto an iron ring (Figure 6.1.1). By connecting a battery to one of the coils he created a strong current through one coil which therefore created a strong magnetic field. He expected to then detect the creation of an electric current in the second coil. No matter how strong the magnetic field, he could not detect an electric current in the other coil.

One day he noticed that the galvanometer (a type of sensitive ammeter) attached to the second coil flickered when he turned on the current that created the magnetic field. It gave another flicker, in the opposite direction, when he turned the current off. It was not the strength of the magnetic field that mattered, but the change in the magnetic field.

The creation of an electric current in a conductor due to a change in the magnetic field acting on that conductor is now called **electromagnetic induction**. This section focuses on this concept.

PHYSICSFILE

Models and theories

Michael Faraday was not alone in the discovery of electromagnetic induction. Joseph Henry (1797–1878), an American physicist, independently discovered the phenomenon of electromagnetic induction a little ahead of Michael Faraday, but Faraday was the first to publish his results. Henry later improved the design of the electromagnet by using a soft iron core wrapped in many turns of wire. He also designed the first reciprocating electric motor. Henry is credited with first discovering the phenomenon of self-induction, and the unit of inductance is named after him. He also introduced the electric relay, which made the sending of telegrams possible. Henry was the first director of the Smithsonian Institution.

Although Faraday will be largely referred to throughout this text, it is worth noting that there can be a number of contributors who together build on the understanding of key ideas. Joseph Henry's contributions should not be forgotten.

CREATING AN ELECTRIC CURRENT

Faraday was unable to produce an electric current when he used a constant magnetic field, but he was able to observe the creation of an electric current whenever there was a change in the magnetic field. This current is produced by an induced **emf**, ε. Although the term 'emf' is derived from the name '**electromotive force**', it is a voltage, or potential difference, rather than a force. Figure 6.1.2 indicates the induction of emf, and therefore current in a complete circuit, caused by the perpendicular movement of a conducting wire relative to a magnetic field.

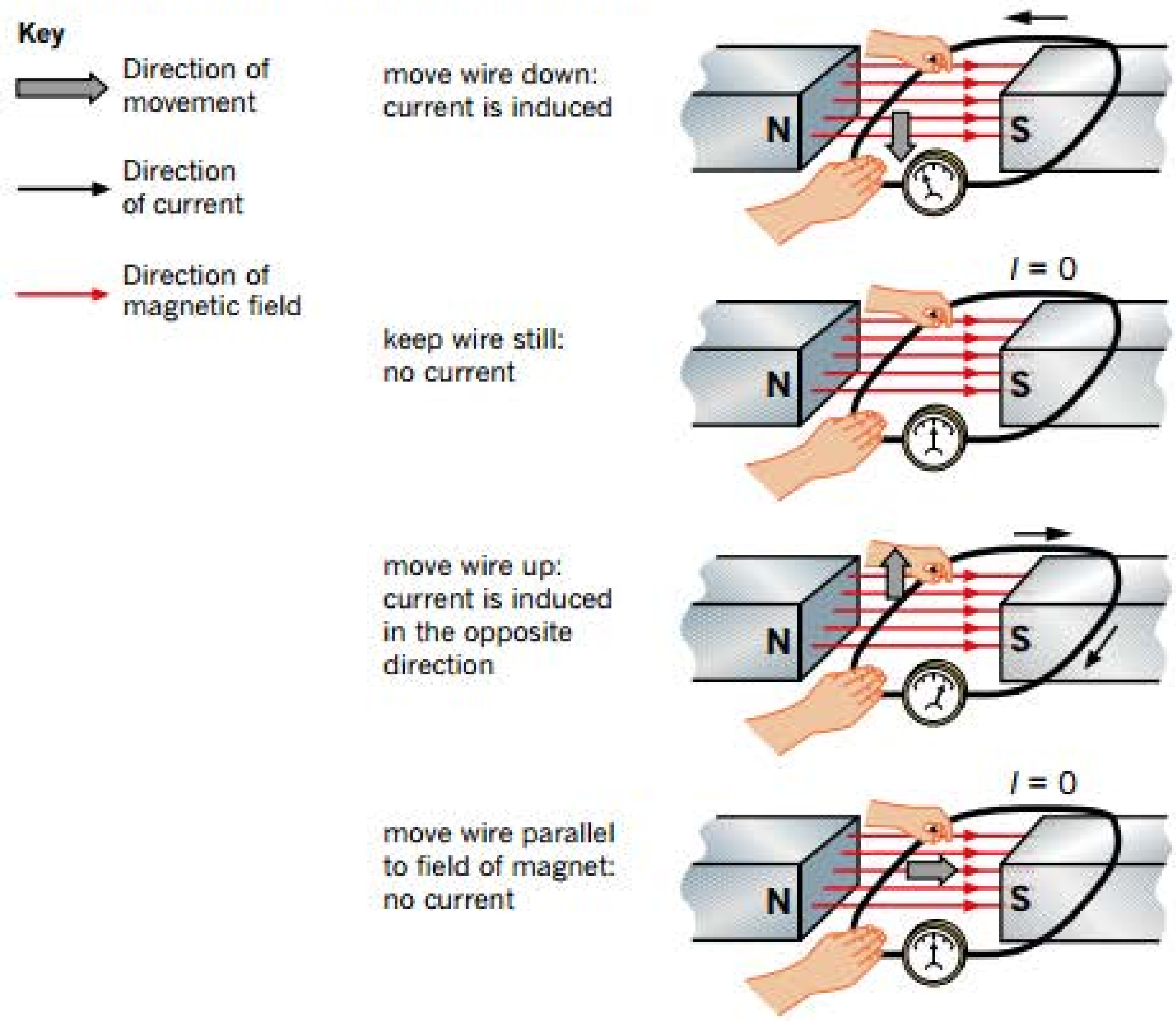


FIGURE 6.1.2 An electromotive force (emf) is induced in a wire when the wire moves perpendicular to a magnetic field.

INDUCED EMF IN A MOVING CONDUCTOR

It was discovered that a change in the magnetic field, when a magnet is moved closer to a conductor, leads to an induced emf that in turn produces an **induced current** in a complete circuit. Although Faraday largely based his investigations on induced currents in coils, another way of inducing an emf is by moving a straight conductor in a magnetic field. It's not hard to understand why this is the case, when you know that charges moving in a magnetic field will experience a force.

In Chapter 5, it was established that when a charge, q, moves at a speed, v, perpendicular to a magnetic field, B, the charge experiences a force, F, equal to qvB. So:

$F = qvB$

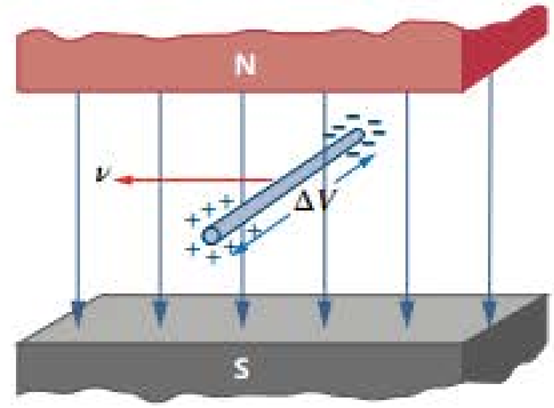


FIGURE 6.1.3 A potential difference, ΔV, will be produced across a straight wire moving to the left in a downward-pointing magnetic field.

Considering the direction of movement shown in Figure 6.1.3, the force on the positive charges within the moving conductor would be along the conductor and out of the page. The force on the negative charges within the conductor would be along the conductor but into the page.

As the charges in Figure 6.1.3 move apart due to the force they are experiencing from the magnetic field, one end of the conductor will become more positive, the other more negative and a potential difference, ΔV, or emf will be induced between the ends of the conductor.

Consider now an electron moving along the conductor. The force from the magnetic field will do work on the electron as it moves along a length, ℓ. To calculate the work done:

$W = \text{force} \times \text{distance} = qvB \times \ell$

The emf is equal to the work done per unit charge, so:

$$\varepsilon = \frac{W}{q} = \frac{q\ell vB}{q}$$

and thus:

$\varepsilon = \ell vB$

where ε is the induced emf (V)
ℓ is the length of the conductor (m)
v is the speed of the conductor perpendicular to the magnetic field ($m\,s^{-1}$)
B is the strength of the magnetic field (T)

Worked example 6.1.1

ELECTROMOTIVE FORCE ACROSS AN AIRCRAFT'S WINGS

Will a moving aeroplane develop a dangerous emf between its wing tips solely from the Earth's magnetic field? An aircraft with a wingspan of 64.0 m is flying at a speed of $990\,km\,h^{-1}$ at right angles to the Earth's magnetic field of 5.00×10^{-5} T.

Thinking	Working
Identify the quantities required in the correct units.	$\varepsilon = ?$ $\ell = 64.0\,m$ $B = 5.00 \times 10^{-5}\,T$ $v = 990\,km\,h^{-1}$ $= 990 \times \frac{1000}{3600} = 275\,m\,s^{-1}$
Substitute into the appropriate formula and simplify.	$\varepsilon = \ell vB$ $= 64.0 \times 275 \times 5.00 \times 10^{-5} = 0.880\,V$
State your answer as a response to the question.	$\varepsilon = 0.880\,V$ This is a very small emf and would not be dangerous.

Worked example: Try yourself 6.1.1

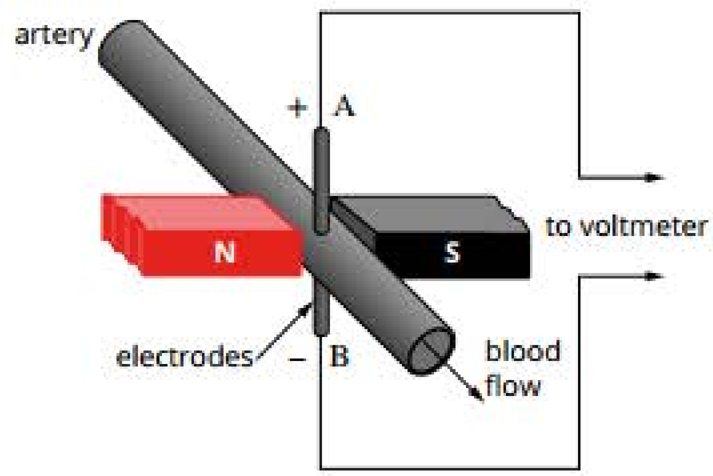
ELECTROMOTIVE FORCE ACROSS AN AIRCRAFT'S WINGS

Will a moving aeroplane develop a dangerous emf between its wing tips solely from the Earth's magnetic field? A fighter jet with a wingspan of 25.0 m is flying at a speed of 2000 km h^{-1} at right angles to the Earth's magnetic field of 5.00×10^{-5} T.

Worked example 6.1.2

FLUID FLOW MEASUREMENT

The rate of fluid flow within a vessel can be measured using the induced emf when the fluid contains charged ions. A small magnet and a sensitive voltmeter calibrated to measure speed are used. This can be applied to measure the flow of blood (which contains iron in solution) in the human body. If the diameter of a particular artery is 2.00 mm, the strength of the magnetic field applied is 0.100 T and the measured emf is 0.100 mV, what is the speed of the flow of the blood within the artery?

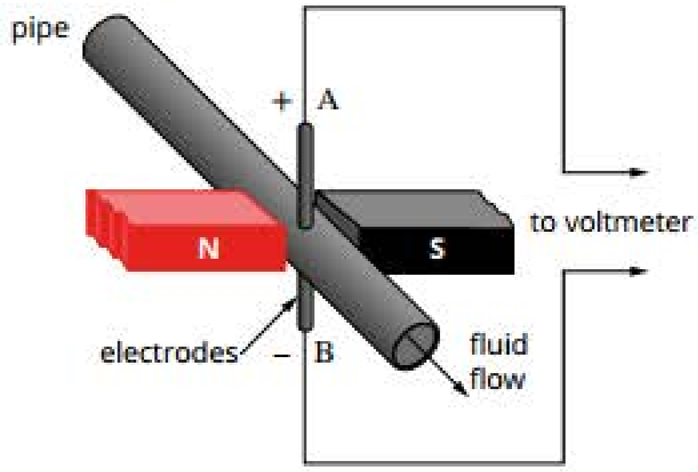


Thinking	Working
Identify the quantities required and put them into SI units.	$\varepsilon = 0.100\ \text{mV} = 1.00 \times 10^{-4}\ \text{V}$ $\ell = 2.00\ \text{mm} = 2.00 \times 10^{-3}\ \text{m}$ $v = ?$ $B = 0.100\ \text{T}$
Rearrange the appropriate formula, substitute and simplify.	$\varepsilon = \ell v B$ $v = \frac{\varepsilon}{\ell B}$ $= \frac{1.00 \times 10^{-4}}{2.00 \times 10^{-3} \times 0.100}$ $= 0.500\ \text{m s}^{-1}$
State your answer with the correct units.	$v = 0.500\ \text{m s}^{-1}$ The speed of the flow of the blood within the artery is 0.500 m s^{-1}.

Worked example: Try yourself 6.1.2

FLUID FLOW MEASUREMENT

The rate of fluid flow within a vessel can be measured using the induced emf when the fluid contains charged ions. A small magnet and a sensitive voltmeter calibrated to measure speed are used. This can be applied to measure fluid flow in small pipes. If the diameter of a particular small pipe is 1.00 cm, the strength of the magnetic field applied is 0.100 T and the measured emf is 0.500 mV, what is the speed of the fluid flow?



When v and B are not perpendicular, then the component of B perpendicular to the magnetic field needs to be used.

$B_{\perp} = B\cos\theta$

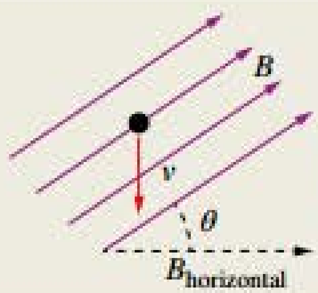


FIGURE 6.1.4 A horizontal rod is moving vertically at speed, v, through the Earth's magnetic field, which is at an angle of dip θ to the horizontal.

Worked example: 6.1.3

EMF ACROSS CONDUCTOR MOVING AT AN ANGLE TO THE MAGNETIC FIELD

A horizontal conducting rod of length 10.0 cm and facing east–west is dropped vertically where the Earth's magnetic field is 5.00×10^{-5} T at an angle of dip of 60.0°. Calculate the induced emf between its ends when it reaches a velocity of 20.0 m s^{-1}.

Thinking	Working
Identify the component of magnetic field that will cause an emf to be induced.	If the rod is falling vertically, it will not cut the vertical component of magnetic field. The horizontal component of the magnetic field needs to be found.
Calculate the horizontal component of the magnetic field.	$B_h = B\cos\theta$ $= 5.00 \times 10^{-5}\cos 60°$ $= 2.50 \times 10^{-5}$ T
Identify the quantities required and put them into SI units.	$\varepsilon = ?$ $\ell = 10.0$ cm $= 0.100$ m $v = 20.0$ m s^{-1} $B_h = 2.50 \times 10^{-5}$ T
Use the appropriate formula, substitute and calculate answer.	$\varepsilon = \ell vB$ $= 0.100 \times 20.0 \times 2.50 \times 10^{-5}$ $= 5.00 \times 10^{-5}$
State your answer with the correct units.	$\varepsilon = 5.00 \times 10^{-5}$ V

Worked example: Try yourself 6.1.3

EMF ACROSS CONDUCTOR MOVING AT AN ANGLE TO THE MAGNETIC FIELD

A car driving from east to west at $100\,km\,h^{-1}$ has a vertical aerial of length 50.0 cm on its roof where the Earth's magnetic field is 5.00×10^{-5} T at an angle of dip of 30.0°. Calculate the induced emf between its ends.

6.1 Review

SUMMARY

- An induced emf, ε, is produced by a changing magnetic flux in a process called electromagnetic induction.
- The induced emf in a straight conductor moving in a magnetic field, B, is given by $\varepsilon = \ell v B_{\perp}$, where $v \perp B$.

KEY QUESTIONS

1 Which of the following scenarios will *not* induce an emf in a long, straight conductor?
A A magnet is stationary alongside the conductor.
B A magnet is brought near the conductor.
C The conductor is brought into a magnetic field.
D The conductor is rotated within a magnetic field.

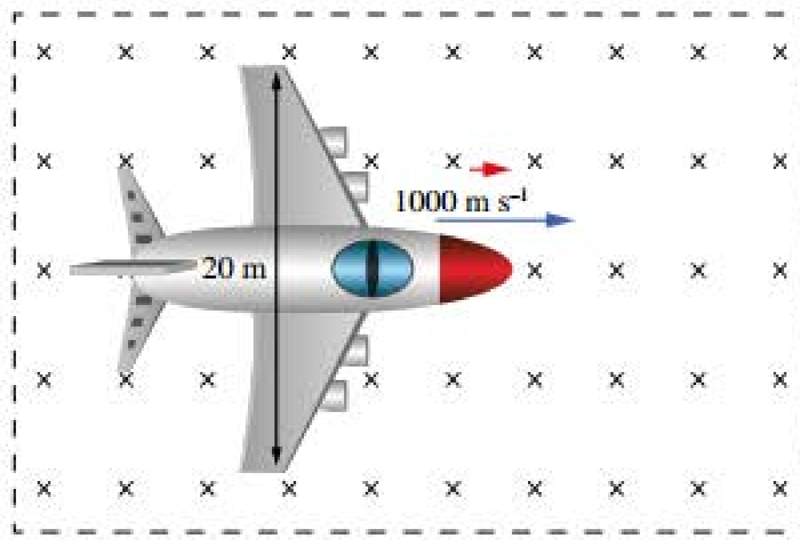
2 A metal rod 12.0 cm long is being moved at a speed of $0.150\,m\,s^{-1}$ perpendicular to a magnetic field, B. If the strength of the magnetic field is 0.800 T, calculate the induced emf in the rod.

3 A metal rod is 13.2 cm long. It generates an emf of 100 mV while moving perpendicular to a magnetic field of strength 0.900 T. Calculate the speed at which it is moving.

4 A metal rod generates an emf of 80.0 mV while moving at a speed of $1.60\,m\,s^{-1}$ perpendicular to a magnetic field of strength 0.500 T. Determine the length of the metal rod.

5 A rod of length 10.0 cm and very small diameter is held vertically and dropped downwards through a magnetic field of strength 0.800 T that is directed vertically upwards. If the rod is dropped from rest, what emf is induced across the rod at an instant 5.00 s later?

6 Calculate the magnitude of the induced emf between the ends of the wings of an aircraft whose wingspan is 20.0 m, given that the aircraft is moving at a speed of $1000\,m\,s^{-1}$ in the magnetic field of the Earth in a plane perpendicular to the lines of the field, where the flux density is 2.50×10^{-5} T.



7 A 0.750 m aerial is placed vertically on the roof of a car and is insulated from the car. The car is then driven east at $60.0\,km\,h^{-1}$, in a region near Perth where the horizontal component of the Earth's magnetic field is 2.50×10^{-5} T. Calculate the emf induced across the aerial.

The following information relates to questions 8–10.

A Boeing 747 with a wingspan of 60.0 m flies due south at a constant altitude at $260\,m\,s^{-1}$ in an area where the vertical component of the Earth's magnetic field is 4.00×10^{-5} T and the horizontal component of the Earth's magnetic field is 2.00×10^{-5} T.

8 Calculate the emf induced between the wing tips while the aircraft flies horizontally.

9 The aircraft now dives at 10.0° to the horizontal, maintaining its speed. Calculate the induced emf due to the horizontal component of its motion.

10 Calculate the induced emf between the wings due to the vertical component of its motion and hence the total induced emf.

6.2 Induced emf from a changing magnetic flux

Faraday's early experiments centred largely on investigating electromagnetic induction in coils, or multiple loops, of wire. Faraday found that if a magnet is quickly moved into a coil, an emf is induced that causes a current to flow in the complete circuit containing the coil. If the magnet is removed, then a current flows in the coil in the opposite direction. Alternatively, if the magnet is held steady and the coil is moved in such a way that changes the magnetic flux, then once again an emf is induced and an electric current flows in the complete circuit. It doesn't matter whether the coil or the magnet is moved—it is a *change* in flux that is required to induce the emf (Figure 6.2.1). This discovery led Faraday to his law of induction. Faraday's law of induction is the focus of this section.

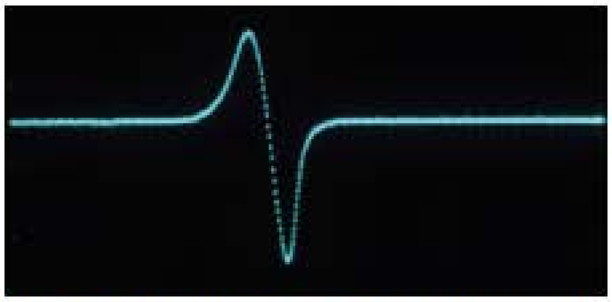


FIGURE 6.2.1 Oscilloscope trace from an electric coil, showing the voltage across the coil as a magnet is dropped through it.

MAGNETIC FLUX

To be able to develop ideas about the change in a magnetic field that induces an emf, which can then create (or induce) a current, it is useful to be able to describe the 'amount of magnetic field'. This amount of magnetic field is referred to as the **magnetic flux**, a scalar quantity, denoted by the symbol Φ. Faraday pictured a magnetic field as consisting of many lines of force. The density of the lines represents the strength of the magnetic field. Magnetic flux can be related to the total number of these lines that pass within a particular area. A strong magnetic field acting over a small area can produce the same amount of magnetic flux as a weaker field acting over a larger area, as shown in Figure 6.2.2. For this reason, magnetic field strength, B, is also referred to as **magnetic flux density**. B can be thought of as being proportional to the number of magnetic field lines per unit area perpendicular to the magnetic field. The magnetic flux will be at a maximum when the area examined is perpendicular to the magnetic field and zero when the area being examined is parallel to the magnetic field.

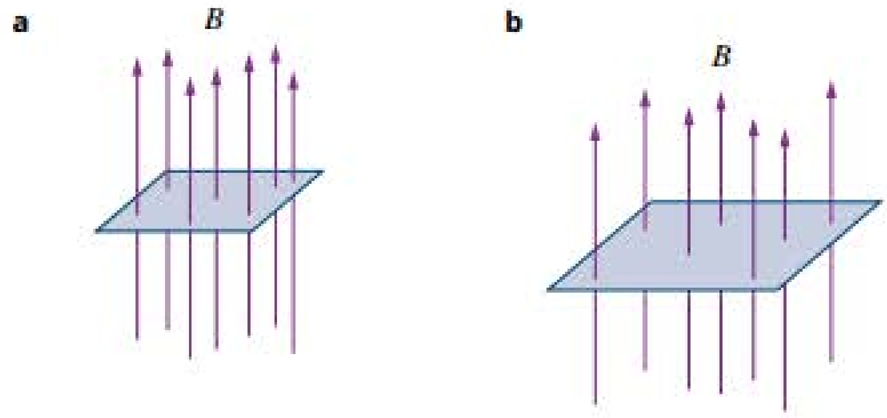


FIGURE 6.2.2 Magnetic flux. A strong magnetic field acting over a small area (a) will have the same magnetic flux as a weaker magnetic field acting over a larger area (b).

Based on this, magnetic flux is defined as the product of the strength of the magnetic field, B, and the area of the field perpendicular to the field lines.

$$\Phi = BA_{\perp}$$

where Φ is the magnetic flux. The unit for magnetic flux is the weber, abbreviated to Wb, where $1\,\text{Wb} = 1\,\text{T}\,\text{m}^{-2}$

B is the strength of the magnetic field in tesla (T)

A is the area perpendicular to the magnetic field, measured in square metres (m^2)

The subscript $_{\perp}$ is included in the formula to indicate that the plane of the area referred to is perpendicular to the magnetic field.

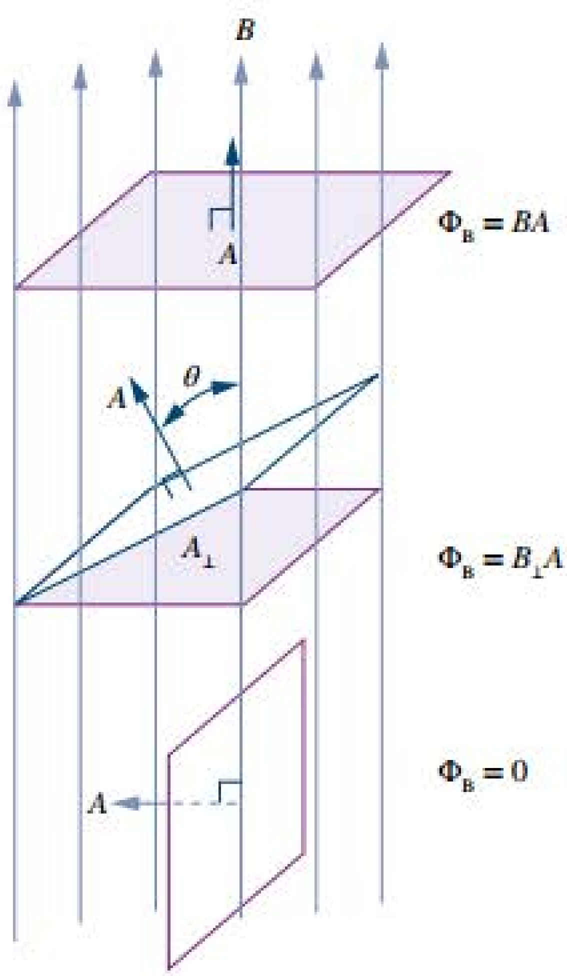


FIGURE 6.2.3 The magnetic flux is the strength of the magnetic field, B, multiplied by the area perpendicular to the magnetic field, given by $A\cos\theta$ and shown as the shaded areas in the diagrams above.

Since it is the area perpendicular to the magnetic field, the angle between the magnetic field and the area through which the field passes will obviously affect the amount of magnetic flux. As the angle increases or decreases from 90°, the amount of magnetic flux will decrease until it reaches zero, when the area under consideration is parallel to the magnetic field. Referring to Figure 6.2.3, the relationship between the amount of magnetic flux and the angle θ to the field is:

$$\Phi = BA\cos\theta$$

It is important to note that θ is not the angle between the plane of the area, as shown by the arrow labelled A in Figure 6.2.3, and the magnetic field. Rather, it is the angle between a normal to the area and the direction of the magnetic field; hence the use of $\cos\theta$. When the area is at right angles to the magnetic field, the angle θ between the normal and the field is 0° and $\cos 0° = 1$ (top diagram in Figure 6.2.3). When the area is parallel to the magnetic field, the angle θ between the normal and the field is 90° and $\cos 90° = 0$ (bottom diagram in Figure 6.2.3).

Worked example 6.2.1

MAGNETIC FLUX

A student places a horizontal square coil of wire of side length 5.00 cm into a uniform vertical magnetic field of 0.100 T. Calculate the magnetic flux 'threading' the coil.

Thinking	Working
Calculate the area of the coil perpendicular to the magnetic field.	side length = 5.00 cm = 0.0500 m area of the square = $(0.0500\,\text{m})^2$ = $0.002\,50\,\text{m}^2$
Calculate the magnetic flux.	$\Phi = B_\perp A$ $= 0.100 \times 0.002\,50$ $= 0.000\,250$ Wb
State the answer in an appropriate form.	$\Phi = 2.50 \times 10^{-4}$ Wb or 0.250 mWb

Worked example: Try yourself 6.2.1

MAGNETIC FLUX

A student places a horizontal square coil of wire of side length 4.00 cm into a uniform vertical magnetic field of 0.0500 T. Calculate the magnetic flux 'threading' the coil.

Note that in Worked example 6.2.1 an area of $5.00\,\text{cm} \times 5.00\,\text{cm} = 25.0\,\text{cm}^2$ was considered, and this corresponds to $0.002\,50\,\text{m}^2$ or $25.0 \times 10^{-4}\,\text{m}^2$, so:

$$1\,\text{cm}^2 = 1 \times 10^{-4}\,\text{m}^2$$

FACTORS AFFECTING INDUCED EMF

Faraday quantitatively investigated the factors affecting the size of the emf induced in a coil. First, emf will be induced by a change in the magnetic field. A simple example of this is to witness the emf induced when a magnet is brought towards or moved away from a wire coil. The greater the change in magnetic field, the greater the induced emf will be.

However, it is not only a change in the strength of a magnetic field, B, that induces an emf. It was noticed that an emf can be induced by changing A, the area perpendicular to the magnetic field through which the magnetic field lines pass, while keeping B constant. An example of this is to witness the emf induced when a wire coil is rotated in the presence of a fixed magnetic field. This discovery indicates that the requirement for an induced emf is to have a *changing magnetic flux*, Φ.

Finally, Faraday discovered that the faster the change in magnetic flux, the greater the induced emf. This can be seen in the oscilloscope trace of a magnet falling through a coil as shown in the Figure 6.2.1 (page 187). The magnet is accelerated by gravity as it drops through the coil. Hence, the peak emf induced when the magnet first enters the coil at a relatively lower speed is noticeably less than the peak emf when the magnet leaves the coil at a faster speed. Thus, it is the *rate of change* of magnetic flux that determines the induced emf.

FARADAY'S LAW OF INDUCTION

Faraday's investigations led him to conclude that the average emf induced in a conducting loop in which there is a changing magnetic flux is proportional to the rate of change of flux.

This is now known as **Faraday's law** of induction and it is one of the basic laws of electromagnetism.

Magnetic flux is defined as $\Phi = BA_{\perp}$.

If the flux through N turns (or loops) of a coil changes from Φ_1 to Φ_2 during a time t, then the average induced emf during this time will be:

$$\varepsilon = -N\frac{(\Phi_2 - \Phi_1)}{t}$$

and if the change in magnetic flux $\Phi_2 - \Phi_1 = \Delta\Phi$, then

$$\varepsilon = -N\frac{\Delta\Phi}{t} = -N\frac{\Delta(BA_{\perp})}{t}$$

The negative sign is placed there as a reminder of the direction of the induced emf. This is discussed further in the next section 'Lenz's law'. When calculating the magnitude of the emf, ignore the negative sign or any negative quantities in a calculation.

PHYSICSFILE

Musicians rely on Faraday's law of induction

Microphones

The so-called 'dynamic' microphone uses a tiny coil attached to a diaphragm, which vibrates with the sound (Figure 6.2.4). The coil vibrates within the magnetic field of a permanent magnet, thus producing an induced emf that varies with the original sound.

Electric guitars

Electric guitars use pickups that transform the mechanical energy of a vibrating guitar string into electrical energy by way of electromagnetic induction (Figure 6.2.5). The pickup uses permanent magnets and pole pieces to form a steady magnetic field near the individual steel guitar strings. The guitar string becomes magnetised when it is above its pole piece; when it is plucked and vibrates, the magnetic field is no longer constant—it changes at the same frequency as the vibration. Coils, with thousands of turns, are fitted around the poles to pick up an induced current and voltage that can then be sent to an amplifier.

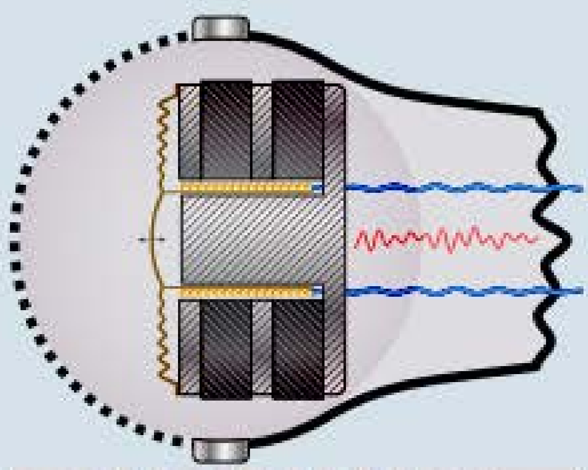


FIGURE 6.2.4 The electrodynamic microphone (acoustic to electric).

FIGURE 6.2.5 Close-up view of strings on a jazz guitar.

If the ends of the coil are connected to an external circuit, then a current, I, will flow. The magnitude of the current is found using Ohm's law:

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

where R is the resistance and V is the emf of the coil.

A coil not connected to a circuit will act like a battery not connected to a circuit. There will still be an induced emf but no current will flow.

Worked example 6.2.2

INDUCED EMF IN A COIL

A student winds a coil of area 40.0 cm^2 with 20 turns and places it horizontally in a vertical uniform magnetic field of 0.100 T.

a Calculate the magnetic flux perpendicular to the coil.	
Thinking	**Working**
Identify the quantities to calculate the magnetic flux through the coil and convert to SI units where required.	$\Phi = BA_{\perp}$ $B = 0.100\,\text{T}$ $A = 40.0\,\text{cm}^2 = 40.0 \times 10^{-4}\,\text{m}^2$
Calculate the magnetic flux and give your answer with appropriate units.	$\Phi = BA_{\perp} = 0.100 \times 40.0 \times 10^{-4}$ $= -4.00 \times 10^{-4}\,\text{Wb}$
b Calculate the magnitude of the average induced emf in the coil when the coil is removed from the magnetic field in a time of 0.500 s.	
Thinking	**Working**
Identify the quantities for determining the induced emf. Ignore the negative sign.	$\varepsilon = -N\frac{\Delta\Phi}{t}$ $N = 20$ turns $\Delta\Phi = \Phi_2 - \Phi_1$ $= 0 - 4.00 \times 10^{-4}$ $= -4.00 \times 10^{-4}\,\text{Wb}$ $t = 0.500\,\text{s}$
Calculate the magnitude of the average induced emf.	$\varepsilon = -N\frac{\Delta\Phi}{t}$ $= -20 \times \frac{-4.00 \times 10^{-4}}{0.500}$ $= 0.0160\,\text{V}$

Worked example: Try yourself 6.2.2

INDUCED EMF IN A COIL

A student winds a coil of area 50.0 cm^2 with 10 turns and places it horizontally in a vertical uniform magnetic field of 0.100 T.

a Calculate the magnetic flux perpendicular to the coil.

b Calculate the magnitude of the average induced emf in the coil when the coil is removed from the magnetic field in a time of 1.00 s.

Worked example 6.2.3

NUMBER OF TURNS IN A COIL

A coil of cross-sectional area $1.00 \times 10^{-3}\,m^2$ experiences a change in the strength of a magnetic field from 0.00 to 0.200 T over 0.500 s. If the magnitude of the average induced emf is measured as 0.100 V, how many turns must be on the coil?

Thinking	Working
Identify the quantities needed to calculate the magnetic flux through the coil when in the presence of the magnetic field and convert to SI units where required.	$\Phi = BA_{\perp}$ $B = 0.200\,T$ $A = 1.00 \times 10^{-3}\,m^2$
Calculate the magnetic flux when in the presence of the magnetic field.	$\Phi = BA_{\perp}$ $= 0.200 \times 1.00 \times 10^{-3}$ $= 2.00 \times 10^{-4}\,Wb$
Identify the quantities from the question required to complete Faraday's law.	$\varepsilon = -N\frac{\Delta\Phi}{t}$ $N = ?$ $\Delta\Phi = \Phi_2 - \Phi_1$ $= 2.00 \times 10^{-4} - 0$ $= 2.00 \times 10^{-4}\,Wb$ $\Delta t = 0.500\,s$ $\varepsilon = 0.100\,V$
Rearrange Faraday's law and solve for the number of turns on the coil. Ignore the negative sign.	$\varepsilon = -N\frac{\Delta\Phi}{t}$ $N = -\frac{\varepsilon t}{\Delta\Phi}$ $= \frac{0.100 \times 0.500}{2.00 \times 10^{-4}}$ $= 250$ turns

Worked example: Try yourself 6.2.3

NUMBER OF TURNS IN A COIL

A coil of cross-sectional area $2.00 \times 10^{-3}\,m^2$ experiences a change in the strength of a magnetic field from 0.00 to 0.200 T over 1.00 s. If the magnitude of the average induced emf is measured as 0.400 V, how many turns must be on the coil?

6.2 Review

SUMMARY

- Magnetic flux is defined as the product of the strength of the magnetic field, B, and the area of the field perpendicular to the field lines, i.e. $\Phi = BA_{\perp}$.
- The amount of magnetic flux varies with the angle of the field to the area under investigation. It is a maximum when the normal to the area is parallel (0°) to the magnetic field, and zero when the normal to the area is perpendicular (90°) to the field, i.e. $\Phi = BA\cos\theta$.
- The unit for magnetic flux is the weber, Wb; 1 Wb = 1 T m^2.
- The emf induced in a conducting loop in which there is a changing magnetic flux is proportional to the negative rate of change of flux.
- $\varepsilon = -N\frac{\Delta\Phi}{t}$, which is Faraday's law of induction.
- The negative sign in Faraday's law indicates direction (see Section 6.3 Lenz's law). For questions involving only magnitudes, you should not use the negative sign or any negative quantities.

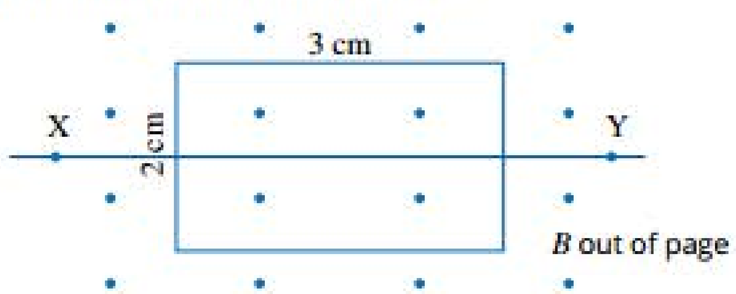
KEY QUESTIONS

1 A student places a 4.00 cm square coil of wire parallel to a uniform vertical magnetic field of 0.0500 T. Calculate the magnetic flux 'threading' the coil.

2 A circular coil of wire, of radius 5.00 cm, is perpendicular to a region of uniform magnetic field, B = 1.60 mT. Calculate the amount of magnetic flux passing through the loop.

The following information relates to questions 3–5.

A single rectangular wire loop is located with its plane perpendicular to a uniform magnetic field of 2.00 mT, directed out of the page, as shown below. The loop is free to rotate about a horizontal axis XY.



3 Calculate the amount of magnetic flux threading the loop in this position.

4 The loop is rotated about the axis XY, through an angle of 90.0°, so that its plane becomes parallel to the magnetic field. Calculate the amount of flux threading the loop in this new position.

5 If the loop completes one-quarter of a rotation in 40.0 ms, determine the average induced emf in the loop.

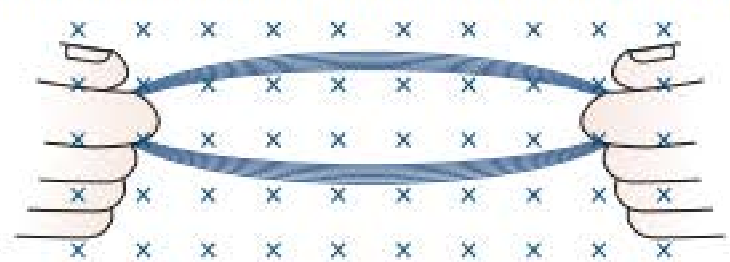
The following information relates to questions 6 and 7.

A coil of 500 turns, each of area 10.0 cm^2, is wound around a square frame. The plane of the coil is initially parallel to a uniform magnetic field of 80.0 mT. The coil is then rotated through an angle of 90° so that its plane becomes perpendicular to the field. The rotation is completed in 20.0 ms.

6 Calculate the average emf induced in each turn during this time.

7 Calculate the average emf induced in the coil in Question 6 during the time the coil rotated.

8 A student stretches a flexible wire coil of 30 turns and places it in a uniform magnetic field of strength 5.00 mT, directed into the page, as shown. While it is in the field, the student allows the coil to regain its original shape. In doing so, the area of the coil changes at a constant rate from 250 cm^2 to 50.0 cm^2 in 0.500 s.



Calculate the average emf induced in the coil during this time.

9 A wire coil consisting of a single turn is placed perpendicular to a magnetic field that experiences a decrease in strength of 0.100 T in 0.0500 s. If the emf induced in the coil is 0.0200 V, determine the area of the coil.

10 A wire coil consisting of 100 turns with an area of 50.0 cm^2 is placed inside a vertical magnetic field of strength 0.400 T, and then rotated about a horizontal axis. For each quarter turn, the average emf induced in the coil is 1600 mV. Calculate the time taken for a quarter turn of the coil.

6.3 Lenz's law

In the previous section Faraday's law of induction was introduced, including the equation:

$$\varepsilon = -N\frac{\Delta\Phi}{t}$$

The negative sign is there to remind you in which direction the induced emf acts—that is, in which direction current flows in a complete circuit as a result of the induced emf.

Lenz's law, which is the focus of this section, is a common way of understanding how electromagnetic induction obeys the principles of conservation of energy and explains the direction of the induced emf. It is named after Heinrich Lenz, whose research put a definite direction to the current created by the induced emf resulting from a changing magnetic flux.

Understanding the direction of the current resulting from an induced emf and how it is produced has allowed electromagnetic induction to be used in a vast array of devices that have transformed modern society, in particular in electrical generators. A metal detector is another example of a device that uses Lenz's law (Figure 6.3.1).

FIGURE 6.3.1 A diver using a metal detector. If a metal object is found underneath the coil of the detector, an emf will be induced, which creates a current that will affect the original current. The direction of the induced current is predicted by using Lenz's law.

DIRECTION OF AN INDUCED EMF

Lenz's law states that an induced emf always gives rise to a current whose magnetic field will oppose the original change in flux.

Figure 6.3.2 applies the law to the relative motion between a magnet and a single coil of wire. Moving the magnet towards or away from the coil will induce an emf in the coil, as there is a change in flux. The induced emf will produce a current in the coil, and this induced current will then produce its own magnetic field that interacts with the permanent magnet to oppose its motion. It is worth noting that Lenz's law is a necessary consequence of the law of conservation of energy: if Lenz's law were not true then the new magnetic field created by a changing flux would encourage that change, which would have the effect of adding energy to the universe.

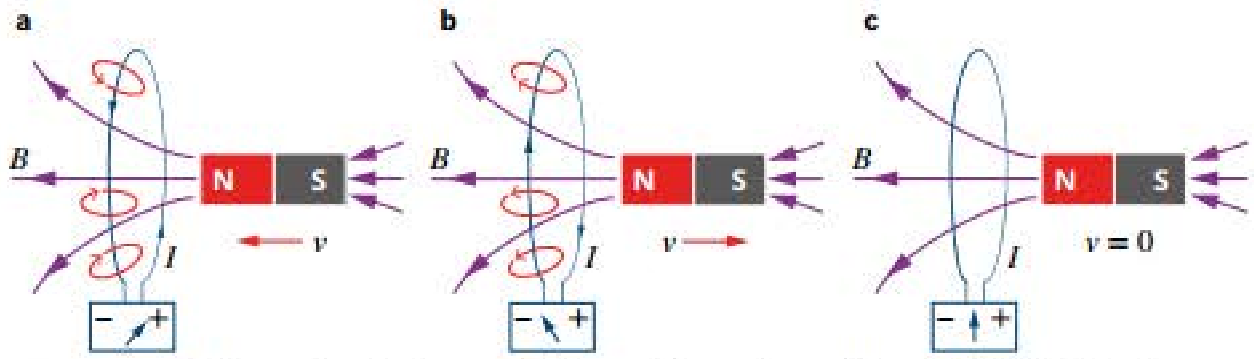


FIGURE 6.3.2 (a) The north end of a magnet is brought towards a coil from right to left, inducing a current that flows anticlockwise. (b) Pulling the north end of the magnet away from the coil from left to right induces a current in a clockwise direction. (c) Holding the magnet still creates no change in flux and hence no induced current.

Applying Lenz's law, the magnetic field created by the induced current will oppose the approaching field caused by the movement of the magnet. When the north end of a magnet is brought towards the loop from the right, as shown in Fig 6.3.2(a), the magnetic flux from right to left through the loop increases. The induced emf produces a current that flows anticlockwise around the loop when viewed from the right. The magnetic field created by this current, shown by the little circles around the wire, is directed from left to right through the loop. It opposes the magnetic field of the approaching magnet. Also note that it is opposing the change in the magnet's flux through the loop by attempting to oppose the increasing flux.

If the magnet is moved away from the loop, as in Figure 6.3.2(b), the magnetic flux from right to left through the loop decreases. The induced emf produces a clockwise current when viewed from the right. This creates a magnetic field that is directed from right to left through the loop. This field is in the same direction as the original magnetic field of the retreating magnet, which causes an attractive force. Also note that it is opposing the change in the magnet's flux through the loop by attempting to replace the declining flux.

And when the magnet is held stationary, as in Figure 6.3.2(c), there is no change in flux to oppose and so no current is induced.

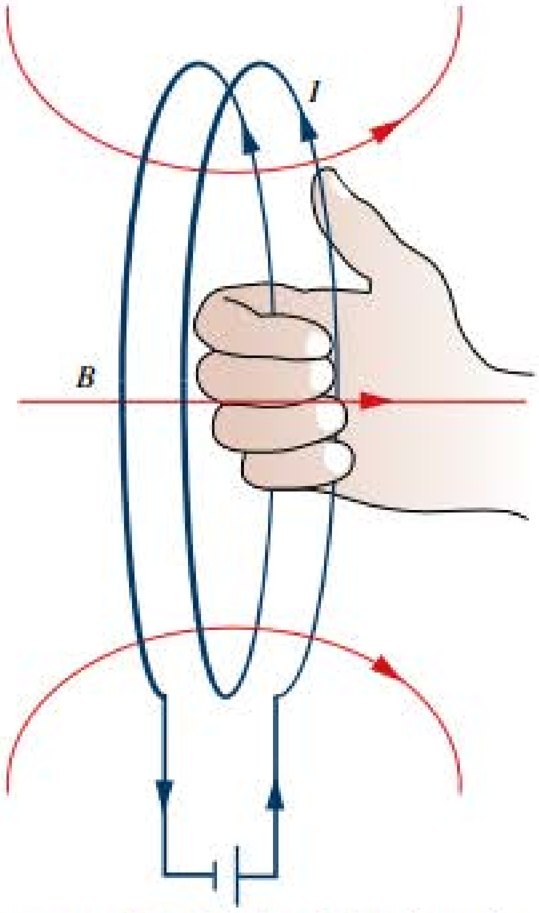


FIGURE 6.3.3 The right-hand grip rule can be used to determine the direction of a magnetic field from a current or vice versa. Your thumb points in the direction of the conventional current in the wire and your curled fingers indicate the direction of the magnetic field through the coil.

Right-hand grip rule and induced current direction

The right-hand grip rule (see page 156) can be used to find the direction of the induced current. Keep in mind that the current must create a magnetic field that opposes the change in flux due to the relative motion of the magnet and conductor. Curl your fingers around the wire by moving your fingers through the loop in the direction of the field that is *opposing* the change and your thumb will then indicate the direction of the conventional current, as shown in Figure 6.3.3.

There are three distinct steps to determine the induced current direction according to Lenz's law:

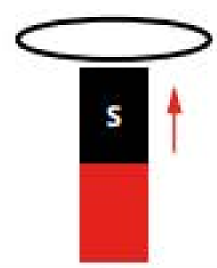
1. What is the change that is happening?
2. What will *oppose* the change and/or restore the original conditions?
3. What must the current direction be to match this opposition?

These steps will be further examined using Worked example 6.3.1.

Worked example 6.3.1

INDUCED CURRENT IN A COIL FROM A PERMANENT MAGNET

The south pole of a magnet is brought upwards towards a horizontal coil initially held above it. In which direction will the induced current flow in the coil?

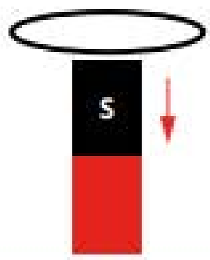


Thinking	Working
Consider the direction of the change in magnetic flux.	The magnetic field direction from the magnet will be downwards towards the south pole. The downwards flux from the magnet will increase as the magnet is brought closer to the coil. So the change in flux is increasing downwards.
What will oppose the change in flux?	The induced magnetic field that opposes the change would act upwards.
Determine the direction of the induced current required to oppose the change.	In order to oppose the change, the current direction would be anticlockwise when viewed from above (using the right-hand grip rule).

Worked example: Try yourself 6.3.1

INDUCED CURRENT IN A COIL FROM A PERMANENT MAGNET

The south pole of a magnet is moved downwards away from a horizontal coil held above it. In which direction will the induced current flow in the coil?



Worked example 6.3.2

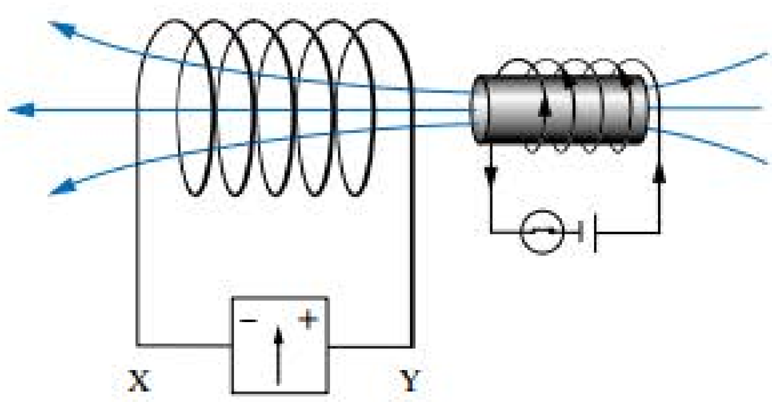
INDUCED CURRENT IN A COIL FROM AN ELECTROMAGNET

Instead of using a permanent magnet to change the flux in the loop in Worked example 6.3.1, an electromagnet (on the right, in the diagram below) could be used. What is the direction of the current induced in the solenoid when the electromagnet is:

(i) switched on?

(ii) left on?

(iii) switched off?



Thinking	Working
Consider the direction of the change in magnetic flux for each case.	**(i)** Initially there is no magnetic flux through the solenoid. When the electromagnet is switched on, it creates a magnetic field directed to the left. So the change in flux through the solenoid is increasing to the left. **(ii)** While the current in the electromagnet is steady, the magnetic flux through the solenoid is constant and the flux is not changing. **(iii)** In this case, initially there is a magnetic flux through the solenoid from the electromagnet directed to the left. When the electromagnet is switched off, there is no longer a magnetic flux through the solenoid. So the change in flux through the solenoid is decreasing to the left.
What will oppose the change in flux for each case?	**(i)** The magnetic field that opposes the change in flux through the solenoid is directed to the right. **(ii)** There is no change in flux and so there will be no opposition needed and no magnetic field created by the solenoid. **(iii)** The magnetic field that opposes the change in flux through the solenoid is directed to the left.
Determine the direction of the induced current required to oppose the change for each case.	**(i)** In order to oppose the change, the current will flow through the solenoid in the direction from X to Y (or through the meter from Y to X), using the right-hand grip rule. **(ii)** There will be no induced emf or current in the solenoid. **(iii)** In order to oppose the change, the current will flow through the solenoid in the direction from Y to X (or through the meter from X to Y), using the right-hand grip rule.

Worked example: Try yourself 6.3.2

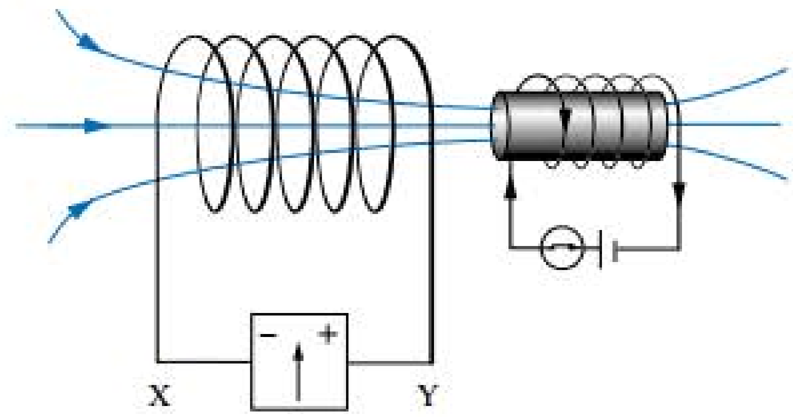
INDUCED CURRENT IN A COIL FROM AN ELECTROMAGNET

What is the direction of the current induced in the solenoid when the electromagnet is:

(i) switched on?

(ii) left on?

(iii) switched off?



INDUCED CURRENT DIRECTION BY CHANGING AREA

It is very important to note that an induced emf is created while there is a change in flux, no matter how that change is created. As magnetic flux $\Phi = BA_{\perp}$, a change can be created by any method that causes a relative change in the strength of the magnetic field, *B*, and/or the area of the coil perpendicular to the magnetic field. So an induced emf can be created in three ways:

- by changing the strength of the magnetic field
- by changing the area of the coil within the magnetic field
- by changing the orientation of the coil with respect to the direction of the magnetic field.

Figure 6.3.4 illustrates an example of the direction of an induced current that results during a decrease in the area of a coil.

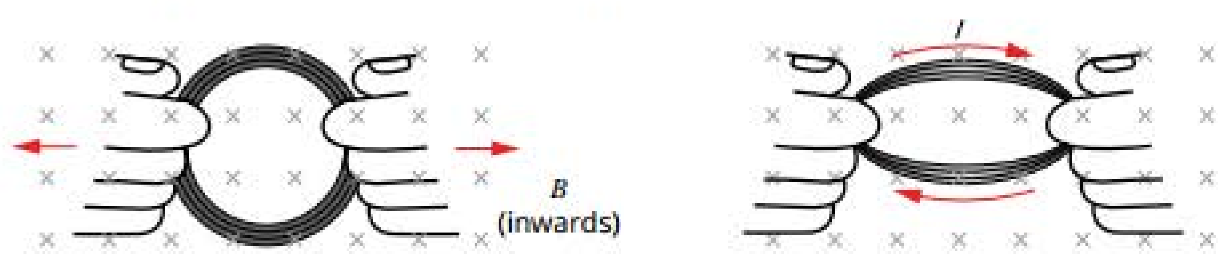


FIGURE 6.3.4 Inducing a current by changing the area of a coil. The amount of flux (the number of field lines) through the coil is reduced and an emf is therefore induced during the time that the change is taking place. The current flows in a direction that creates a field to oppose the reduction in flux into the page.

As the area of the coil decreases due to its changing shape, the flux through the coil (which is directed into the page) also decreases. Applying Lenz's law, the direction of the induced current would oppose this change and will be such that it acts to increase the magnetic flux through the coil into the page. Using the right-hand grip rule, a current would therefore flow in a clockwise direction while the area is changing.

In Figure 6.3.5, the coil is being rotated within the magnetic field. The effect is the same as reducing the area. The amount of flux flowing through the coil is reduced as the coil changes from being perpendicular to the field to being parallel to the field. An induced emf would be created while the coil is being rotated. This becomes particularly important in determining the current direction in a generator.

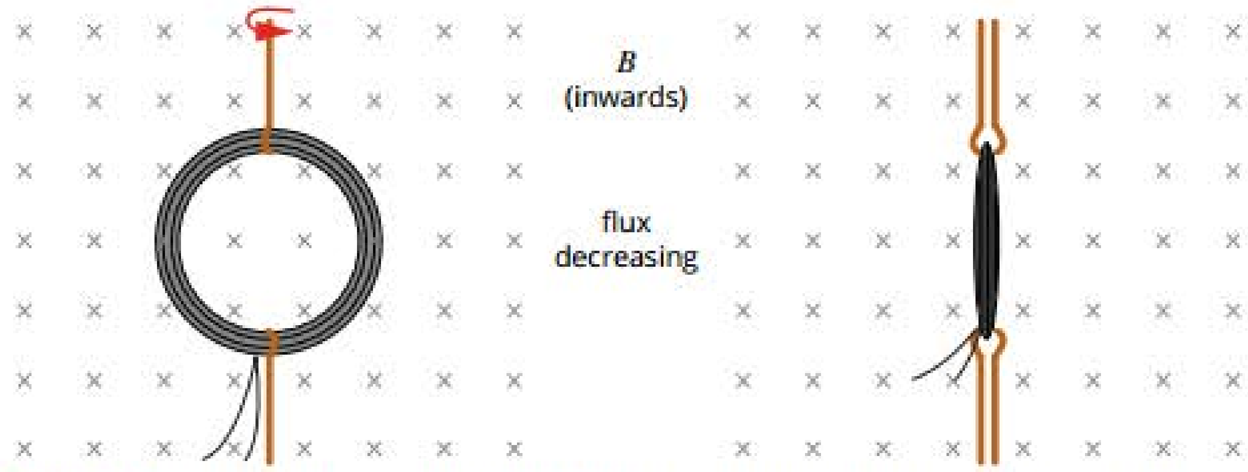
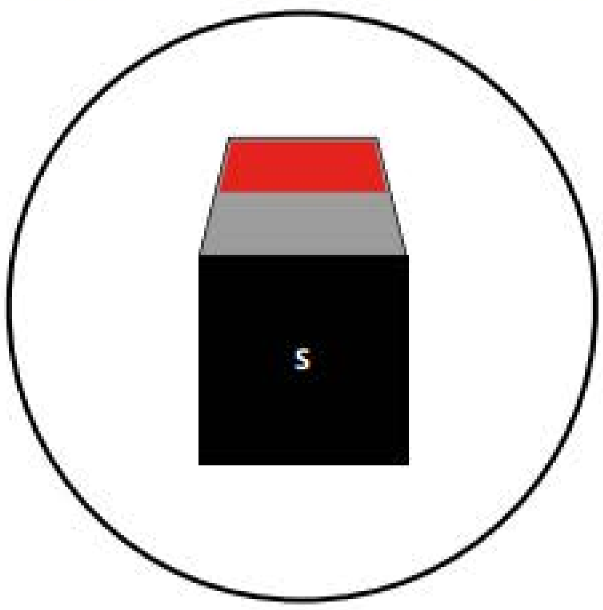


FIGURE 6.3.5 Changing the orientation of a coil within a magnetic field by rotating it reduces the amount of flux through the coil and so induces an emf in the coil while it is being rotated.

Worked example 6.3.3

FURTHER PRACTICE WITH LENZ'S LAW

The north pole of a magnet is moving towards a coil, into the page (the south pole is shown at the top looking down). In which direction will the induced current flow in the coil while the magnet is moving towards the coil?

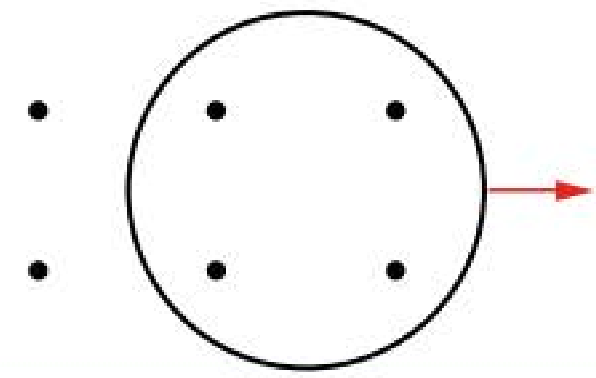


Thinking	Working
Consider the direction of the change in magnetic flux.	The magnetic field direction from the magnet will be away from the north pole, into the page. The flux from the magnet will increase as the magnet is brought closer to the coil. So the change in flux is increasing into the page.
What will oppose the change in flux?	The magnetic field that opposes the change would act out of the page.
Determine the direction of the induced current required to oppose the change.	In order to oppose the change, the current direction would be anticlockwise when viewed from above (using the right-hand grip rule).

Worked example: Try yourself 6.3.3

FURTHER PRACTICE WITH LENZ'S LAW

A coil is moved to the right and out of a magnetic field that is directed out of the page. In which direction will the induced current flow in the coil while the magnet is moving?



EDDY CURRENTS

Lenz's law is important for many practical applications such as metal detectors, induction stoves and regenerative braking. These all rely on an **eddy current**, which is a circular electric current induced within a conductor by a changing magnetic field.

Applying Lenz's law, an eddy current will be in a direction that creates a magnetic field that opposes the change in magnetic flux that created it. Thus eddy currents can be used to apply a force that opposes the source of the motion of an external magnetic field. For example, if a metal plate is dragged out of a magnetic field, an eddy current will form within the plate that opposes the change in flux through the area of the plate, and thus opposes the motion of the plate itself due to the interaction of the magnetic fields (Figure 6.3.6).

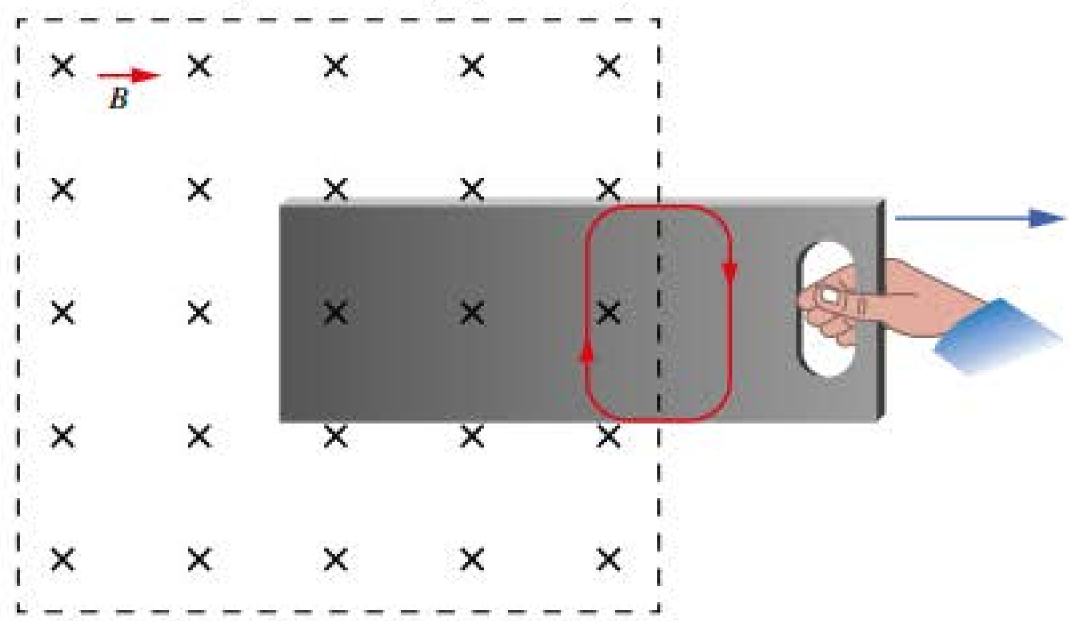


FIGURE 6.3.6 As the metal plate is moved towards the right, out of the magnetic field that is directed into the page, an eddy current forms in a clockwise direction. This eddy current resists the motion of the plate.

This is the basis of regenerative braking, in which the drag of the opposing magnetic field is utilised as a braking force. An eddy current flowing through a conductor with some resistance will also lose energy to the conductor by heating it. This makes eddy currents useful for an induction stovetop, but a potentially major source of energy loss within an alternating current (AC) generator, motor or transformer. Laminated cores with insulating material between the thin layers of iron are used in these applications to reduce the overall conductivity and suppress eddy currents.

The Earth's magnetic field is also a result of eddy currents. The energy that drives the Earth's dynamo comes from the enormous heat produced by radioactive decay deep in the Earth's core. The heat causes huge swirling convection currents of molten iron in the outer core. These convection currents of molten iron act rather like a spinning disc. They are moving in the Earth's magnetic field and so eddy currents are induced in them. These eddy currents produce their own magnetic fields, which contribute to the Earth's field, and so a heat-driven process called the geodynamo is created.

PHYSICS IN ACTION

Induction stoves

In contrast to conventional gas or electric stoves, which heat via radiant heat from a hot source, an **induction hotplate** heats via an eddy current in the metal pot in which the food is being cooked. A coil of copper wire is placed within the cooktop (Figure 6.3.7). The AC electricity supply produces a changing magnetic field in the coil. This induces an eddy current in the conductive metal pot. The resistance of the metal in the pot, in which the eddy current flows, transforms electrical energy into heat and cooks the food.

Although induction cooktops have only reached the domestic market in relatively recent times, the first patents for induction cookers where issued in the early 1900s. They have significant advantages over traditional electric cooktops in that they allow instant control of cooking power (similar to gas burners), they lose less energy through ambient heat loss and heating time, and they have a lower risk of burn injuries. Overall, the heating efficiency of an induction cooktop is about 12% better than traditional electric cooktops and twice that of gas.

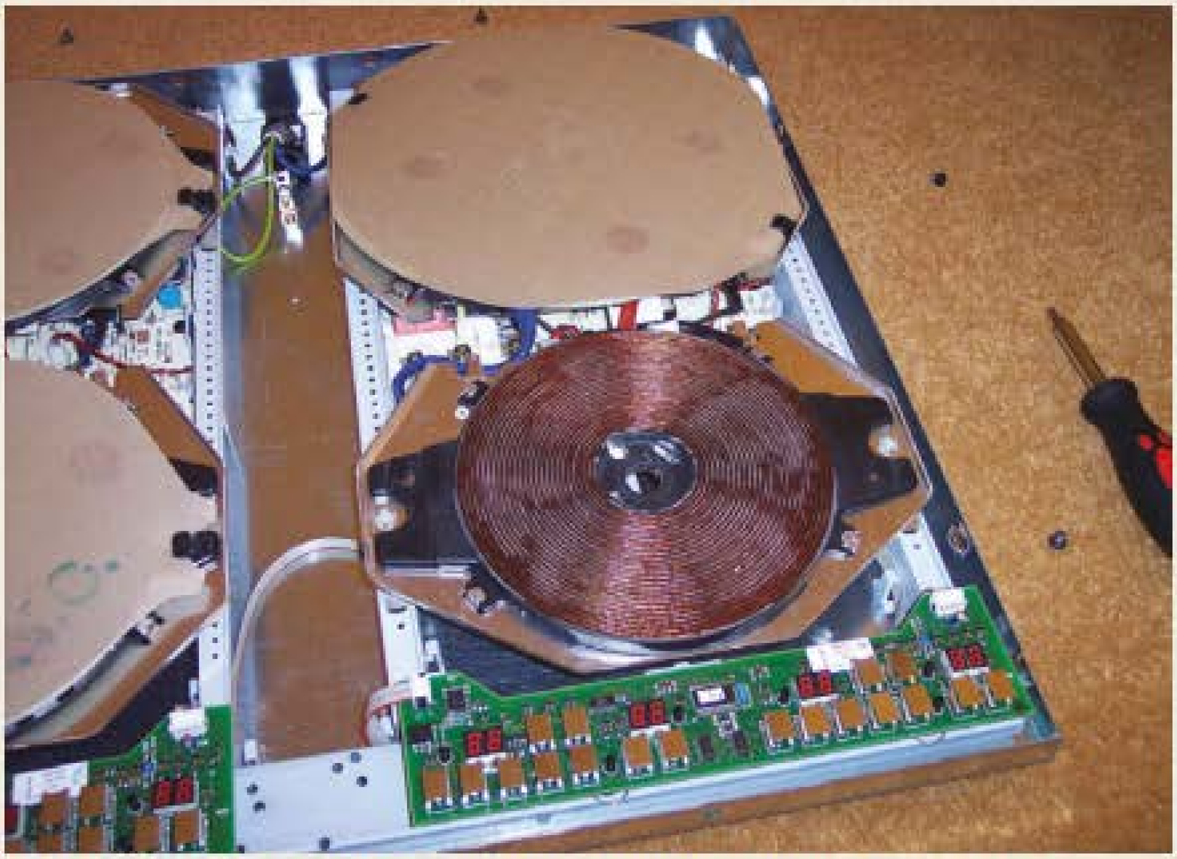


FIGURE 6.3.7 The coil of an induction zone within an induction cooktop. The large copper coil creates an alternating magnetic field.

EXTENSION

Superconductors and superconducting magnets

Technological breakthroughs have often led to advances in physics. This was the case in 1908 when Kamerlingh Onnes, at the University of Leiden in the Netherlands, succeeded in liquefying helium. Helium liquefies at 4.2 K (−268.9°C). It was known that the electrical resistance of metals decreases as they cool, so one of the first things that Onnes and his assistant did was to measure the resistance of some metals at these very low temperatures.

Onnes was hoping to find that as the temperature of mercury dropped towards absolute zero its resistance would also gradually drop towards zero. What they found, however, was a complete surprise. At 4.2 K its resistance vanished completely!

Onnes coined the word 'superconductivity' to describe this phenomenon. Soon he found that some other metals also became superconducting at extremely low temperatures: lead at 7.2 K and tin at 3.7 K, for example. Curiously, metals such as copper and gold, which are very good conductors at normal temperatures, do not become superconducting at all. Onnes was awarded the 1913 Nobel Prize in Physics for his work in low-temperature physics.

Much of the great promise of superconductivity has to do with the magnetic properties of superconductors. In a superconductor an induced current does not fade away. As the resultant field opposes the changing flux, the magnet is repelled. This gives rise to the 'magnetic levitation' effect that is by now well known (Figure 6.3.8). On a large scale this could perhaps one day provide a virtually frictionless maglev (magnetic levitation) train.

Unfortunately, the early superconducting metals lost their superconductivity in moderate magnetic fields of about 0.1 T or less. However, in the 1940s it was found that some alloys of elements such as niobium had higher 'critical temperatures' and, more particularly, retained their properties in stronger magnetic fields. By 1973 the niobium–germanium alloy Nb_3Ge held the record with a critical temperature of 23.2 K in a magnetic field of 38 T, an extremely strong field.

In 1986 an entirely new and exciting class of superconductors was discovered. Georg Bednorz and Karl Müller, working in Switzerland, found that compounds of some rare earth elements and copper oxide had considerably higher critical temperatures. They received the 1987 Nobel Prize in Physics for their work.

These new 'warm superconductors' are ceramic materials made by powdering and baking the metal compounds. Most ceramics are insulators; it was a combination of good science and inspired guesswork that led Müller to try such unlikely candidates for superconductivity. So far, the record is held by bismuth and thallium oxides with a critical temperature around 125 K—still rather cold, but significantly above the temperature of readily available liquid nitrogen (77 K).

Superconductivity, particularly in the newer materials, is still not fully understood. It can really only be discussed in terms of quantum physics, but one rather picturesque way of thinking about it is that electrons pair up and 'surf' electrical waves set up by each other in the crystal lattice of the material.

The promise of superconductivity is enormous: low friction transport, no-loss transmission of electricity, and smaller and more powerful electric motors and generators. Superconducting magnets might be used to contain the extremely hot plasma needed to bring about hydrogen fusion, producing almost pollution-free energy in much the same way that the Sun does. Superconducting magnets of between 2 T and 10 T are regularly used in medical applications such as magnetic resonance imaging and in research.

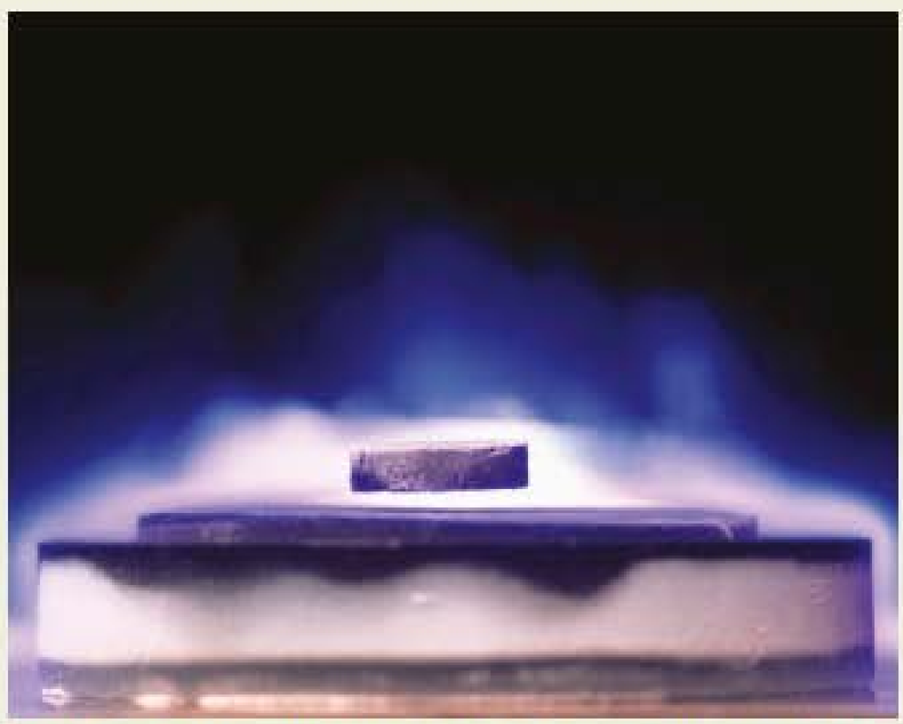


FIGURE 6.3.8 A disc magnet is repelled by a superconductor because the magnet induces a permanent current in the superconductor, which results in an opposing field.

6.3 Review

SUMMARY

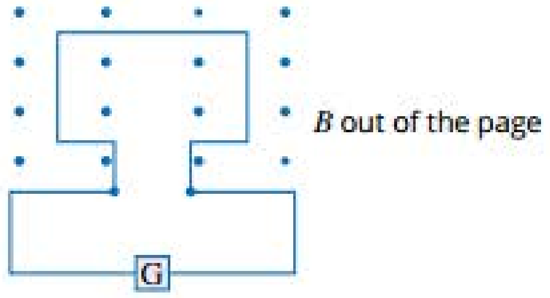
- An induced emf always gives rise to a current whose magnetic field will oppose the original change in flux.
- There are three distinct steps to determine the induced current direction according to Lenz's law:
  1. What is the change that is happening?
  2. What will oppose the change and/or restore the original conditions?
  3. What must be the current direction to match this opposition?
- An induced emf can be created in three ways:
  - by changing the strength of the magnetic field
  - by changing the area of the coil within the magnetic field
  - by changing the orientation of the coil with respect to the direction of the magnetic field.
- An eddy current is a circular electric current induced within a conductor by a changing magnetic field.
- Eddy currents can be useful for creating resistive forces such as in regenerative braking in cars and damping in moving-coil galvanometers.
- Eddy currents can be useful to produce heating effects such as in induction cooktops.

KEY QUESTIONS

1 A conducting loop is located in an external magnetic field whose direction (but not necessarily magnitude) remains constant. A current is induced in the loop. Which of the following best describes the direction of the magnetic field due to the induced current?

A It will always be in the same direction as the external magnetic field.
B It will always be in the opposite direction to the external magnetic field.
C It will be in the same direction as the external magnetic field if the external magnetic field gets weaker, and it will be in the opposite direction to the external magnetic field if the external magnetic field gets stronger.
D The direction can't be determined from the information supplied.

2 A rectangular conducting loop forms the circuit shown below. The plane of the loop is perpendicular to an external magnetic field whose magnitude and direction can be varied. The initial direction of the field is out of the page.



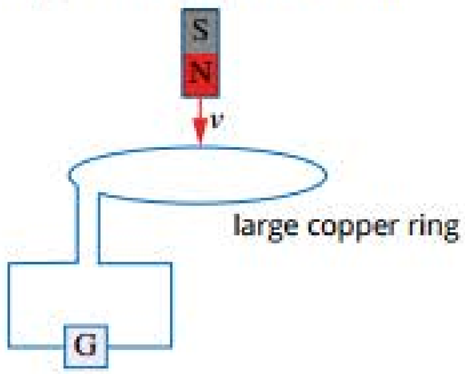
a What will be the direction of the magnetic field due to the induced current when the magnetic field is switched off?

A out of the page
B into the page
C clockwise
D anticlockwise
E left to right
F right to left

b What will be the direction of the magnetic field due to the induced current when the magnetic field is reversed?

A out of the page
B into the page
C clockwise
D anticlockwise
E left to right
F right to left

3 A bar magnet is falling towards the centre of a horizontal copper coil, as shown below.



a State the direction (as seen from above) of the induced current in the coil when the magnet is in the position shown in the diagram.

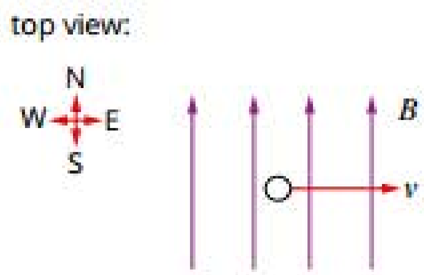
b List four factors that would influence the magnitude of the induced current in the copper ring.

6.3 Review *continued*

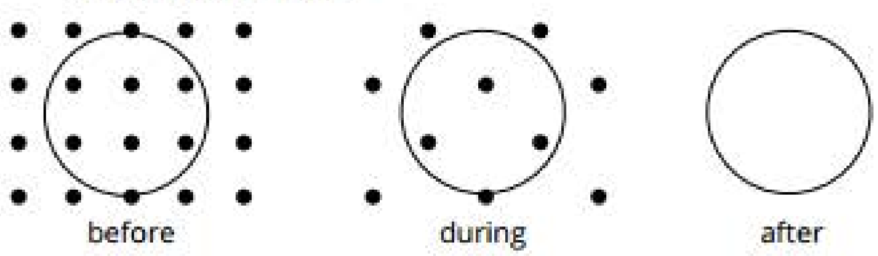
4 A copper wire falls through a horizontal magnetic field with its ends sliding smoothly down two thick vertical rails, the top ends of which are connected by a wire of negligible resistance. Outline why the wire reaches a terminal velocity.

5 A magnetised compass needle is allowed to oscillate first above a sheet of glass and then above a sheet of aluminium. Use Lenz's law to describe what happens in each case.

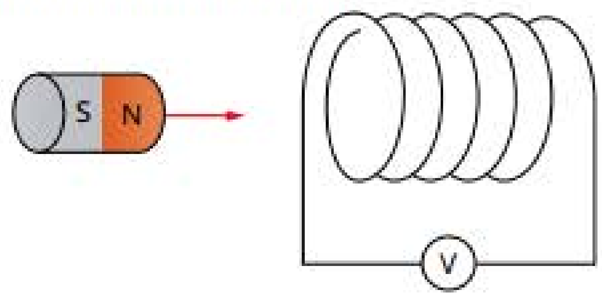
6 A 0.750 m aerial is placed vertically on the roof of a car and is insulated from the car. The driver then drives the car east at 60.0 km h^{-1}, in a region near Perth where the horizontal component of the Earth's magnetic field is 2.50×10^{-5} T. Which end of the aerial is at a positive potential?



7 A circular loop of area 1.00×10^{-2} m^2 is placed at right angles to a uniform magnetic field generated by a solenoid with a strength 25.0 mT, as shown in the diagram below. The current in the solenoid is switched off and the field drops to zero in a time of 0.0500 seconds. In which direction does the induced current flow around the coil during the change in flux?

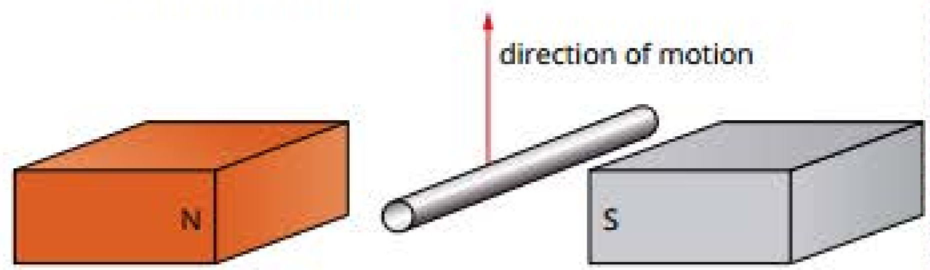


8 The north pole of a magnet is brought from a distance to rest inside the coil, as shown.



Determine the direction of the induced current. Justify your answer.

9 Determine the positive end of the metal bar shown in the diagram below.



10 Outline one useful application of eddy currents.

6.4 Transforming voltage using changing magnetic field

When Faraday first discovered electromagnetic induction, he had effectively invented generators and transformers. A transformer is a device for increasing and decreasing an alternating voltage. Transformers can be found in almost all electrical devices. They are an essential part of any electrical distribution system and are the focus of this section (Figure 6.4.1).

FIGURE 6.4.1 (a) Transformers at an electrical substation. The substation takes electricity from the distribution grid and converts it to lower voltages used by industrial or residential equipment. (b) More common are the smaller distribution transformers found on every suburban street. See if you can locate at least one on your street.

ELECTRIC POWER GENERATORS

We take the supply of electric power to our homes, schools and businesses for granted, yet it was only in 1894 that the Perth Gas Company connected the first commercial supply of electricity in Perth.

The electric **generator** is probably the most important practical application of Faraday's discovery of electromagnetic induction. The principle of electric power generators is the same whether the result is alternating current (AC) or direct current (DC). Relative motion between a coil and a magnetic field induces an emf in the coil. In small generators, the coil is rotated within a magnetic field, but in large power stations, car alternators, and other industrial-level production the coils are stationary and an electromagnet rotates inside them.

This might all sound quite similar to the way electric motors work (see Chapter 5). In fact, it is—a generator is basically just the inverse of a motor.

Induced emf in an alternator or generator

A basic electric generator, or **alternator**, consists of many coils of wire wound on an iron core framework. This is called an armature and it is made to rotate in a magnetic field. The axle is turned by some mechanical means—mechanical energy is being converted to electrical energy—and an emf is induced in the rotating coil.

Consider a single loop of wire in the generator shown in Figure 6.4.2. The loop is rotated clockwise in a uniform magnetic field, B. The amount of flux threading through the loop will vary as it rotates. It is this change in flux that induces the emf.

While sides AD and BC are moving, because they are running parallel to the magnetic field, the force on the electrons within the wire is towards the sides of the wire, not along their lengths. The emf is only generated in sides AB and CD where the wire is 'cutting flux'.

Lenz's law tells us that as the flux in the loop decreases from position (a) to (b) in Figure 6.4.2, the induced current will be in a direction such as to create a magnetic field in the same direction, relative to the loop, as the external field. The right-hand grip rule can then be used to show that the direction of the induced current will be D → C → B → A.

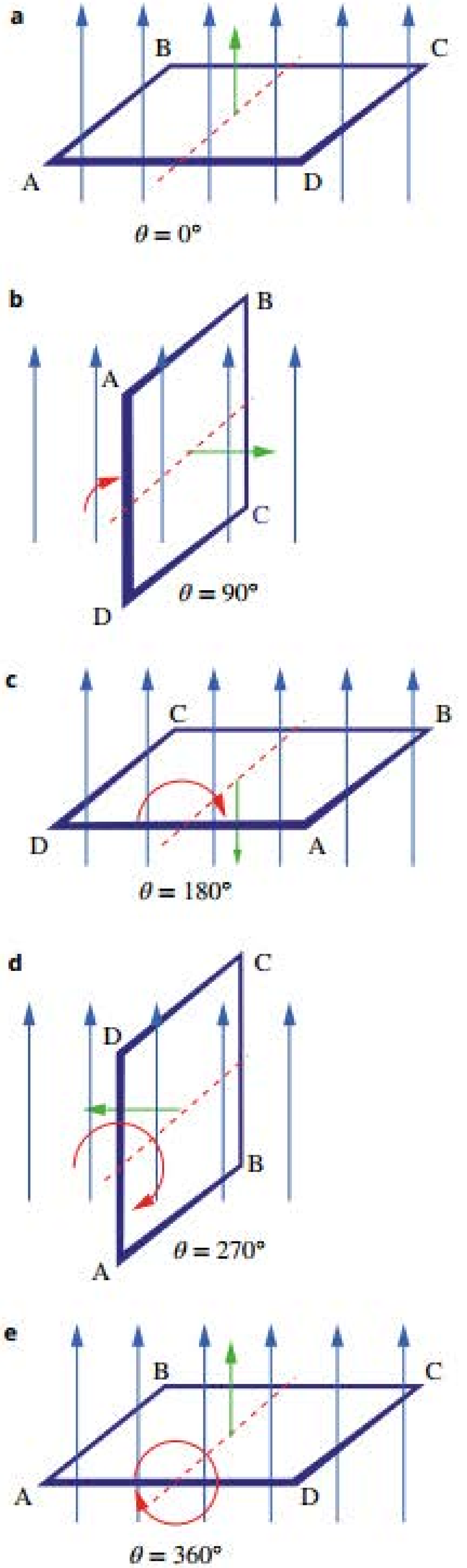


FIGURE 6.4.2 A single loop of a generator rotating in a magnetic field. (a) The plane of the area of the loop is perpendicular to the field B and the amount of flux $\Phi = BA$ is at a maximum. (b) The loop has turned one quarter of a turn and is parallel to the field; $\Phi = 0$. (c) As the loop continues to turn, the flux increases to a maximum but in the opposite sense relative to the loop in (a); $\Phi = -BA$. (d) The flux then decreases to zero again as the loop is parallel to the field before repeating the cycle again from (e) onwards.

The direction of the induced current will reverse every time the plane of the loop reaches a point perpendicular to the field. The magnitude of the induced emf will be determined by the rate at which the loop is rotating. It will be a maximum when the rate of change of flux is a maximum. This is when the loop is moving through a position parallel to the magnetic field where the flux through the loop is zero, i.e. the gradient of the flux versus time graph shown in Figure 6.4.3 is a maximum.

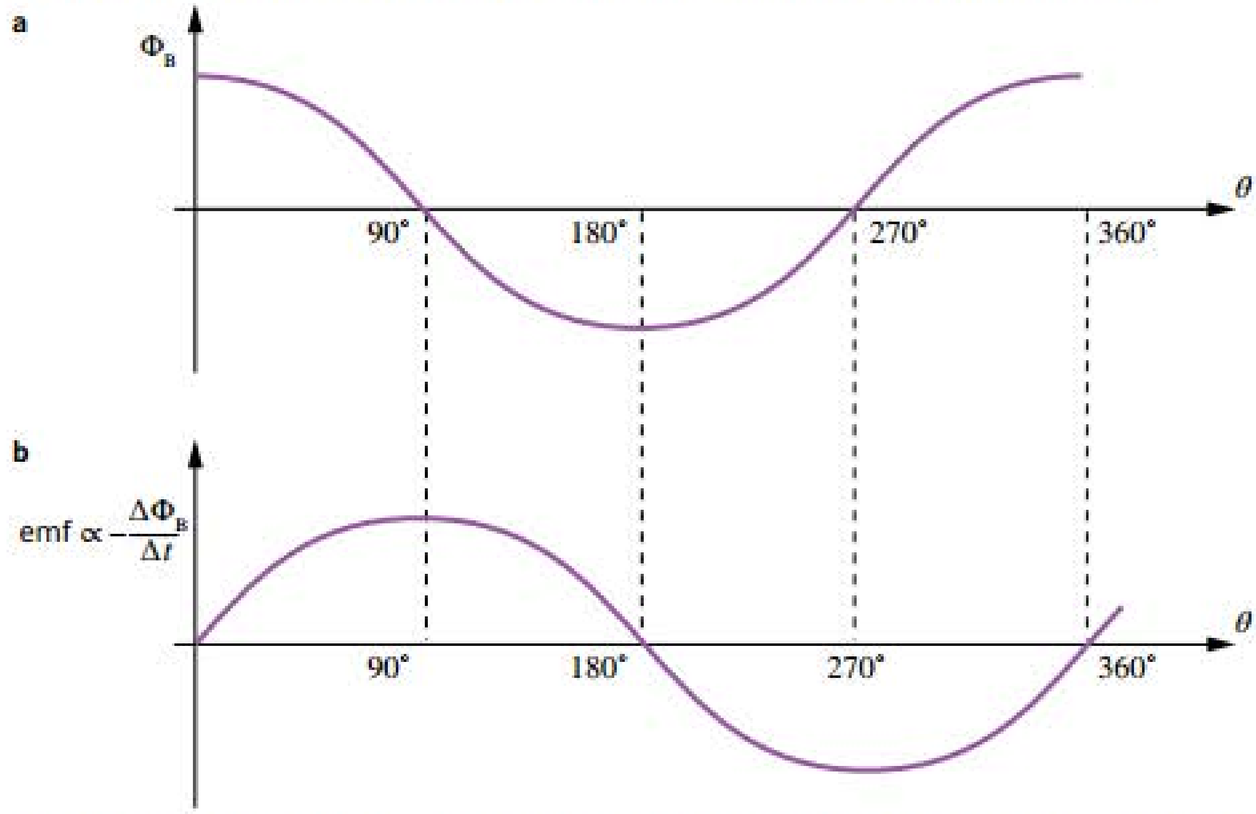


FIGURE 6.4.3 (a) The flux, ϕ, through the loop of Figure 6.4.2 as a function of the angle between the field and the normal to the plane of the area, θ. (b) The rate of change of flux and hence emf through the loop as a function of the angle between the field and the normal to the plane of the loop, θ. The loop is rotating at a constant speed.

An alternative way to think about how the emf changes as the loop rotates is to remember that the emf is actually created as the wires AB and CD cut across the magnetic field lines. Maximum emf occurs when these wires cut the magnetic field lines perpendicularly, when θ is 90° or 270°, and zero emf occurs when the motion of these wires is parallel to the field lines when θ is 0°, 180° or 360°.

The magnitude of the maximum emf on each side can be expressed in terms of the velocity of the loop perpendicular to the magnetic field:

$$\varepsilon = -B\ell v_{\perp}$$

where ℓ is the length in metres of either side of the loop that is cutting flux, and the negative sign is there, once again, to remind you in which direction the maximum emf is produced.

The total maximum emf, adding the emf in both sides of the loop, for a coil of loops will then be:

$$\varepsilon_{max} = -2NB\ell v_{\perp}$$

In more general terms, if the coil of any generator of area A, no matter what shape, consists of loops rotating at a constant rate, then the magnitude of the maximum induced emf becomes:

$$\varepsilon_{max} = -2\pi NBA_{\perp}f$$

where f is the frequency of rotation in hertz (Hz). Throughout Australia, the frequency is maintained very precisely at 50 Hz.

Notice that the maximum or peak value of the induced emf is proportional to the strength of the magnetic field, the area of the coil, the number of turns on the coil and the rate at which the coil is turning. Notice also that this gives you the maximum or peak value of the emf in the cycle, not the average emf over a period of time in the cycle.

Worked example 6.4.1

AC GENERATOR

The armature of a 50 Hz AC generator is rotating within a 0.150 T magnetic field. If the area of the coil is $2.00 \times 10^{-2}\,m^2$, how many loops must the coil contain for the peak output to be 340 V? Give your answer to two significant figures.

Thinking	Working
The is an application of $\varepsilon_{max} = -2\pi NBA_{\perp}f$. Identify the known and unknown quantities in the equation.	$\varepsilon_{max} = 340\,V$ $N = ?$ $B = 0.150\,T$ $A = 2.00 \times 10^{-2}\,m^2$ $f = 50\,Hz$
Rearrange the equation in terms of the unknown, *N*. The negative sign can be left out—it is a reminder that the direction of the induced emf opposes the change that created it.	$\varepsilon_{max} = -2\pi NBA_{\perp}f$ $N = \frac{\varepsilon}{2\pi BA_{\perp}f}$ $N = \frac{340}{2\pi \times 0.150 \times 2.00 \times 10^{-2} \times 50}$ $N = 360.75$ or 360 turns

Worked example: Try yourself 6.4.1

AC GENERATOR

The armature of a 50 Hz AC generator is rotating within a 0.200 T magnetic field. If the coil has 500 turns, what must the area be, in m^2, for the peak output to be 340 V?

AC generators and alternators

The construction of a generator is basically the same as that of a motor. The main components of an AC generator are shown in Figure 6.4.4.

Consider a coil, or armature, with a number of turns, being rotated in a magnetic field, inducing an emf as shown previously in Figure 6.4.2 (page 203). The resultant emf alternates in direction as shown by the graph going above and below the zero emf line in Figure 6.4.3. This type of emf or voltage produces an AC in the coil. How this AC in the coil is harnessed determines if the device is an AC alternator or a DC generator.

As was stated earlier, in many industrial generators the coils remain static and the electromagnet rotates within them. The principle of inducing an emf is the same. The coil itself may take a variety of shapes, sizes and positions.

If the output from the coils is transferred to a circuit via continuous **slip rings**, the alternating current in the coil will be maintained at the output. The slip rings also allow the coil to rotate without tangling the connecting wires. Carbon **brushes** press against the slip rings to allow a constant output to be transferred to a circuit without a fixed point of connection.

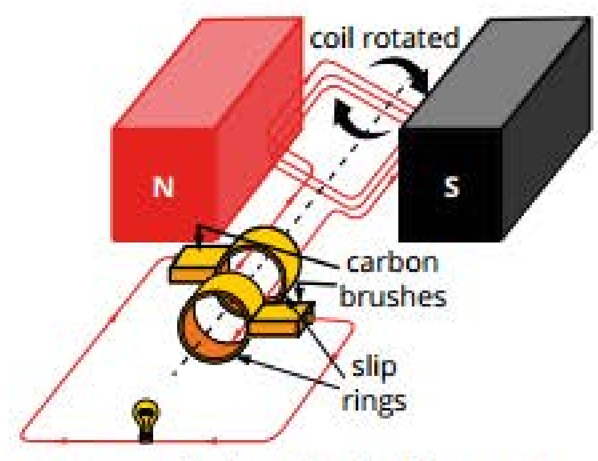


FIGURE 6.4.4 A schematic of an AC generator showing the main features.

PHYSICS IN ACTION

Three-phase generators

Many industrial applications require a more constant maximum voltage than is possible from a single coil. These applications require a three-phase power supply. The coils are arranged such that the emfs vary at the same frequency, but with the peaks and troughs of their waveforms offset to provide three complementary currents with a phase separation of one-third of a cycle, or 120°. The resulting output of all three phases maintains an emf near the maximum voltage more continuously. Standard electrical supplies include three phases, but most home applications only require a single phase to be connected.

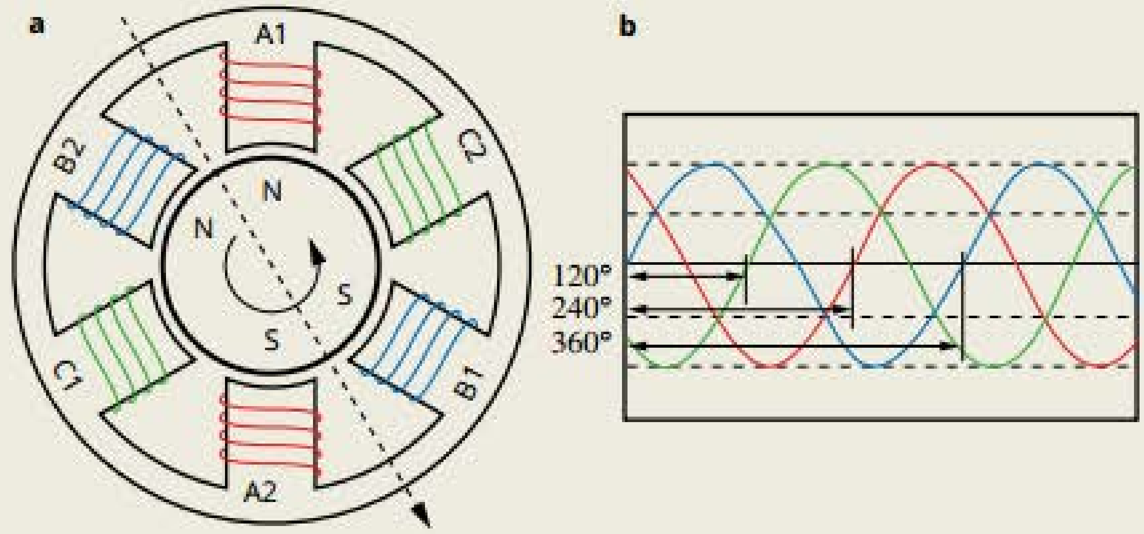


FIGURE 6.4.5 (a) A three-phase power supply has three coils, each producing an output 120° out of phase with the adjoining coil. (b) The resulting output can be combined for a more constant supply voltage.

DC generators

A DC generator is much like an AC generator in basic design. The continuous slip rings are replaced by a **split-ring commutator**. That is, the ring picking up the output from the coils has two breaks (or splits) in it at opposite sides of the ring. The direction of the output is changed by the commutator every half turn so that the output current is always in the same direction (Figure 6.4.6(a)). The output will still vary from zero to a maximum every half cycle. The output can be smoothed by placing a capacitor in parallel with the output. More commonly, the use of multiple armature windings and more splits in the split ring commutator can smooth the output by ensuring that the output is always connected to an armature that is in the position for generating maximum emf (Figure 6.4.6(b)).

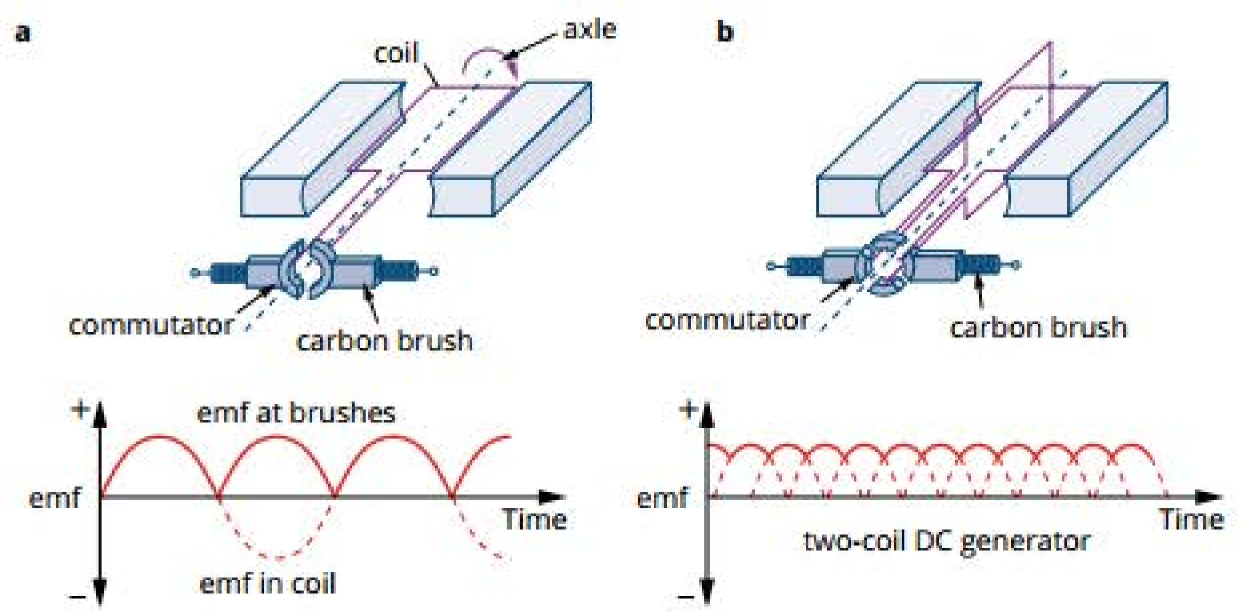


FIGURE 6.4.6 (a) A DC generator has a commutator to reverse the direction of the alternating current every half cycle and so produce a DC output. (b) Multiple armature windings can smooth the output.

In the past, cars used DC generators to power ancillary equipment. More common now is the use of AC generators or alternators, which avoid the problems of wear and sparking across the commutator, inherent in the design of DC generators, by using a moving electromagnet inside a set of stationary coils to generate current.

PHYSICS IN ACTION

Back emf in DC motors

The description of the construction and operation of a generator shows that a DC motor and a generator share a lot in common and can function in either way; so every motor can also be used as a generator. The motors of electric trains, for instance, work as generators when a train is slowing down, converting kinetic energy to electrical energy and putting it back into the electrical supply grid. Regenerative brakes in cars work in a similar way. A DC motor will also generate an emf when running normally. This is termed the '**back emf**'.

The back emf generated in a DC motor is the result of current produced in response to the motion of a conducting coil inside the motor in the presence of an external magnetic field. This back emf, according to Lenz's law, opposes the change in magnetic flux that created it, so this induced emf will be in the opposite direction to the applied emf creating the motion in the first place. The net emf used by the motor is thus always less than the applied voltage:

$$\varepsilon_{net} = V - \varepsilon_{back}$$

As the motor increases speed, the induced back emf will also increase, thus reducing the net emf and therefore reducing the applied current in the motor coils. This has the effect of reducing the force on the motor and so it will accelerate the motor less and less. Eventually the back emf will equal the applied emf and, theoretically, there will be no net emf and no current flowing in the coil. As the current causes the force on the wire there will be no force and consequently no acceleration; the motor is at its maximum speed.

When a load is applied to the motor, such as a when a drill is drilling into a piece of wood, the speed of the motor will generally reduce. This will reduce the back emf and increase the net emf and hence the current in the motor. This increase in current causes a force on the motor coil that acts to do work on the wood. If the load brings the motor to a sudden halt—say, an electric drill bit getting stuck in some wood—the sudden decrease in back emf and increase in net emf will cause the current to increase rapidly such that it may be high enough to burn out the motor and the motor windings. To protect the motor, a resistor is placed in series. It is switched out of the circuit when the current drops below a predetermined level and is switched back into the circuit for protection once the level is exceeded.

Regenerative braking

Every time a traditional car brakes, it is wasting energy. When the car slows down, the kinetic energy that was propelling it forwards is largely lost as heat. That energy, and the original fuel that was used to create it, is essentially wasted. **Regenerative braking** overcomes this problem by recapturing much of the car's kinetic energy and converting it to electrical energy to recharge the car's battery, as shown in Figure 6.4.7.

Regenerative braking is nothing new. Trams have been utilising regenerative brakes for many years, feeding excess energy back into the electrical supply grid. With more focus on energy efficiency and hybrid cars, regenerative brakes are now becoming more common in cars.

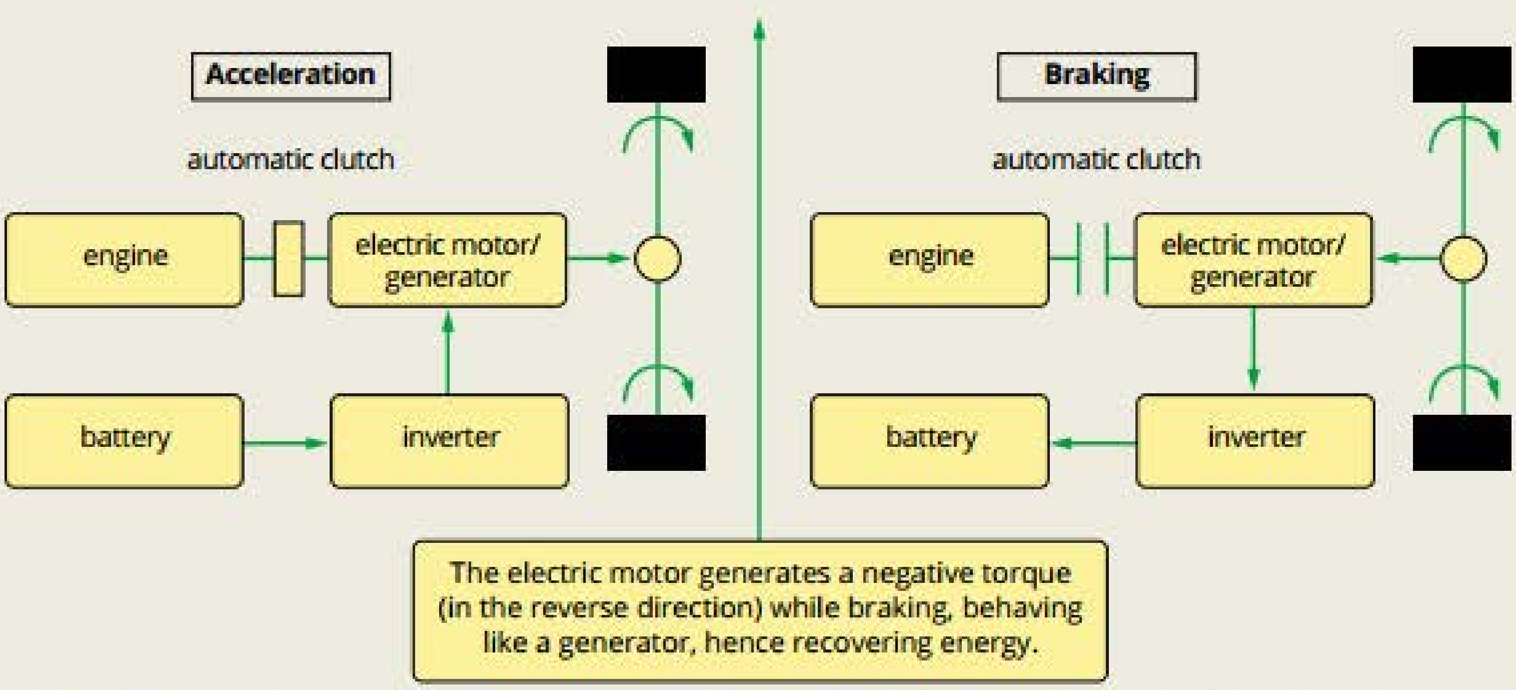


FIGURE 6.4.7 A regenerative braking system, showing the flow of electrical energy during acceleration and braking.

PHYSICS IN ACTION *(continued)*

Traditional brake systems rely on a brake pad producing friction against a metal rotor or disc to slow or stop the vehicle, with additional friction to bring the vehicle to a full stop coming from the action between tyres and road. Friction turns the car's kinetic energy into heat. With induction-based regenerative brakes, it's actually the same system that propels the car that does most of the braking, with the kinetic energy being recovered to be used again.

Regenerative braking is largely used in electric and hybrid cars that make use of electric motors to propel the car and have batteries to store charge. The braking system runs the car's electric motor in reverse, converting mechanical energy into electrical energy and then into the car's batteries for storage.

During acceleration, the system uses electricity from the battery, converts it to the voltage required via an inverter transformer and sends it to the electric motor to power the car. During braking, the process is reversed. Mechanical energy from the moving car is converted by the motor into electrical energy and fed back through the inverter for storage in the car's batteries.

Specialised electronic circuits determine when the motor should reverse and direct the electricity produced back to the batteries or, in some cases, to a series of capacitors for shorter term storage. The process also provides very effective braking via the eddy currents induced to oppose the change in flux.

ALTERNATING VOLTAGE AND CURRENT

An AC generator produces an alternating current that varies **sinusoidally** over time with the change in magnetic flux. The maximum emf is only achieved for particular points in time. In Australia, mains power oscillates at 50 Hz and reaches a peak voltage of ±340 V each cycle or a peak-to-peak voltage of 680 V (Figure 6.4.8).

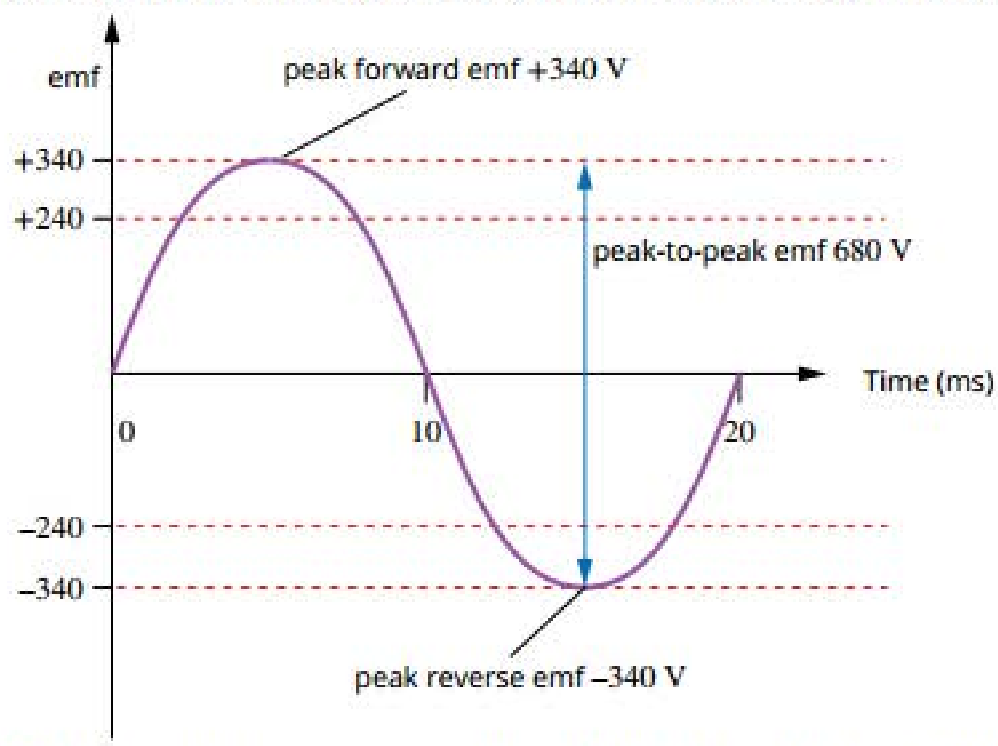


FIGURE 6.4.8 The voltage in Australian power points oscillates between +340 V and −340 V, 50 times each second. The value of a DC supply that would supply the same average power is 240 V.

It is often more useful to know the average power produced in a circuit so that the equations of DC electricity can be used with AC electricity. The average power can be obtained by using a value for the voltage and current equal to the peak values divided by $\sqrt{2}$. This is referred to as the **root mean square** or rms value.

In effect, the rms values are the values of a DC supply that would be needed to provide the same average power as the AC supply. It is the rms value of the voltage $\left(\frac{340}{\sqrt{2}} = 240\right)$ that is normally quoted. This is the effective average value of the voltage and is the value that should be used to find the actual power supplied each cycle by an AC supply.

$$V_{rms} = \frac{V_{peak}}{\sqrt{2}}$$

$$I_{rms} = \frac{I_{peak}}{\sqrt{2}}$$

$$P_{rms} = V_{rms} \times I_{rms} = \frac{1}{2}V_{peak}I_{peak}, \text{ and}$$

$$P_{peak} = \sqrt{2}V_{rms} \times \sqrt{2}I_{rms} = 2V_{rms}I_{rms}$$

EXTENSION

Deriving the root mean square formulae

In an AC circuit, the power produced in a resistor is equal to $\frac{V^2}{R}\sin^2\theta$.

The average power will be given by:

$$\frac{1}{2}\frac{V_{peak}^2}{R}$$

If this same power was to be supplied by a steady (DC) source, the voltage V_{av} of this source would have to be such that:

$$\frac{V_{av}^2}{R} = \frac{1}{2}\frac{V_{peak}^2}{R}$$

Simplifying:

$$V_{av}^2 = \frac{V_{peak}^2}{2}$$

$$V_{av} = \frac{V_{peak}}{\sqrt{2}}$$

This average voltage is known as the root mean square voltage or V_{rms}. It is the value of a steady voltage that would produce the same power as an alternating voltage with a peak value equal to $\sqrt{2}$ times as much.

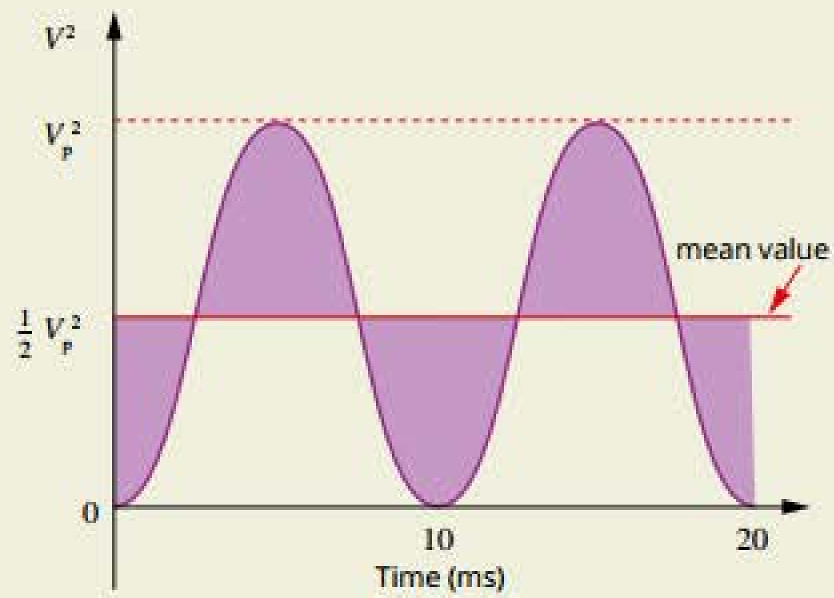


FIGURE 6.4.9 The power transmitted is proportional to the area under a V^2 graph. Clearly, the power transmitted by an AC circuit (with V_{peak}) is the same as that in a DC circuit with a voltage equal to the square root of $\frac{1}{2}V_{peak}^2$, that is $\frac{V_{peak}}{\sqrt{2}}$.

Worked example 6.4.2

PEAK AND RMS AC VALUES

A 60.0 W light globe is connected to a 240 V AC circuit. Calculate the peak power use of the light globe.

Thinking	Working
Note that the values given in the question represent rms values. Power is $P = VI$ so both V and I must be known to calculate the power use. The voltage V is given, and the current I can be calculated from the rms power supplied.	$P_{rms} = V_{rms}I_{rms}$ $I_{rms} = \frac{P_{rms}}{V_{rms}}$ $= \frac{60.0}{240} = 0.250\text{ A}$
Substitute in known quantities and solve for peak power.	$P_{peak} = \sqrt{2}V_{rms} \times \sqrt{2}I_{rms} = 2V_{rms}I_{rms}$ $= 2 \times V_{rms} \times I_{rms}$ $= 2 \times 240 \times 0.250$ $= 120\text{ W}$

Worked example: Try yourself 6.4.2

PEAK AND RMS AC VALUES

A 1000 W kettle is connected to a 240 V AC power outlet. Calculate the peak power use of the kettle.

TRANSFORMERS

A transformer is a device for increasing and decreasing an AC voltage. Transformers can be found in most electrical devices and are essential parts of any electrical distribution system.

The workings of a transformer

A **transformer** works on the principle that a changing magnetic flux induces an emf. No matter what the size or application, a transformer will consist of two coils known as the *primary* and *secondary* coils. The changing flux originates with the alternating current supplied to the primary coil. The changing magnetic flux is directed to the secondary coil where the changing flux will induce an emf in that coil (Figure 6.4.10).

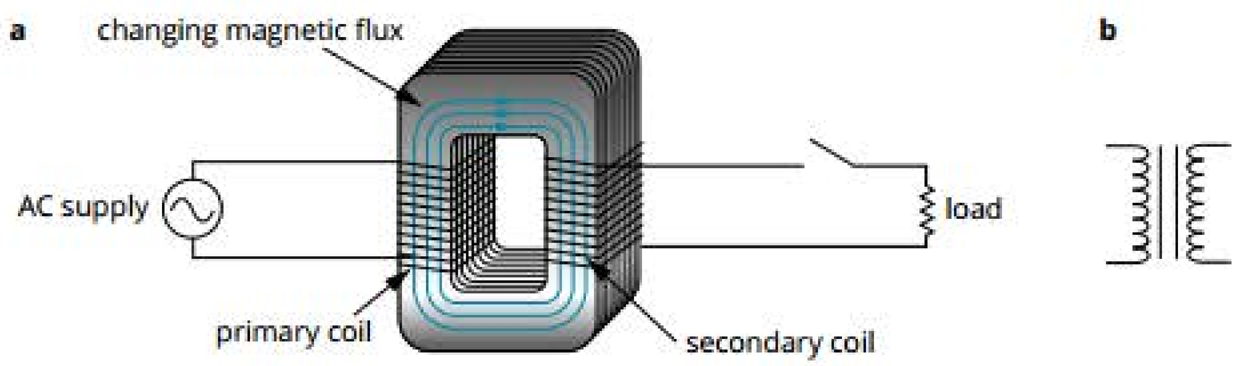


FIGURE 6.4.10 (a) In an ideal transformer, the iron core ensures that all the flux generated in the primary coil also passes through the secondary coil. (b) The symbol used in circuit diagrams for an iron-core transformer.

The two coils can be interwoven using insulated wire or they can be linked by a soft iron core, laminated to minimise eddy current losses. Transformers are designed so that nearly all of the magnetic flux produced by the primary coil will pass through the secondary coil. In an **ideal transformer** the assumption is that this will be 100% efficient and energy losses can be ignored. In a real transformer, this assumption remains a good approximation. Transformers are one of the most efficient devices around, with practical efficiencies often being better than 99%.

PHYSICSFILE

Laminations

Eddy currents that are set up in the iron core of transformers can generate a considerable amount of heat. Energy that has been lost from the electrical circuit and the transformer as heat may become a fire hazard. To reduce eddy current losses, the core is made of laminations, which are thin plates of iron electrically insulated from each other and placed so that the insulation between the laminations interrupts the eddy currents.

AC VERSUS DC

The power distribution system works on alternating current. That may seem odd when many devices run on direct current, but one of the primary reasons is the ease with which alternating current can be transformed from one voltage to another.

A transformer works on the basis of a changing current in the primary coil inducing a changing magnetic flux. This in turn induces an emf and current in the secondary coil circuit. For this to work, the original current must be constantly changing, as it does in an AC supply.

A DC voltage has a constant, unchanging current. With no change in the size of the current, no changing magnetic flux will be created by the primary coil and, hence, no emf or current is induced in the secondary coil circuit. Transformers do not work with the constant current of a DC electrical supply. There will be a very brief induced current when a DC supply is turned on, and a change occurs from zero current to the supply level. There is a similar spike if the DC supply is switched off, but while the DC supply is constant there is no change in magnetic flux to induce an emf or current in the secondary coil circuit.

THE TRANSFORMER EQUATION

When an AC emf is connected to the primary coil of a transformer, the changing magnetic field will induce an AC emf in the secondary coil of the same frequency as the primary emf. The value of the emf in the secondary coil will be different and depends upon the number of turns in each coil.

From Faraday's law, the average emf in the primary coil, V_p, will affect the rate at which the magnetic flux changes:

$$V_p = N_p \frac{\Delta\Phi}{t}$$

or

$$\frac{\Delta\Phi}{t} = \frac{V_p}{N_p}$$

where N_p is the number of turns in the primary coil.

The induced emf in the secondary coil, V_s, will be

$$V_s = N_s \frac{\Delta\Phi}{t}$$

and

$$\frac{\Delta\Phi}{t} = \frac{V_s}{N_s}$$

where N_s is the number of turns in the secondary coil.

Assuming that there is little or no loss of flux between the primary and secondary coil, then the flux in each will be the same and

$$\frac{V_p}{N_p} = \frac{V_s}{N_s}$$

or

$$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$

The transformer equation relating emf and number of turns in each coil is:

$$\frac{V_p}{N_p} = \frac{V_s}{N_s} \text{ or } \frac{V_p}{V_s} = \frac{N_p}{N_s}$$

where the subscript 'p' refers to the primary or first coil, and the subscript 's' refers to the secondary coil.

The transformer equation explains how the secondary (output) emf is related to the primary input emf. Either the rms emf for both or the peak emf for both can be used.

A **step-up transformer** increases the secondary emf compared with the primary emf. The secondary emf is greater than the primary emf and the number of turns in the secondary coil is greater than the number of turns in the primary coil, i.e. if $N_s > N_p$ then $V_s > V_p$.

A **step-down transformer** decreases the secondary emf compared with the primary emf. The secondary emf is less than the primary emf and the number of turns in the secondary coil is less than the number of turns in the primary coil, i.e. if $N_s < N_p$ then $V_s < V_p$.

Worked example 6.4.3

TRANSFORMER EQUATION—EMF

A transformer is built into a portable radio to reduce the 240 V supply to the required 12.0 V for the radio. If the number of turns in the secondary coil is 100, what is the number of turns required in the primary coil?

Thinking	Working
State the relevant quantities given in the question. Choose a form of the transformer equation with the unknown quantity in the top left position.	$V_s = 12.0\text{ V}$ $V_p = 240\text{ V}$ $N_s = 100$ turns $N_p = ?$ $\frac{V_p}{V_s} = \frac{N_p}{N_s}$
Substitute the quantities into the equation, rearrange and solve for N_p.	$\frac{N_p}{100} = \frac{240}{12.0}$ $N_p = \frac{100 \times 240}{12.0}$ $= 2000$ turns

Worked example: Try yourself 6.4.3

TRANSFORMER EQUATION—EMF

A transformer is built into a phone charger to reduce the 240 V supply voltage to the required 6.00 V for the charger. If the number of turns in the secondary coil is 100, what is the number of turns required in the primary coil?

POWER OUTPUT

Although a transformer very effectively increases or decreases an AC emf, energy conservation means that the output power cannot be any greater than the input power. Since a well-designed transformer with a laminated core can be more than 99% efficient, the power input can be considered equal to the power output, making it an 'ideal' transformer.

As power supplied is $P = VI$, then:

$$V_pI_p = V_sI_s$$

The transformer equation can then be written in terms of current, I.

Note that the number-of-turns ratio for currents is the *inverse* of that for the transformer equation written in terms of emf.

The transformer equation, relating current and the number of turns in each coil:

$$\frac{I_p}{I_s} = \frac{N_s}{N_p} \text{ or } \frac{I_s}{I_p} = \frac{N_p}{N_s} \text{ or } \frac{I_p}{N_s} = \frac{I_s}{N_p}$$

PHYSICSFILE

Standby power

Because very little current will flow in the primary coil of a good transformer to which there is no load connected, it will use little power when not in use. However, this 'standby power' can add up to around 10% of power use. This is why devices such as TVs and computers should be switched completely off when not in use. Over the whole community, standby power amounts to megawatts of wasted power and unnecessary greenhouse emissions. Special switches, such as the 'Ecoswitch' shown in Figure 6.4.11, have been developed that can be connected between the power outlet and the device to make it easier to remember to turn devices completely off when not in use.

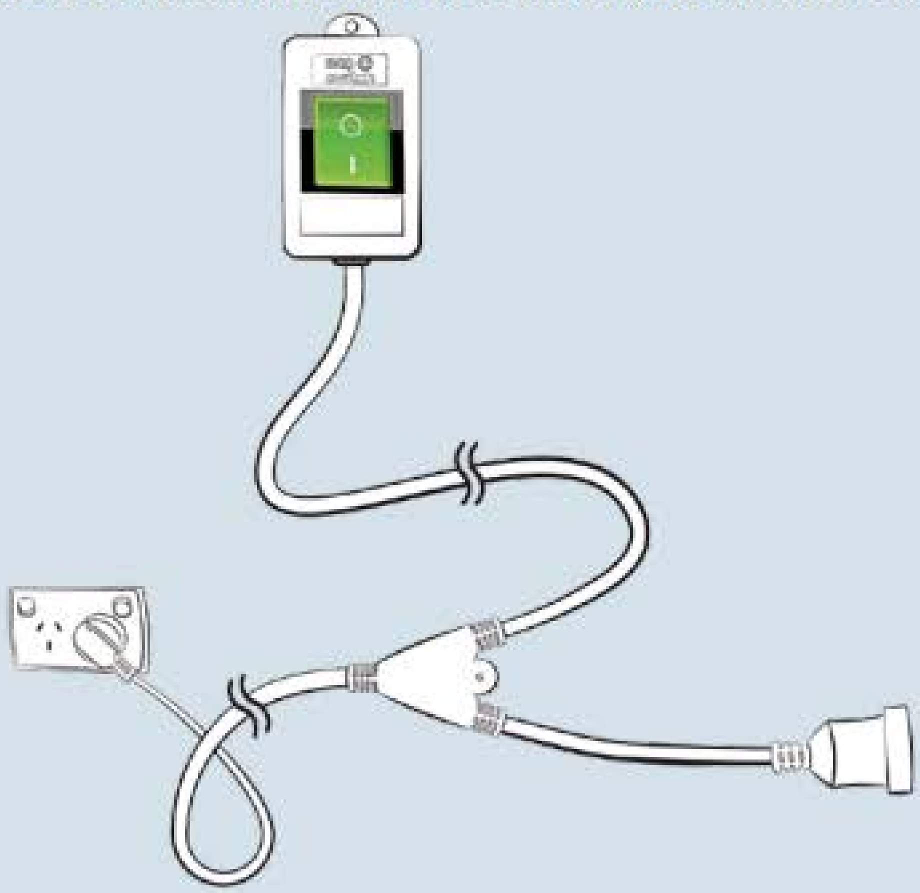


FIGURE 6.4.11 Standby switches such as the 'Ecoswitch' make it easier and more convenient to turn devices completely off when not in use, saving up to 10% on power bills.

PHYSICSFILE

Overload

A transformer will be overloaded if too much current is drawn and the resistive power loss in the wires becomes too great. There will be a point at which the transformer starts to overheat rapidly. For this reason, it is important not to exceed the rated capacity of a transformer.

Worked example 6.4.4

TRANSFORMER EQUATION—CURRENT

A radio with 2000 turns in the primary coil and 100 turns in its secondary coil draws a current of 4.00A. What is the current in the primary coil?

Thinking	Working
State the relevant quantities given in the question. Choose a form of the transformer equation with the unknown quantity in the top left position.	N_s = 100 turns N_p = 2000 turns I_s = 4.00 A I_p = ? $\frac{I_p}{I_s} = \frac{N_s}{N_p}$
Substitute the quantities into the equation, rearrange and solve for I_p.	$\frac{I_p}{4.00} = \frac{100}{2000}$ $I_p = \frac{4.00 \times 100}{2000}$ $= 0.200\,A$

Worked example: Try yourself 6.4.4

TRANSFORMER EQUATION—CURRENT

A phone charger with 4000 turns in the primary coil and 100 turns in its secondary coil draws a current of 0.500 A. What is the current in the primary coil?

Worked example 6.4.5

TRANSFORMERS—POWER

The power drawn from the secondary coil of the transformer by a portable radio is 48.0 W. What power is drawn from the mains supply if the transformer is an ideal transformer?	
Thinking	**Working**
The energy efficiency of an ideal transformer can be assumed to be 100%. The power in the secondary coil will be the same as that in the primary coil.	The power drawn from the mains supply is the power in the primary coil, which will be the same as the power in the secondary coil: $P = 48.0\,W$

Worked example: Try yourself 6.4.5

TRANSFORMERS—POWER

The power drawn from the secondary coil of a transformer by a phone charger is 3.00 W. What power is drawn from the mains supply if the transformer is an ideal transformer?

POWER FOR CITIES: LARGE-SCALE AC SUPPLY

In your school experiments using electric circuits, it is likely that you have ignored the resistance of the connecting wires because the wires (generally made from copper) are good conductors, and so the resistance is very small over short distances. However, over large distances, even relatively good electrical conductors such as copper have a significant resistance.

Modern cities use huge amounts of electrical energy, most of which is supplied from power stations built at a considerable distance from the metropolitan areas. The efficient transmission of the electrical energy with the least amount of power loss over that distance is therefore a very important consideration for electrical engineers, particularly given the vast distances between population centres in Australia.

The power lost in an electric circuit is given by $P_{loss} = \Delta VI$, where ΔV is the potential difference or voltage drop across the load. Recalling Ohm's law, $\Delta V = IR$, and substituting, the power can be expressed in terms of either current and load resistance or potential difference and load resistance:

$$P_{loss} = \Delta VI = I^2R = \frac{\Delta V^2}{R}$$

By considering the form of the equation including the current carried by the circuit and its electrical resistance ($P_{loss} = I^2R$), it is clear that transmitting large amounts of power using a large current will create very large power losses. If the current in the power lines can be reduced, it will significantly reduce the power loss. As the power loss is proportional to the square of the current, if the current is reduced by a factor of 3, for example, the power loss will be reduced by a factor of 3^2 or 9.

The challenge, then, is to transmit the large amounts of power being produced at power stations using a very low current. Transformers are the most common solution to this problem. Using a step-up transformer near the power station, the emf or voltage is increased by a certain factor and, importantly, the current is decreased by the same factor. Due to the $P_{loss} = I^2R$ equation, the power lost during transmission is reduced by the square of that factor.

At this point you might be confused by the alternative equation for power loss: $P_{loss} = \frac{\Delta V^2}{R}$. A simple misunderstanding could make you think that increasing the voltage through the use of a step-up transformer would actually lead to greater power loss, if you use this equation to calculate power loss. However, ΔV represents the potential difference or voltage drop in a circuit. You must be careful not to confuse the potential (V) or voltage at a point in the wires with the potential difference or voltage *drop across* the wires. So, even though the potential or voltage at the beginning is increased through the use of a step-up transformer, the potential difference or voltage drop across the wires would be reduced. Since the current is reduced, the potential difference, $\Delta V = IR$, and thus the power loss would also be reduced.

AC power from the generator is readily stepped up by a transformer to somewhere between 240 kV and 500 kV prior to transmission. Once the electrical lines reach the city, the potential or voltage is stepped down in stages at electrical substations for distribution. The power lines in streets will have a potential of around 2400 V, before being stepped down via small distribution transformers to a potential of 240 V for home use.

Worked example 6.4.6

TRANSMISSION-LINE POWER LOSS

300 MW is to be transmitted from the Collie power station to Perth along a transmission line with a total resistance of 1.00 Ω. What would be the total transmission power loss if the initial voltage at Collie was 250 kV?

Thinking	Working
Convert the values to SI units.	$P = 300\text{ MW} = 300 \times 10^6\text{ W}$ $V = 250\text{ kV} = 250 \times 10^3\text{ V}$
Determine the current in the line based on the required voltage.	$P = VI$ $\therefore I = \frac{P}{V}$ $I = \frac{300 \times 10^6}{250 \times 10^3}$ $= 1200\text{ A}$
Determine the corresponding power loss.	$P_{loss} = I^2R$ $= 1200^2 \times 1.00$ $= 1.44 \times 10^6\text{ W}$ or 1.44 MW

Worked example: Try yourself 6.4.6

TRANSMISSION-LINE POWER LOSS

300 MW is to be transmitted from the Collie power station to Perth along a transmission line with a total resistance of 1.00 Ω. What would be the total transmission power loss if the initial voltage at Collie was now to be 500 kV?

Worked example 6.4.7

VOLTAGE DROP ALONG A TRANSMISSION LINE

Power is to be transmitted from the Garden Island wave-power station to Fremantle along a transmission line with a total resistance of 1.00 Ω. The current is 1200 A. What initial voltage would be needed at the Garden Island end of the transmission line to achieve a supply voltage of 250 kV?

Thinking	Working
Determine the voltage drop along the transmission line.	$\Delta V = IR$ $= 1200 \times 1.00$ $= 1200\,\text{V}$
Determine the initial supply voltage.	$V_{\text{initial}} = V_{\text{supplied}} + \Delta V$ $= 250 \times 10^3 + 1200$ $= 251\,200\,\text{V}$ or $251.2\,\text{kV}$

Worked example: Try yourself 6.4.7

VOLTAGE DROP ALONG A TRANSMISSION LINE

Power is to be transmitted from the Garden Island wave-power station to Fremantle along a transmission line with a total resistance of 1.00 Ω. The current is 600 A. What initial voltage would be needed at the Garden Island end of the transmission line to achieve a supply voltage of 500 kV?

PHYSICS IN ACTION

Large-scale electrical distribution systems

Large-scale energy transmission is done through an interconnected grid between the power stations and the population centres where the bulk of the electrical energy is used. A wide-area synchronous grid, also known as an interconnection, directly connects a number of generators, delivering AC power with the same relative phase, to a large number of consumers.

In Western Australia, the South West Interconnected System (SWIS) extends from Kalbarri in the north to Kalgoorlie in the east and Albany in the south, and includes coal and gas-fired power stations along with others based on waves, wind, solar and hydro.

No matter the source, the path the electrical power takes to the final consumer is very similar. Step-up transformers in a large substation near the power station will raise the voltage from that initially generated to 240 000 V (240 kV) or more. The electrical power will then be carried via high-voltage transmission lines to a number of substations near key centres of demand. Substations with step-down transformers then reduce the voltage to safer levels for distribution underground or via the standard 'electricity pole' you would be familiar with around city and country areas. Each group of 10–15 houses will be supplied by a smaller distribution transformer, mounted on the poles, which reduces the voltage down to the 240 V AC rms voltage that home and business installations are designed to run on (Figure 6.4.12).

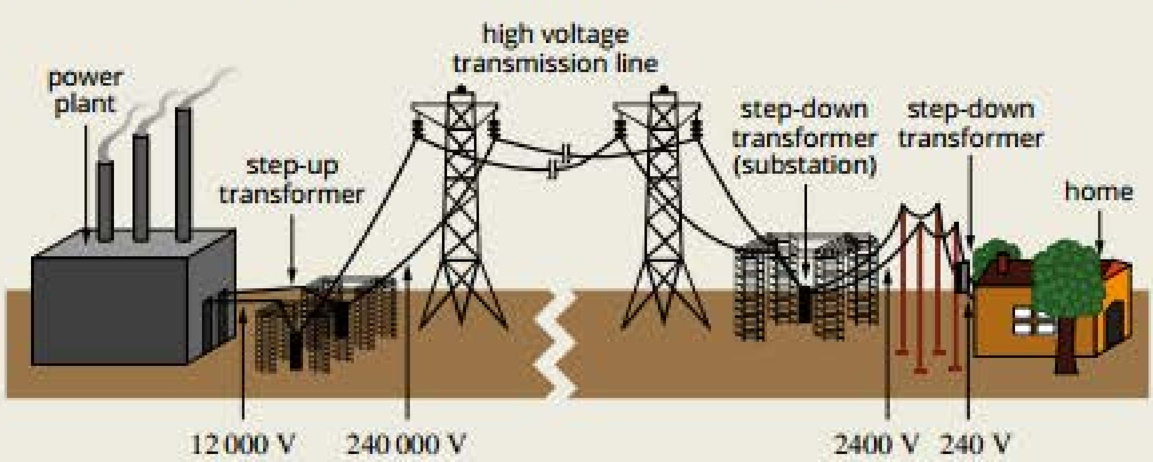


FIGURE 6.4.12 Transmitting electric power from generator to home uses AC power, so transformers can be used to minimise power losses through the system.

The use of AC as the standard for distribution allows highly efficient and relatively cheap transformers to convert the initial voltages created at the power station to much higher levels. The same power transmitted at a higher voltage requires less current and therefore there is less power loss. If it were not for this, the resistance of the transmission wires would need to be significantly reduced, which would require more copper in order to increase their cross-sectional area. This is both expensive and heavy. Less metal makes cables lighter and thinner, and the supporting towers themselves can be comparatively shorter, cheaper and lighter to build.

DC TRANSMISSION

Although AC power is now universal in large-scale power distributions, there is a limit to how high the voltage of an AC system can go and still be efficient. Above approximately 100 kV, corona loss (due to the high voltage ionising air molecules) begins to occur, and above 500 kV it becomes no longer feasible to transmit electric power, due to these effects.

To transmit the same power as a DC system, an AC system would need to operate with a higher peak voltage. During the portion of the cycle when the AC is lower than peak voltages, efficiency is compromised, because the higher the voltage the better. Until recently, the expense of alternative methods to raise and lower the voltage at either end of the transmission line more than negated this negative aspect of AC systems.

High DC voltage levels can now be reached more easily with new technology by using small, high-frequency switching converters. Projects such as the Three Gorges Dam in China (Figure 6.4.13) and undersea transmission lines are now planning to use DC transmission. There are some other benefits, with many of the AC/DC transformers and three-phase industrial power currently in use becoming unnecessary. However, there is a whole range of other devices to be considered that would need to be allowed for, and there are major issues with safety. For example, safety switches won't work with DC power.

FIGURE 6.4.13 At the Three Gorges Dam in China, engineers are considering the use of DC transmission at higher voltages than is possible with AC to further reduce transmission losses.

PHYSICSFILE

The War of Currents

AC and DC power supplies have been in competition for nearly as long as humans have been generating electricity. The heated debates about the benefits and disadvantages of each type of current prompted what has been called the 'War of Currents' in the late 1800s. During this time Thomas Edison, an American inventor and businessmen, had created the Edison Electric Light Company that he hoped would supply electricity to large parts of the USA with his DC generators. Meanwhile, Nikola Tesla, a Serbian–American physicist, had invented the AC induction motor and, with financial support from George Westinghouse, hoped AC would become the dominant power supply. Ultimately, the ease with which AC could be stepped up using transformers for long-distance transmission with minimal power loss (as discussed in detail throughout this chapter) proved to be the prevailing benefit that led to AC winning the 'war'. However, in his attempt to win the competition, Edison attempted to portray the high-voltage AC power as terrifyingly dangerous by using it to electrocute elephants and by inventing the AC-powered electric chair for the US government to execute prisoners on death row.

In some ways, the competition between Edison and Tesla continues. The Edison Electric Company merged in 1892 to become the General Electric Company, which exists to this day as one of the largest and most profitable companies in the world, while Westinghouse is still in business as a large home-appliance brand.

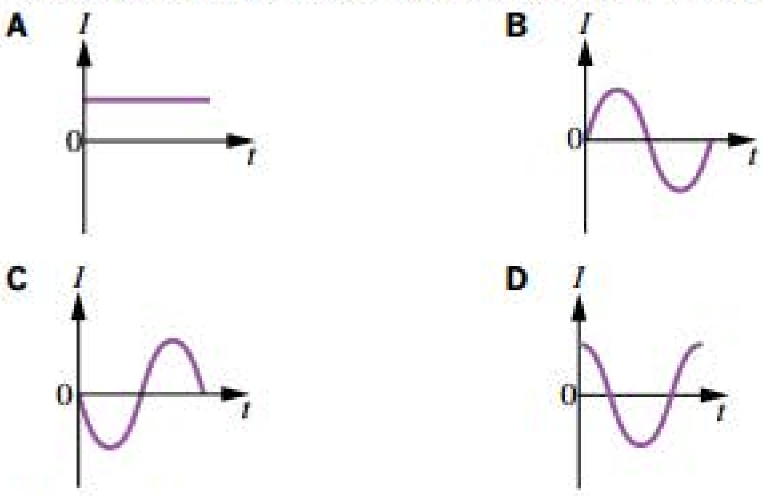
6.4 Review

SUMMARY

- The principle of electric power generators is the same whether the result is alternating current or direct current. Relative motion between a coil and a magnetic field induces an emf in the coil.
- The construction of a generator, or alternator, is very similar to that of an electric motor.
- A coil rotated in a magnetic field will produce an alternating induced current in the coil. How that current is harnessed will determine if the device is an AC alternator or a DC generator.
- An AC alternator has slip rings that transfer the alternating nature of the current in the coil to the output. A DC generator has a split ring commutator to reverse the current direction every half turn so that the output current is always in the same direction.
- Relative motion between a coil and a magnetic field induces an emf in a coil.
- The magnitude of the emf in a generator can be expressed in terms of the velocity of the loop perpendicular to the magnetic field, B:
  $\varepsilon = -\ell v B_{\perp}$
- The total emf, adding the emf in both sides AB and CD, for a coil of N loops is:
  $\varepsilon = -2N\ell vB$ (where $v \perp B$)
- In general terms, if the coil of any generator of area A m^2, no matter what shape, consisting of N loops is rotating at a constant rate then the magnitude of the maximum induced emf is:
  $\varepsilon_{max} = -2\pi NBA_{\perp}f$
- The alternating current produced by power stations and supplied to cities varies sinusoidally at a frequency of 50 Hz. The peak value of the voltage of domestic power (V_{peak}) is ±340 V, and the peak-to-peak voltage ($V_{peak-peak}$) is 680 V.
- The root mean square voltage is the value of an equivalent steady voltage (DC) supply that would provide the same power.
  $V_{rms} = \frac{V_{max}}{\sqrt{2}}$
- The rms value of domestic mains voltage in Australia is 240 V.
- The average power in a resistive AC circuit is:
  $P_{rms} = V_{rms}I_{rms}$
  $= \frac{1}{2} \times V_{peak} \times I_{peak}$
- A transformer works on the principle that a changing magnetic flux induces an emf. No matter what the size or application, it will consist of two coils known as the primary and secondary coils.
- Ideal transformers are 100% efficient; real transformers are often over 99% efficient, and for this reason power losses within the transformer can be ignored in calculations.
- The transformer equation can be written in different versions but is based on:
  $\frac{V_p}{V_s} = \frac{N_p}{N_s}$
- A *step-up* transformer *increases* the secondary voltage compared with the primary voltage.
- A *step-down* transformer *decreases* the secondary voltage compared with the primary voltage.
- The transformer equation can also be written in terms of current:
  $\frac{I_p}{I_s} = \frac{N_s}{N_p}$
- Transformers will not work with DC voltage since it has a constant, unchanging current that creates no change in magnetic flux.
- The power supplied in an electrical circuit is given by $P = VI$.
- The power lost in an electrical circuit is given by
  $P_{loss} = \frac{\Delta V^2}{R} = I^2R$.
- The AC electrical supply from a generator is readily stepped up or down by transformers, hence AC is the preferred form of electrical energy in large-scale transmission systems.

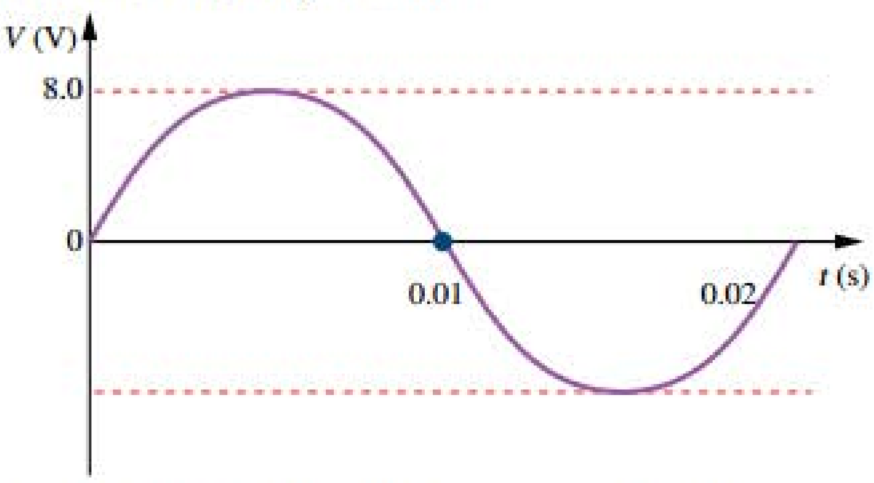
KEY QUESTIONS

1 Assuming that an anticlockwise rotation of a coil starting from $\theta = 0°$ perpendicular to a constant magnetic field initially produces a positive current, which of the graphs below best illustrates the variation of the induced current as a function of time for one full revolution of the coil?



The following information relates to questions 2 and 3.

A simple generator consists of a coil of $N = 1000$ turns, each of radius 10.0 cm, mounted on an axis in a uniform magnetic field of strength B. The following graph shows the voltage output as a function of time when the coil is rotated at a frequency of 50.0 Hz.

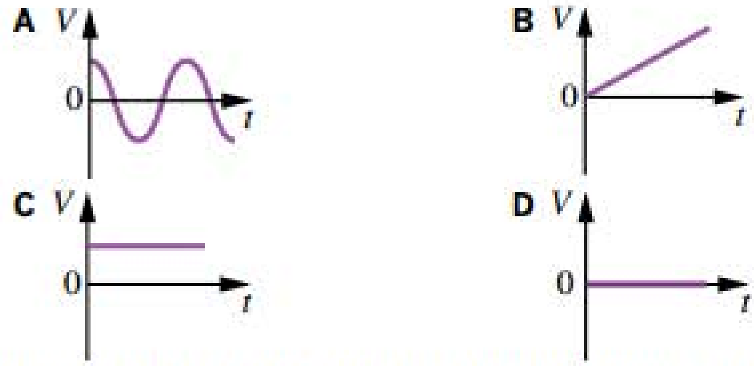


2 Determine the values of V_{peak}, $V_{peak-peak}$ and V_{rms}.

3 The generator is modified so that the magnetic field strength is doubled and the frequency of rotation is increased to 100 Hz. The radius of the coil is halved to 5.00 cm. Draw a line graph representing the new output from the generator.

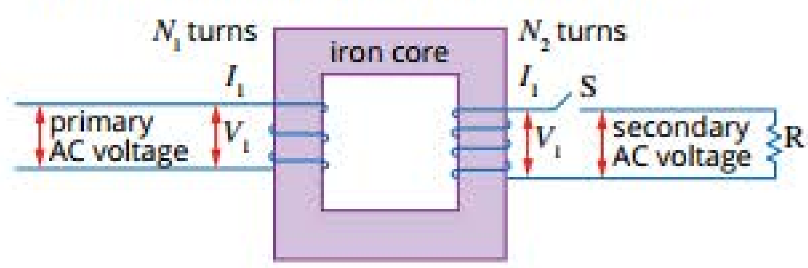
4 An AC supply of frequency 50.0 Hz is connected to a circuit, resulting in an rms current of 1.00 A being observed. Draw a graph that shows one full period of the variation of current with time for this circuit.

5 A voltage sensor is connected to the output of a transformer and a series of different inputs is used. Which of the following graphs is the most likely output displayed on a voltage graph for a steady DC voltage input?



6 A security light is operated from mains voltage of 240 V rms through a step-down transformer with 800 turns on the primary winding. The security light operates normally on an rms voltage of 12.0 V. How many turns are on the secondary coil?

7 The figure below depicts an iron core transformer. An alternating voltage applied to the primary coil produces a changing magnetic flux. The secondary circuit contains a switch, S, in series with a resistor, R. The number of turns in the primary coil is N_1 and in the secondary, N_2. The power in the first coil is P_1 and that in the second coil, P_2. Assume that this is an ideal transformer.



a Write an equation that defines the relationship between the power in the primary coil, P_1, and the power in the secondary coil, P_2.

b Write an equation that defines the relationship between the current in the secondary coil, I_2, and the current in the primary coil, I_1, in terms of the number of turns in each coil.

8 The primary windings of a transformer consist of 20 turns and the secondary of 200 turns. The primary rms voltage input is 8.00 V and a primary rms current of 2.00 A is flowing.

a What is the rms voltage across the load attached to the secondary coil?

b What power is being supplied to the load attached to the secondary coil?

c What is the rms current in the secondary coil?

9 A solar-powered generator produces 5.00 kW of electrical power at 500 V. This power is transmitted to a distant house via twin cables of total resistance 4.00 Ω. What is the total power loss in the cables?

10 A power station generates 500 MW of power to be used by a town 100 km away. The power lines between the power station and the town have a total resistance of 2.00 Ω.

a If the power is transmitted at 100 kV, what current would be required?

b What voltage would be available at the town? Give your answer in kilovolts (kV).

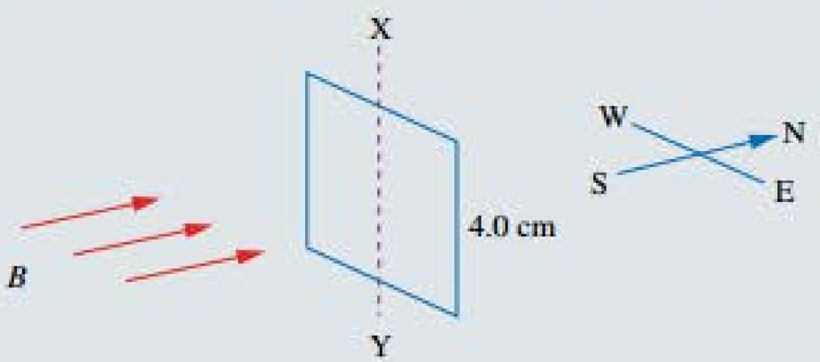
Chapter review

KEY TERMS

alternator
back emf
brushes
eddy current
electromagnetic induction
electromotive force
emf
Faraday's law
generator
ideal transformer
induced current
induction hotplate
Lenz's law
magnetic flux
magnetic flux density
regenerative braking
root mean square
sinusoidal
slip rings
split-ring commutator
step-down transformer
step-up transformer
transformer

The following information relates to questions 1 and 2.

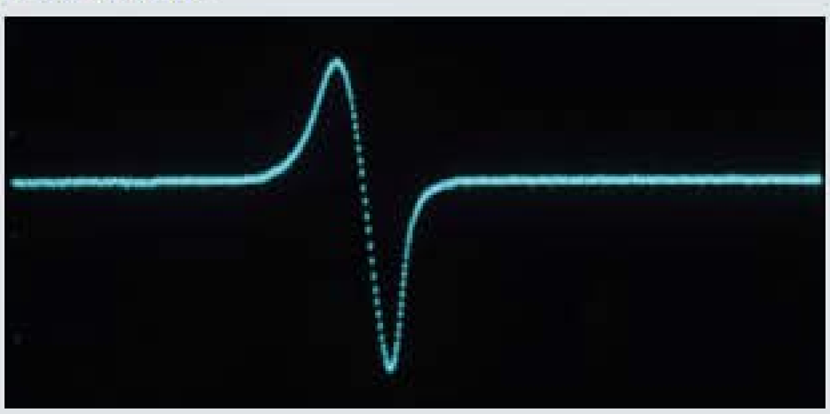
A square loop of wire of side 4.00 cm is in a region of uniform magnetic field, $B = 2.00 \times 10^{-3}$ T north, as in the diagram below. The loop is free to rotate about a vertical axis XY. When the loop is in its initial position, its plane is perpendicular to the direction of the magnetic field.



1 Calculate the magnetic flux passing through the loop.

2 Describe what happens to the amount of magnetic flux passing through the loop as the loop is rotated through one complete revolution.

3 When a magnet is dropped through a coil, a voltage sensor will detect an induced voltage in the coil as shown below.



The area under the curve above zero is exactly equal to the area above the curve below zero because:

A the strength of the magnet is the same

B the area of the coil is the same

C the strength of the magnet and area of the coil are the same

D the magnet speeds up as it falls through the coil

4 A student has a flexible wire coil of variable area of 100 turns and a strong bar magnet, which has been measured to produce a magnetic field of strength $B = 100$ mT, a short distance from it. The student has been instructed to demonstrate electromagnetic induction by using this equipment to light up an LED rated at 1.00 V.

Explain, including appropriate calculations, one method by which she could complete this task.

5 The back emf generated in a DC motor is the result of current produced in response to the rotation of the armature in the motor in the presence of an external magnetic field. As a result of the back emf, will the net emf used by a DC motor be:

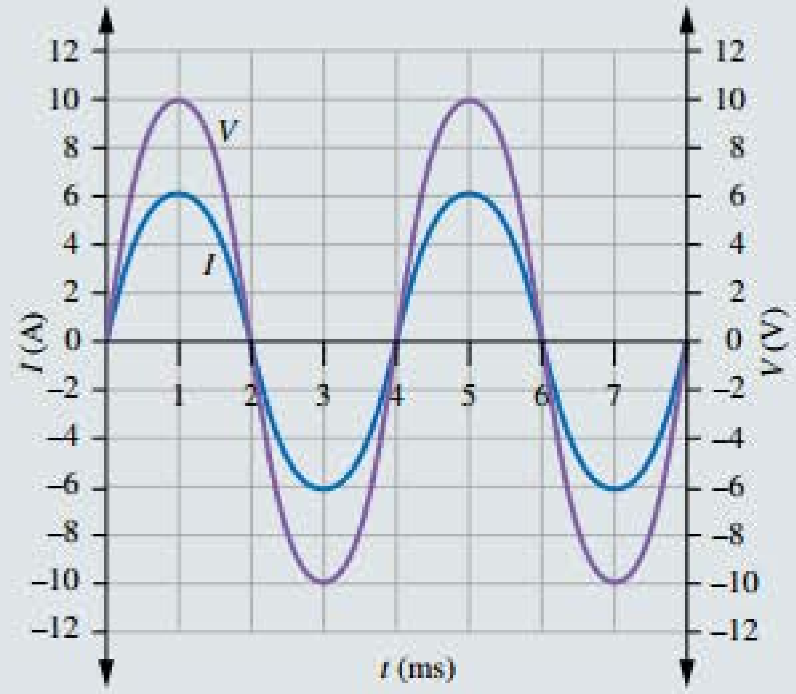
A the same as the supplied voltage

B less than the supplied voltage

C greater than the supplied voltage

D greater or less than the supplied voltage, depending on the speed of the motor

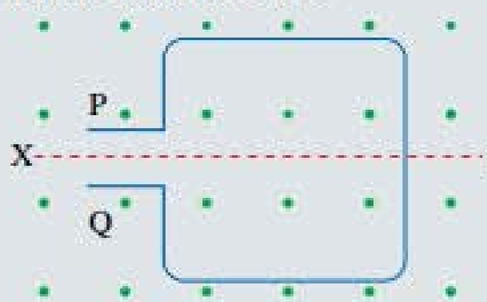
6 A student decides to test the output power of a new amplifier by using a voltage sensor to capture and display the alternating current I and voltage V that it produces. The result is shown below.



What is the rms power rating of the amplifier?

7 An electric toaster designed to operate at a V_{rms} of 240 V has a power rating of 600 W. Determine the peak current flowing through the heating element.

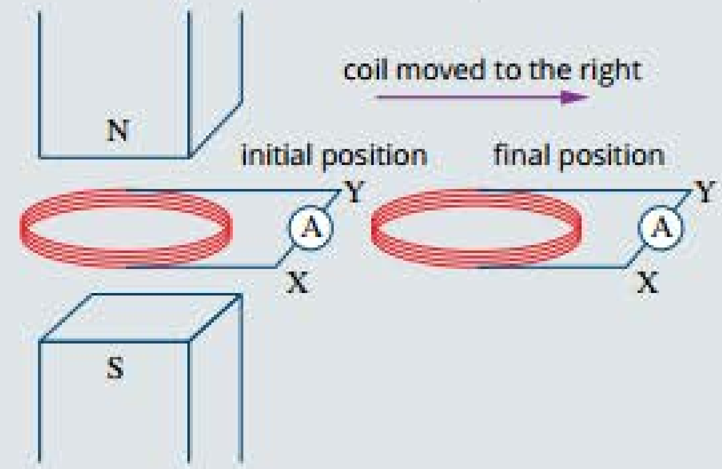
8 A rectangular coil of area 40.0 cm^2 and resistance 1.00 Ω is located in a uniform magnetic field $B = 8.00 \times 10^{-4}$ T that is directed out of the page. The plane of the coil is initially perpendicular to the field as depicted in the diagram below.



a Calculate the magnitude of the emf induced in the coil when the strength of the magnetic field is doubled in a time of 1.00 ms.

b Determine the direction of the current caused by the induced emf in the coil when the strength of the magnetic field is doubled in a time of 1.00 ms.

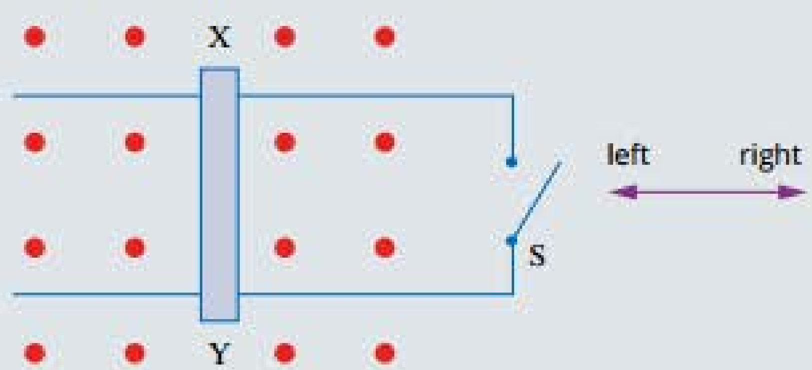
9 During a physics experiment a student pulls a horizontal circular coil from between the poles of two magnets in 0.100 s. The initial position of the coil is entirely in the field while the final position is free of the field. The coil has 40 turns, each of radius 4.00 cm. The field strength between the magnets is 20.0 mT.



a What is the magnitude of the average emf induced in the coil as it is moved from its initial position to its final position?

b What is the direction of the current through the ammeter caused by the induced emf?

10 A copper rod XY, of length 20.0 cm, is free to move along a set of parallel conducting rails as shown in the following diagram. These rails are connected to a switch, S, which completes a circuit when it is closed. A uniform magnetic field of strength 10.0 mT, directed out of the page, is established perpendicular to the circuit. S is closed and the rod is moved to the right with a constant speed of 2.00 m s^{-1}.



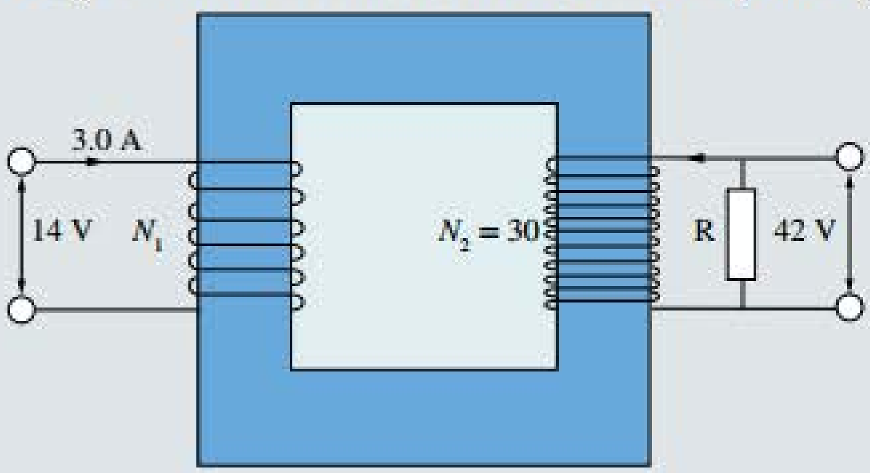
a Calculate the magnitude of the induced emf in the rod in mV.

b Determine the direction of the current through the rod caused by the induced emf.

11 A ship with a vertical steel mast of length 8.00 m is travelling due west at 4.00 m s^{-1} in a region where the Earth's magnetic field is horizontal and is equal to 5.00×10^{-5} T north. What average emf, in mV, would be induced between the top and the bottom of the mast?

12 Coils S_1 and S_2 are close together and linked by a soft iron core. The emf in S_1 varies as shown in the graph below. Draw a line graph to show the shape of the variation of the current in S_2.

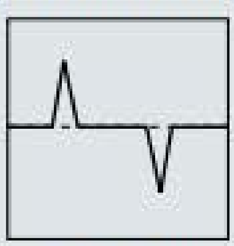
The following information relates to questions 13 and 14.
An ideal transformer is operating with an rms input voltage of 14.0 V and rms primary current of 3.00 A. The output voltage is 42.0 V. There are 30 turns in the secondary winding.



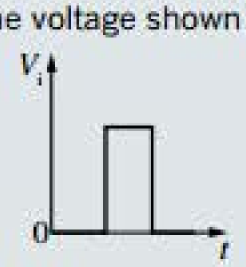
13 What is the rms output current?

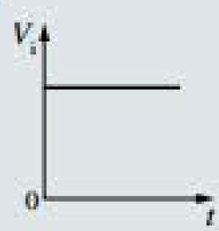
14 How many turns are there in the primary coil?

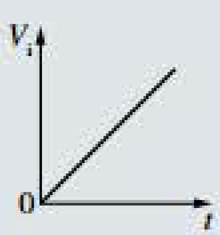
15 The following diagram shows a graph of induced voltage versus time as it appears on the screen of a CRO.



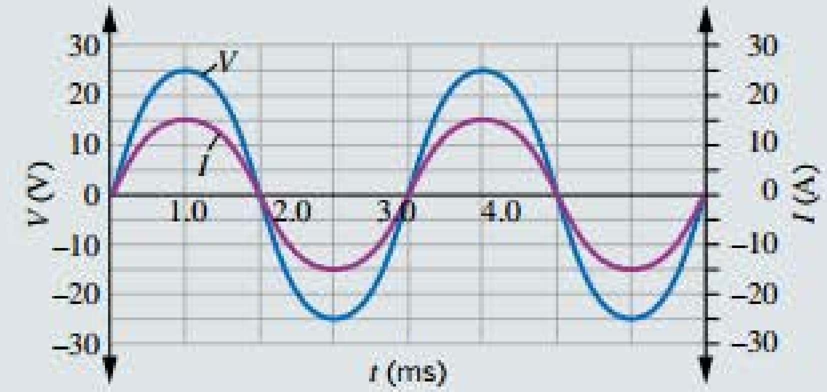
Which of the following input voltages would produce the voltage shown in the CRO display?

A


B


C


16 A physics student uses a voltage/current sensor to display the current, *I*, through, and the voltage, *V*, across the output terminals of a small generator. The graph obtained from the display is shown below.



a What is the approximate rms voltage for the signal?
b Calculate the peak power output of the generator.

17 A student decides to test the power output of a new stereo amplifier. The maximum rms power output guaranteed by the manufacturer (assumed accurate) is 60.0 W. Which set of specifications is consistent with this power output?

	Peak–peak voltage (V)	Peak–peak current (A)
A	20	3.0
B	40	6.0
C	40	12.0
D	20	6.0

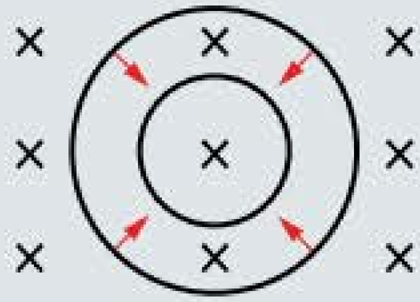
The following information refers to questions 18 and 19.
A student builds a simple alternator consisting of a coil containing 500 turns, each of area 10.0 cm^2, mounted on an axis that can rotate between the poles of a permanent magnet of strength 80.0 mT. The alternator is rotated at a frequency of 50.0 Hz.

18 Calculate the rms emf of the alternator.

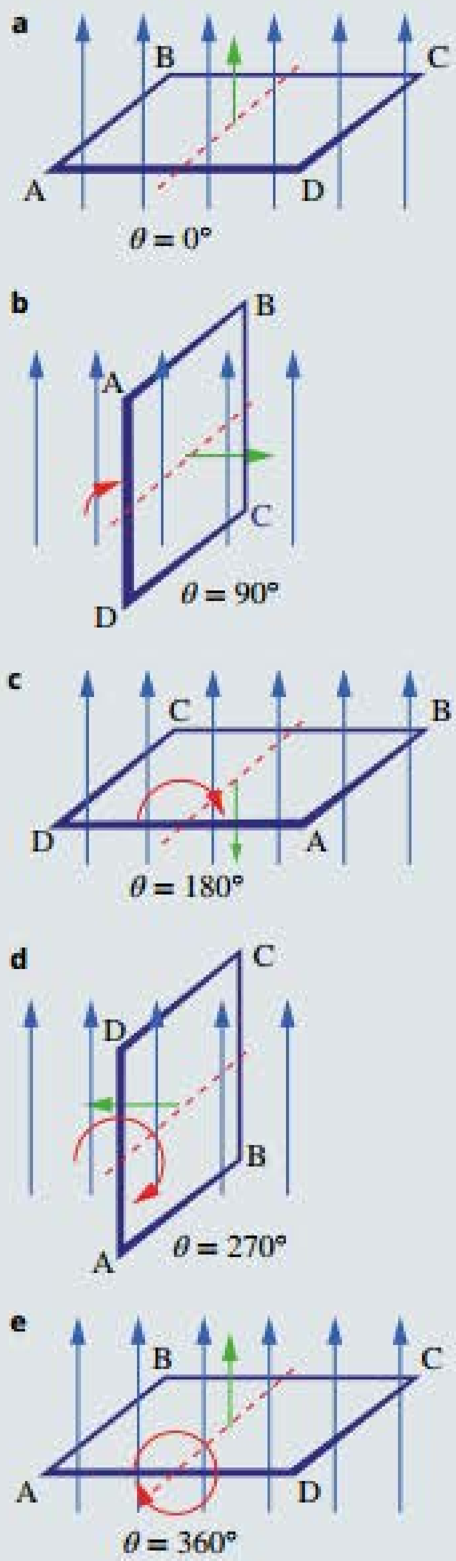
19 Explain what the effect will be on the rms emf when the frequency is doubled to 100 Hz.

20 A generator is to be installed in a farm shed to provide 240 V power for the farmhouse. A twin-conductor power line with total resistance 8.00 Ω already exists between the shed and house. The farmer has seen a cheap 240 V DC generator advertised and is tempted to buy it.
Identify and explain two significant problems that you foresee with using the 240 V DC generator.

21 A coil in a magnetic field directed into the page is reduced in size. In what direction will the induced current flow in the coil while the coil is being reduced in size?



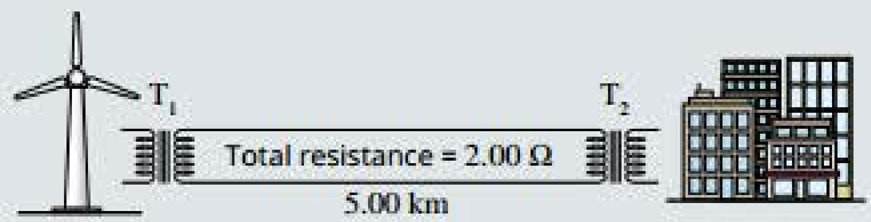
22 A single loop of wire is rotated within a magnetic field, *B*, as shown below.



While the coil rotates, an emf is generated as a result of *which* sides of the coil? Give a reason for your answer.

The following information relates to questions 23–26.

A wind turbine runs a 150 kW generator with an output voltage of 1000 V. The voltage is increased by a transformer T_1 to 10 000 V for transmission to a town 5.00 km away through power lines with a total resistance of 2.00 Ω. At the town, another transformer, T_2, reduces the voltage to 250 V. Assume that there is no power loss in the transformers (i.e. they are 'ideal').



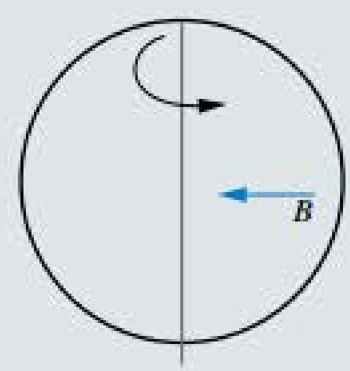
23 Calculate the current in the power lines.

24 Calculate the voltage at the input to the town transformer T_2.

25 Determine the power lost in the power lines.

26 It is suggested that some money could be saved from the scheme by removing the first transformer. Explain, using appropriate calculations, whether this is a good plan.

27 A coil is rotated about its vertical axis such that the left-hand side would be coming out of the page and the right-hand side would be going into it. A magnetic field runs from right to left across the page. In which direction would the induced current in the coil flow?



28 A non-ideal transformer has a slightly smaller power output from the secondary coil than input to the primary coil. The voltage and current in the primary coil are denoted V_1 and I_1 respectively. The voltage and current in the secondary coil are denoted V_2 and I_2 respectively. Which of the following expressions describes the power output in the secondary coil?

A $V_1 I_1$
B $V_2 I_2$
C $V_1 I_2$
D $I_2^2 R$

29 A security light is connected to a mains voltage of 240 V rms. It runs on a voltage of 12.0 V rms and an rms current of 2.00 A. A step-down transformer with 800 turns on the primary winding is used to reduce the voltage from the mains level to the required operating voltage. Assume that the light is operating normally and that there is no power loss in the transformer.

a Calculate the number of turns, to the nearest whole number, in the secondary coil.

b What is the value of the peak current in the primary coil?

c Calculate the rms power input to the primary coil of the transformer.

d During some routine maintenance work, the primary coil of the transformer for the security light is unplugged from the AC mains supply and plugged into a DC supply of 240 V instead. What is the new output (secondary) voltage?

A 0 V

B 12 V

C 24 V

D 240 V

30 A 100 km transmission line made from aluminium cable has a total resistance of 10.0 Ω. The line carries the electrical power from a 500 MW power station to a substation. If the line is operating at 250 kV, what is the power loss in the line?

31 Power loss can be expressed by the formula $P_{loss} = \frac{\Delta V^2}{R} = I^2R$. Select which of the following statements is true, and justify why the other response is incorrect.

A The greater the voltage being transmitted in a transmission line, the greater the power loss.

B The greater the current in the transmission line, the greater the power loss.

REVIEW QUESTIONS

Section 1: Short response

1 Use Newton's law of gravitation to calculate the size of the force between two masses of 24.0 kg and 903 kg with a distance of 0.670 m between their centres.

2 The orbital period of Mars around the Sun is 687 days at an average distance of 227.9 million kilometres. If the distance between Jupiter and the Sun is 778.5 million kilometres, what is Jupiter's orbital period in days?

3 A 21.0 kg child slides down a curved frictionless slide from 3.50 m off the ground with no initial kinetic energy.

a Calculate the speed of the child as she reaches the bottom of the slide.

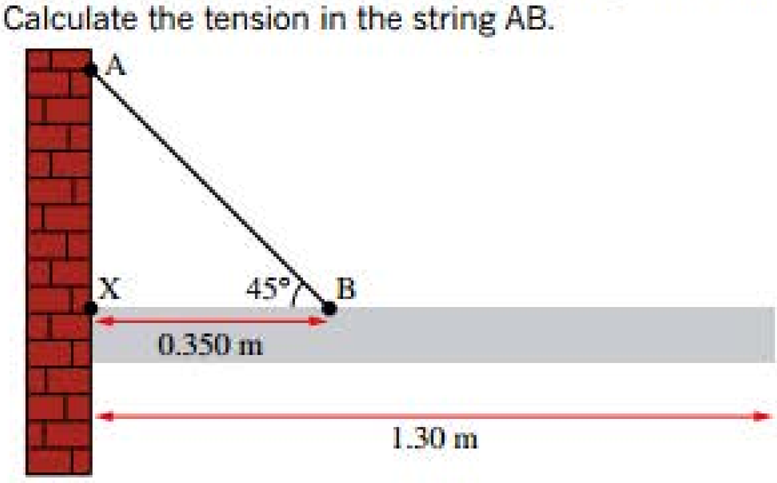
b Assume now that the slide has some friction. If the child is travelling at 5.00 m s^{-1} when she reaches the bottom, calculate the energy lost to friction.

4 Two cannonballs are launched from level ground with a speed of 35.0 m s^{-1} at an angle of 40.0° to the horizontal. Assume there is no air resistance.

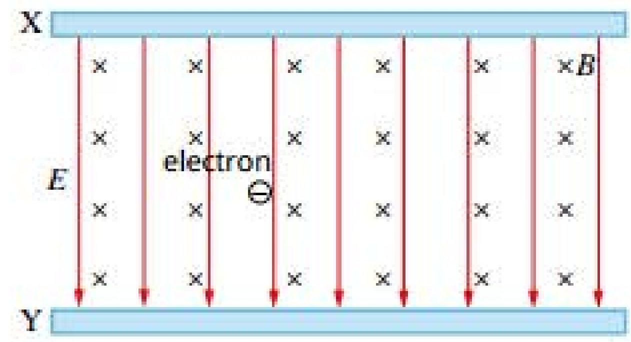
a Determine the speed and direction of the first cannonball the instant before it hits a target on the ground.

b The second cannonball is fired from the edge of a cliff. Using conservati on of energy, calculate the speed of the cannonball when it is 30.0 m below its initial launch level.

5 A cantilever is attached to a wall as shown. The weight of the beam is 100 g and it is connected to a freely rotating hinge on the wall at X. The string is attached 0.350 m along the 1.30 m beam. Assume the string has no mass. Calculate the tension in the string AB.



6 An electron beam travelling through a cathode ray tube at a speed of 2.00 × 10^{7} m s^{-1} is subjected to simultaneous electric and magnetic fields. The electric and magnetic fields are oriented perpendicular to each other such that the electrons emerge with no deflection. Given that the potential difference across the parallel plates X and Y is 3.00 kV, and that the applied magnetic field is of strength 1.60 × 10^{-3} T, calculate the distance between plates X and Y.



7 A long bar magnet is dropped, with its north pole first, through a conducting coil and induces a current. The coil is connected to an oscilloscope, which measures the amplitude and direction of the current. If positive is taken to be a clockwise current when viewed from above, and negative is an anticlockwise current, sketch the shape of the current signal that will be observed on the oscilloscope over time.

8 a Draw a solenoid with at least seven magnetic field lines in each direction. Indicate the direction of current and the north and south poles in your sketch.

b List three ways to improve the strength of the solenoid's magnetic field.

c If two solenoids are placed next to each other with current running in opposite directions, will they attract or repel each other? Why?

Section 2: Problem solving

9 An elaborate remote-control car track is set up at a department store, and the employees are trying to calculate the speeds and specifications necessary for the track to work optimally. The track is set up as shown below with a banked curve at section A and a loop-the-loop section at B. The points g and f have also been labelled at the centre of the banked section and at the top of the loop-the-loop. The mass of the car is 0.800 kg. Section A can be taken as a perfect circle of radius 2.50 m, with a banked angle θ; section B is a vertical loop-the-loop with a radius of 0.500 m.

a Draw the diagram in your workbook. Sketch the force vector acting on one of the cars at point g.

b If the employees want the cars to be going at least 5.00 m s^{-1} in section A of the track, with a guarantee of not sliding off, what is the *minimum* angle at which they should bank the track?

c Calculate the normal force acting on the car as it travels around the banked track.

d If the car was to travel at a speed lower than the critical speed mentioned in part (b), draw the frictional force of the road on the car required for the car to maintain its position on the track.

e The car is now in position f. Modelling the car as a point, calculate the minimum speed at the top of the loop needed for the car to maintain contact with the track.

10 As of February 2016, a total of 72 satellites had been launched for the purpose of maintaining the global positioning system. Twenty-four such satellites are necessary for the total GPS constellation (a group of satellites working in unison); the rest are either in reserve, or no longer operational. These satellites orbit the Earth at a radius slightly more than half of a geosynchronous orbit.

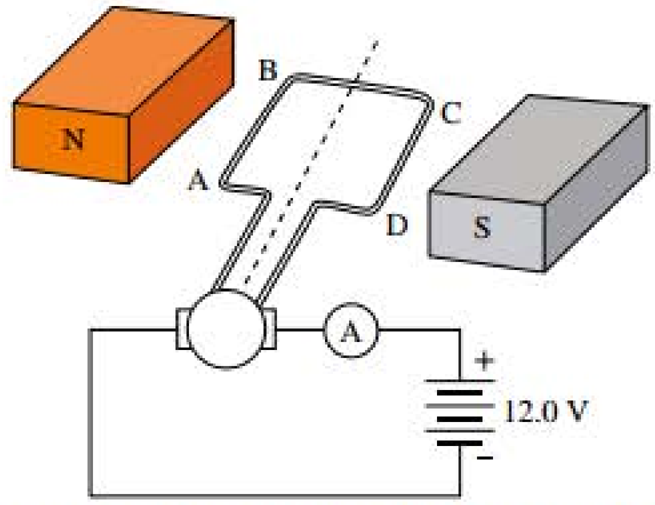
a Briefly define what is meant by a geosynchronous or geostationary orbit.

b Calculate how far in kilometres above the surface of the Earth a satellite would need to be if it was to sustain a geostationary orbit.

c If one of these GPS satellites orbits at an altitude of 20 143 km, calculate its orbital speed.

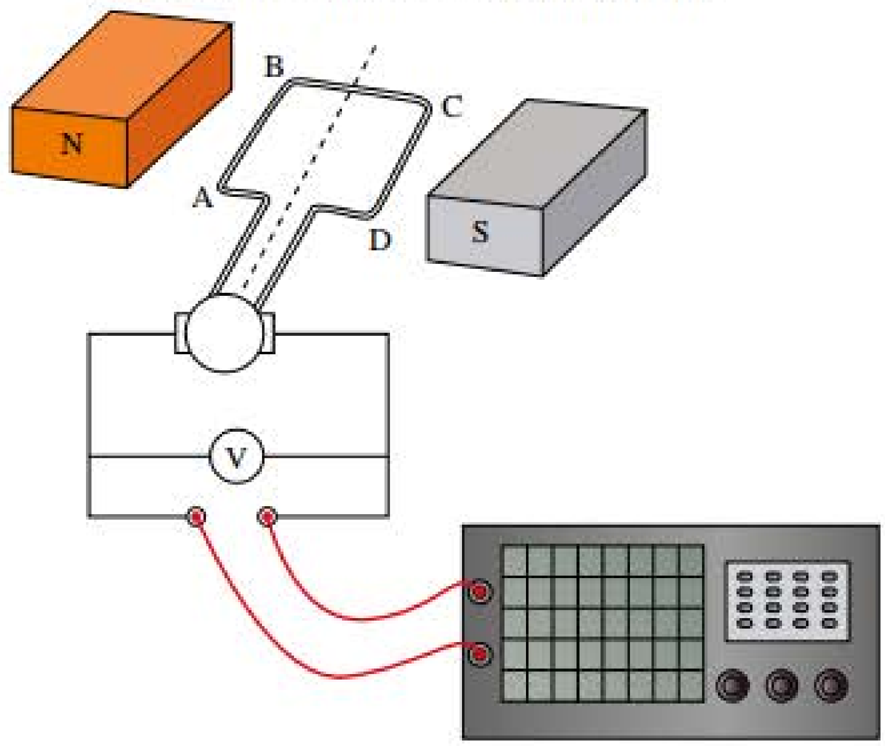
d In order for GPS satellites to calculate position, a satellite must keep time with accuracy of approximately 50 nanoseconds per day.
Find the error in the distance calculated for the radio signal to travel from the satellite to Earth, given the error in time accumulated over a single day.

11 A group of physics students set up a device that can operate as either a motor or a generator, depending on the connections used. The device is set up as shown, with a uniform magnetic field of 1.50×10^{-3} T. ABCD is a square coil of side length 8.00 cm with 25 turns and total resistance $R = 2.00\,\Omega$. The setup is connected to a battery of 12.0 V, with an ammeter and a commutator.



a With the coil as shown in the above diagram, calculate the force on side AB, and give the direction (up, down, left, right).

b The device is now set up as a DC generator, with the circuit now connected to a voltmeter and an oscilloscope as shown. The coil and magnetic field are the same as in part (a). If the coil is rotated at a constant rate, sketch a rough graph of the output expected to be shown on the oscilloscope.



c The shaft and coil makes one half-turn every 20.0 ms. Calculate the magnitude of the average voltage produced in a quarter of a turn.

d The students now decide to replace the split-ring commutator with slip rings and test out their motor again. Explain in detail what will now happen to the coil.

12 A farmer has installed a wind generator on a nearby hill and a power line with a total resistance of 13.0 Ω. The output of the generator is rated as 250 VAC (rms) with a maximum power of 4.00 kW. The farmer connects up the system and finds that the voltage at the house is indeed 250 V. However, when she turns on various appliances so that the generator is running at its maximum power output of 4.00 kW, she finds that the voltage supplied at the house is rather low.

a Explain why the voltage dropped when the farmer turned on the appliances in the house.

b Calculate the voltage and power at the house when the appliances are turned on.

The farmer then decides to install ideal transformers at either end of the same power line so that the voltage transmitted from the generator end (after the transformer) in this system becomes 5000V.

c Describe the essential features of the types of transformers that are needed at either end of the power line.

Assume the generator is operating at full load, i.e. 4000W, when answering parts d–j.

d Calculate the current in the power line with the transformers when the same appliances are turned on.

e Calculate the voltage drop along the power line.

f Calculate the power loss in the power line.

g Calculate the potential delivered to the transformer near the house.

h Calculate the power delivered to the transformer near the house.

i How do the power losses in the system without the transformers compare to those for the system with the transformers, as a percentage of the power generated?

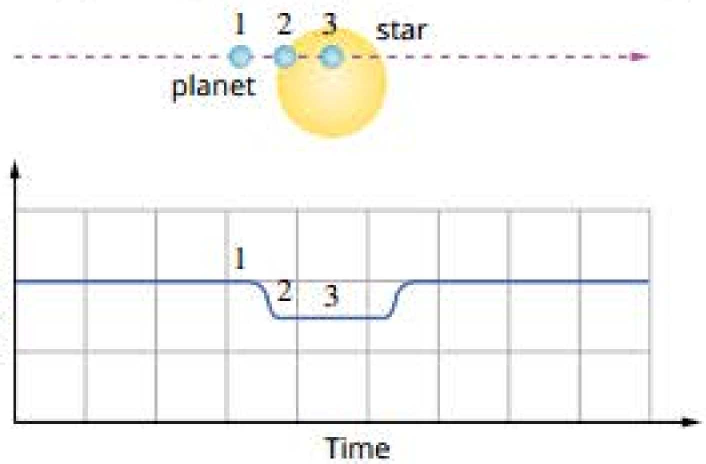
j Explain why the system operated with much lower power losses when the voltage was transmitted at the higher voltage.

Section 3: Comprehension

13 In May of 2016, an international team of scientists led by the University of Liege discovered a set of exoplanets (planets outside of the solar system) around the TRAPPIST-1 stellar system. The properties of the exoplanets were studied by intense surveillance by telescopes around the world—including NASA's Spitzer Space Telescope—and in February 2017 the findings were excitedly announced as the single largest discovery of rocky planets and potentially habitable worlds to date. TRAPPIST-1 is an ultra-cool dwarf star located approximately 39.5 light years away from Earth with just 8% of the mass of the Sun. The initial exoplanet discovery suggested three Earth-sized planets orbiting the star; however, later studies confirmed a total of seven planets in the system, three of which are considered to be in the star's so-called 'habitable zone'. At this stage, the names of the seven planets are TRAPPIST-1b to TRAPPIST-1h in order from closest to furthest away from the star.

Astronomers use a myriad of techniques to measure and calculate the properties of faraway celestial bodies, most of which are surprisingly simple in concept. Each of them relies, however, on knowing properties of their central star, such as mass, distance, size and luminosity, to a high degree of precision. Using these elementary techniques, scientists can calculate the properties of these exoplanets.

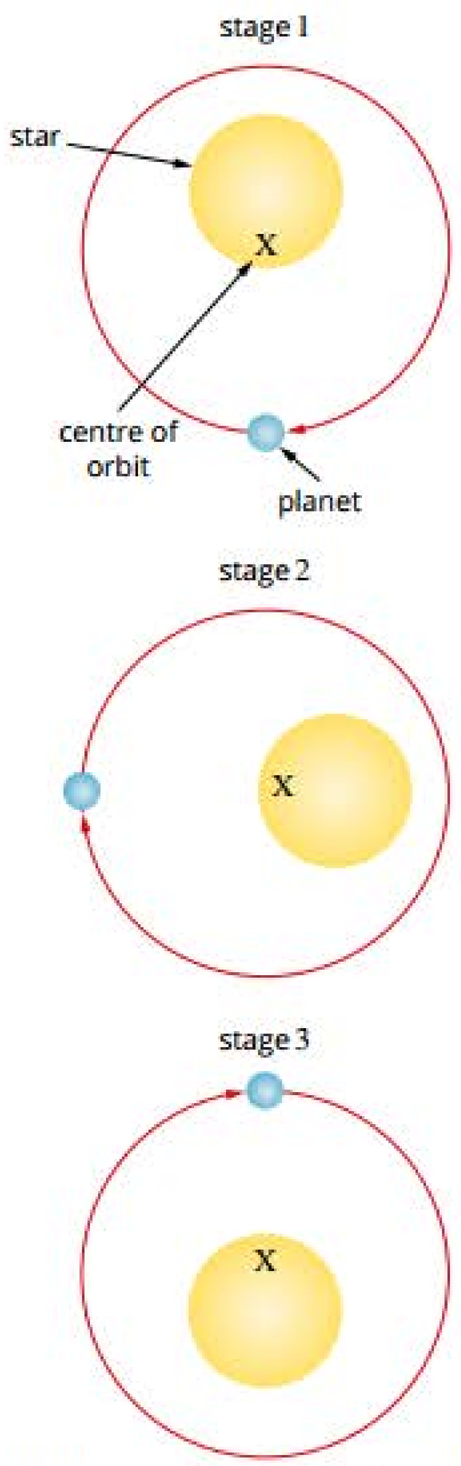
One such of these techniques, called transit photometry, enables us to determine the size of different planets. Astronomers observe the exoplanets making a transit across the face of the star. When these planets pass across the star, the brightness of the star decreases by a predictable amount that is directly related to the size of the planet. By studying the apparent luminosity of the star from Earth for a number of transits of a given planet, astronomers can reliably estimate the size of that planet. It is the most accurate current method of determining planetary size, but has the drawback of needing to wait for a planet to pass across of the face of the star (an event that may last less than an hour).



Transit photometry provides information about the planet's size, but not its mass. Another technique called Doppler spectroscopy provides information about a planet's mass, but not its size. As a planet orbits circularly around a star, the star will also orbit around the shared centre of orbit, which is usually somewhere inside the star itself (see diagram on page 228).

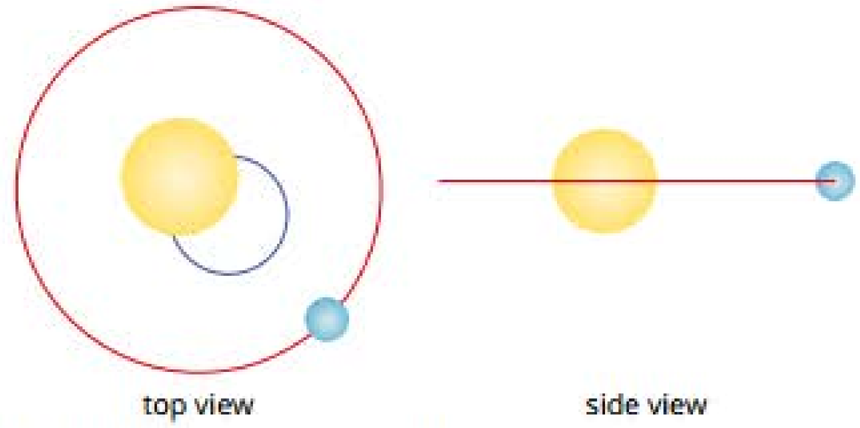
The periodic variations in the star's velocity towards and away from Earth are exactly the period of the orbit, and can be measured by observing the Doppler shift of the star. This radius is then used in Kepler's third law to calculate the distance of the planet to the star, which can be similarly used in Newton's law of gravitation to calculate the velocity of the planet. Finally, the relationship

$\frac{M_{planet}}{v_{star}} = \frac{M_{star}}{v_{planet}}$ can be used to find the mass of the planet, assuming the mass of the star is already known. It is important to note that the velocity detected by astronomers is not necessarily the total velocity of the star, but the *radial* velocity (how fast it moves towards and away from Earth).



These two methods, together, allow scientists to determine the density of an exoplanet, which is important for learning its composition.

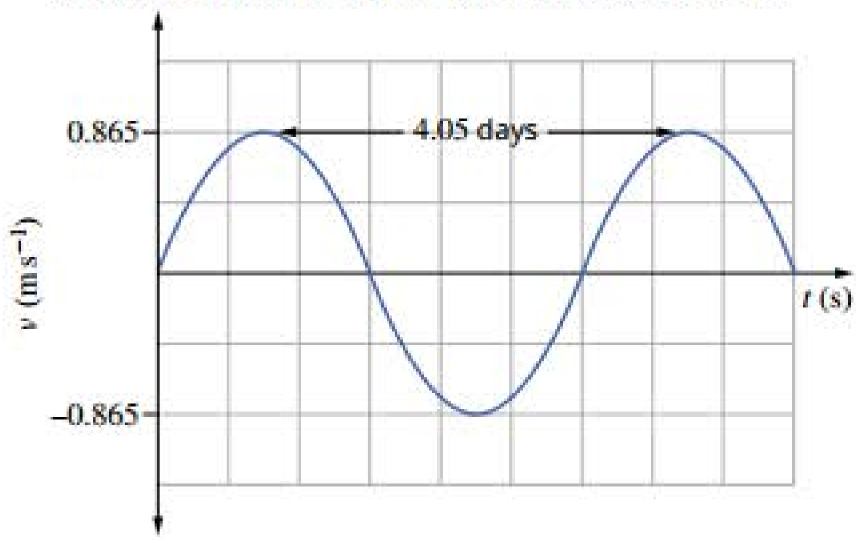
a The orientation of the orbital of the planets with respect to Earth is important in the accuracy of the Doppler spectroscopy method. Which of the following views of a planet orbiting a star would be a more ideal view of the stellar system in this respect? Give reasons for your answer.



b **i** In light of your answer for part (a), explain using words and formulae why the Doppler spectroscopy method can, in general, only provide an estimate for the *minimum* possible mass of an exoplanet for all different orientations of stellar systems.

ii Given that the TRAPPIST researchers observed transits of the different planets in front of the star, in an idealised sense would you expect their mass calculations to be accurate, or have a high margin of error?

c Below is an idealised plot of the hypothetical velocity variation of TRAPPIST-1 in its orbit with TRAPPIST-1d, obtained by using Doppler spectroscopy. Given that the mass of TRAPPIST-1 is 1.59×10^{29} kg, determine a minimum estimate for the mass of TRAPPIST-1d.



UNIT 4 Revolutions in modern physics

The development of the quantum theory and the theory of relativity fundamentally changed our understanding of how nature operates, and led to the development of a wide range of new technologies, including technologies that revolutionised the storage, processing and communication of information. In this unit, students examine observations of relative motion, light and matter that could not be explained by existing theories, and investigate how the shortcomings of existing theories led to the development of the special theory of relativity and the quantum theory of light and matter. Students evaluate the contribution of the quantum theory of light to the development of the quantum theory of the atom, and examine the Standard Model of particle physics and the Big Bang theory.

Learning outcomes

By the end of this unit, students:

- understand the consequences for space and time of the equivalence principle for inertial frames of reference
- understand how the quantum theory of light and matter explains black body radiation, the photoelectric effect, and atomic emission and absorption spectra
- use the Standard Model to describe the nature of and interaction between the fundamental particles that form the building blocks of matter
- understand how models and theories have developed over time, and the ways in which physical science knowledge and associated technologies interact with social, economic, cultural and ethical considerations
- use science inquiry skills to design, conduct, analyse and evaluate investigations into frames of reference, diffraction, black body and atomic emission spectra, the photoelectric effect, and photonic devices, and to communicate methods and findings
- use algebraic and graphical models to solve problems and make predictions related to the theory and applications of special relativity and quantum theory
- evaluate the experimental evidence that supports the theory of relativity, wave–particle duality, the Bohr model of the atom, the Standard Model, and the Big Bang theory
- communicate physics understanding using qualitative and quantitative representations in appropriate modes and genres.

CHAPTER

07 Wave–particle duality and the quantum theory

Discovering the nature of light has been one of the scientific community's greatest challenges and has involved the giants of physics such as Galileo, Newton, Poisson, Young, Fresnel, Maxwell, Heisenberg, Planck, Bohr, de Broglie, Einstein, Schrodinger and others. Over the course of history, light has been compared to a geometric ray, a stream of particles and even a series of waves. However, these relatively simple models have been found to be limited in their ability to explain all of the properties of light. In the 19th century, Young's experiment and other evidence caused scientists to develop a wave model for light. Other observations, such as blackbody radiation, the photoelectric effect and absorption and emission spectra, led scientists such as Einstein and Planck to conclude that light has a particle-like nature and to the development of the field of quantum mechanics. Then, ultimately, scientists were led to conclude that light and matter have more in common than originally thought, and both exhibit a wave–particle duality.

Science as a Human Endeavour

Models that were initially rejected can be revisited as more evidence becomes available. For many years, the presence of the luminiferous ether was proposed as the medium by which light is propagated. Around 1800, Thomas Young showed through experimentation that light passing through a double slit showed interference and thus wave properties. The wave explanation of Young's double slit demonstration was initially rejected until other physicists, including Fresnel and Poisson, showed that light was able to undergo diffraction, a property of waves. Later, in the 1860s, James Clerk Maxwell developed a theory of electromagnetism and showed that electromagnetic waves would travel through space at the speed of light, implying light was an electromagnetic wave.

The use of devices developed from the application of quantum physics, including the laser and photovoltaic cells, have significantly changed many aspects of society.

Science Understanding

- light exhibits many wave properties; however, it cannot only be modelled as a mechanical wave because it can travel through a vacuum
- a wave model explains a wide range of light-related phenomena, including reflection, refraction, dispersion, diffraction and interference; a transverse wave model is required to explain polarisation
- electromagnetic waves are transverse waves made up of mutually perpendicular, oscillating electric and magnetic fields
- oscillating charges produce electromagnetic waves of the same frequency as the oscillation; electromagnetic waves cause charges to oscillate at the frequency of the wave
- atomic phenomena and the interaction of light with matter indicate that states of matter and energy are quantised into discrete values

- on the atomic level, electromagnetic radiation is emitted or absorbed in discrete packets called photons. The energy of a photon is proportional to its frequency. The constant of proportionality, Planck's constant, can be determined experimentally using the photoelectric effect and the threshold voltage of coloured LEDs

  This includes applying the relationships

  $$c = f\lambda,\ E = hf = \frac{hc}{\lambda},\ \ E_k = hf - W,\ \text{de Broglie } \lambda = \frac{h}{p}$$

- a wide range of phenomena, including blackbody radiation and the photoelectric effect, are explained using the concept of light quanta
- atoms of an element emit and absorb specific wavelengths of light that are unique to that element; this is the basis of spectral analysis

  This includes applying the relationships

  $$\Delta E = hf,\ E_2 - E_1 = hf$$

- the Bohr model of the hydrogen atom integrates light quanta and atomic energy states to explain the specific wavelengths in the hydrogen spectrum and in the spectra of other simple atoms; the Bohr model enables line spectra to be correlated with atomic energy-level diagrams
- on the atomic level, energy and matter exhibit the characteristics of both waves and particles. Young's double slit experiment is explained with a wave model but produces the same interference and diffraction patterns when one photon at a time or one electron at a time are passed through the slits

7.1 Properties of waves in two dimensions

One of the great scientific achievements of the 19th century was the development of a comprehensive wave model for light. This model was able to explain a large number of wave properties including reflection, refraction, dispersion (as shown in Figure 7.1.1), diffraction, interference and polarisation. This also led to a deeper understanding of phenomena such as heat and radio transmissions.

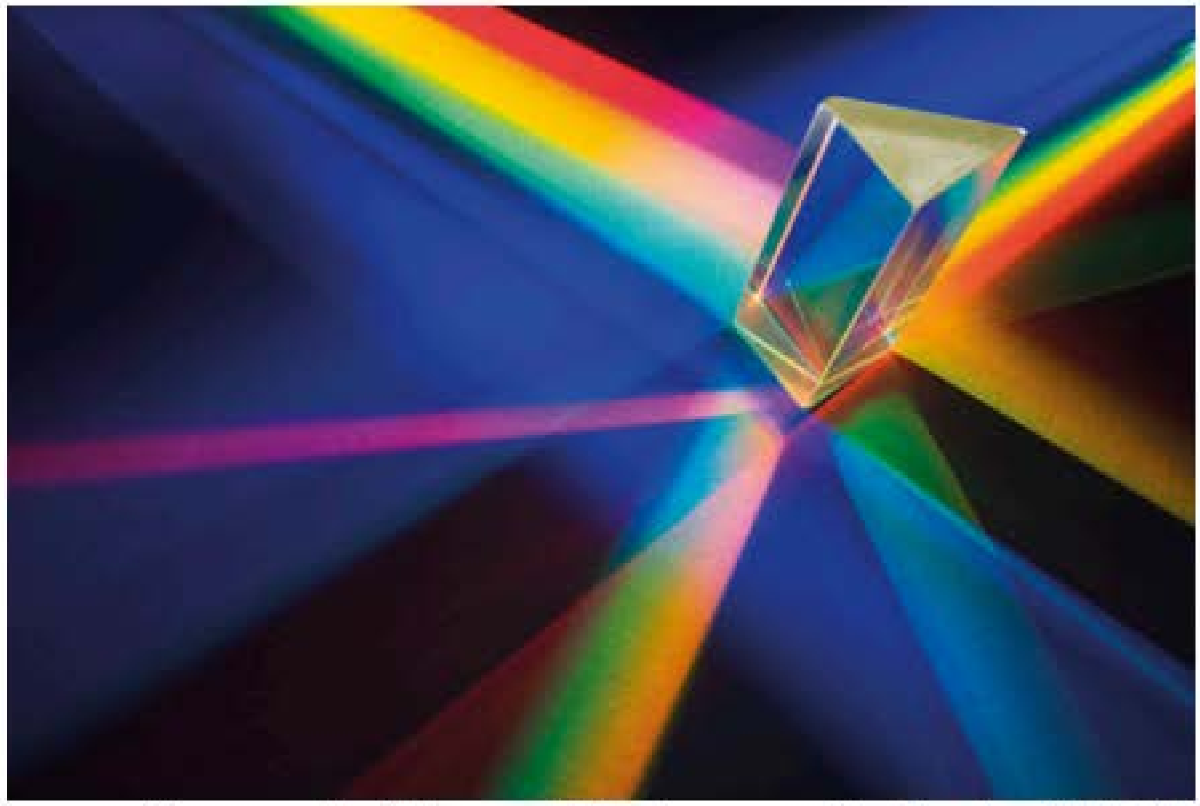


FIGURE 7.1.1 The wave model of light can explain the phenomenon of the dispersion of light into its component colours.

WAVE MODEL VERSUS PARTICLE MODEL

In the late 17th century a debate raged among scientists about the nature of light.

The famous English scientist Sir Isaac Newton explained light in terms of particles or 'corpuscles', with each different colour of the spectrum representing a different type of particle. Scientists Robert Hooke (from England) and Christiaan Huygens (from the Netherlands) proposed an alternative model that described light as a type of wave, similar to the water waves observed in the ocean.

A key point of difference between the two theories was that Newton's 'corpuscular' theory suggested that light would speed up as it travelled through a solid material such as glass. In comparison, the wave theory predicted that light would be slower in glass than in air.

Unfortunately, at that time it was impossible to measure the speed of light accurately, so the question could not be resolved scientifically. Newton's esteemed reputation meant that for many years his corpuscular theory was considered correct.

It was not until the early 19th century that experiments first convincingly demonstrated the wave properties of light.

Today, a modern understanding of light draws on aspects of both theories and is, perhaps, more complex than Newton, Hooke or Huygens could ever have imagined.

HUYGENS' PRINCIPLE

The theoretical basis for wave propagation in two dimensions was first explained by the Dutch scientist Christiaan Huygens. Huygens' principle states that each point on a wavefront can be considered as a source of secondary wavelets (i.e. small waves).

Consider the plane wave shown in Figure 7.1.2. Each point on the initial wavefront can be treated as if it is a point source producing circular waves, some of which are shown in green. After one period, these circular waves will have advanced by a distance equal to one wavelength. Huygens proved mathematically that when the amplitudes of each of the individual circular waves are added, the result is another plane wave as shown by the new wavefront.

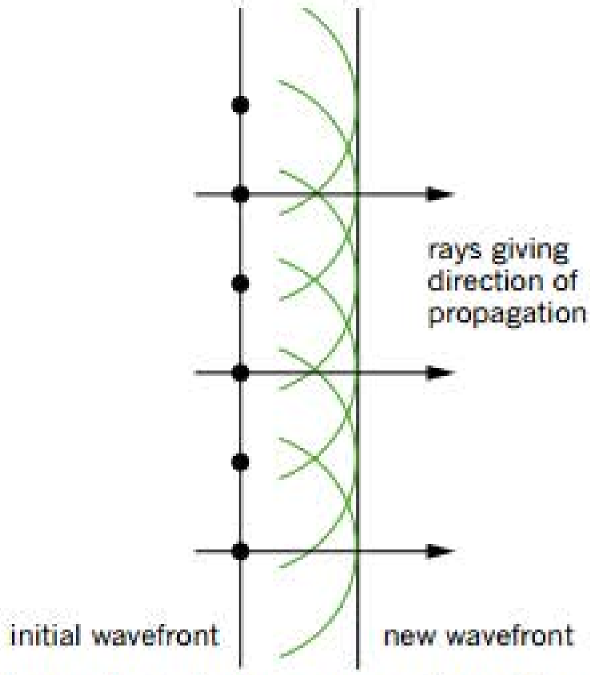


FIGURE 7.1.2 Each point on the wavefront of a plane wave can be considered as a source of secondary wavelets. These wavelets combine to produce a new plane wavefront.

This process is repeated at the new wavefront, causing the wave to propagate in the direction shown.

Circular waves are propagated in a similar way, as shown in Figure 7.1.3.

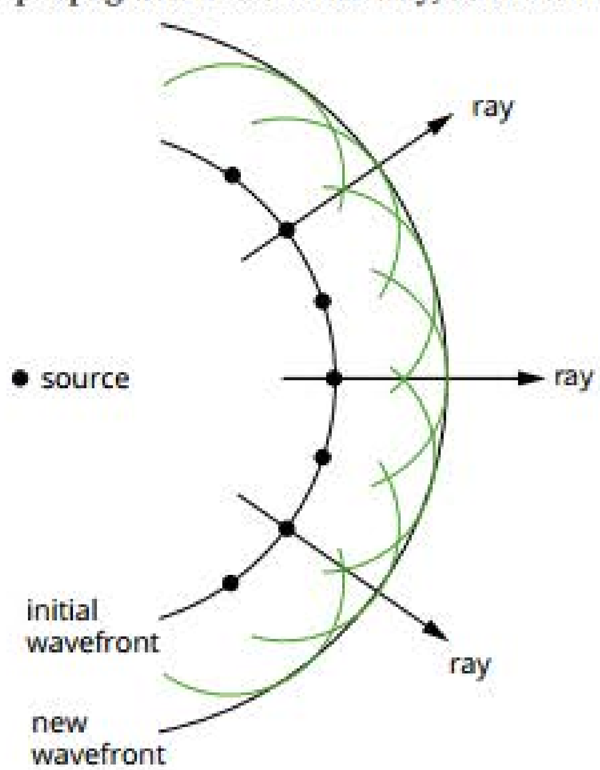


FIGURE 7.1.3 Each point on the wavefront of a circular wave can be considered as a source of secondary wavelets. These wavelets combine to produce a new circular wavefront.

A wave model can be used to explain a number of important properties of light including:

- reflection
- refraction
- dispersion
- diffraction
- polarisation.

REVISION

Properties of transverse waves

Transverse waves oscillate in a direction perpendicular to the direction of propagation. Figure 7.1.4(a) shows a transverse wave plotted against distance from the source. The amplitude is the displacement from the average or rest position. One complete cycle or oscillation is called the wavelength (λ). If the wave is plotted against time one complete cycle is the period (T), as shown in Figure 7.1.4(b).

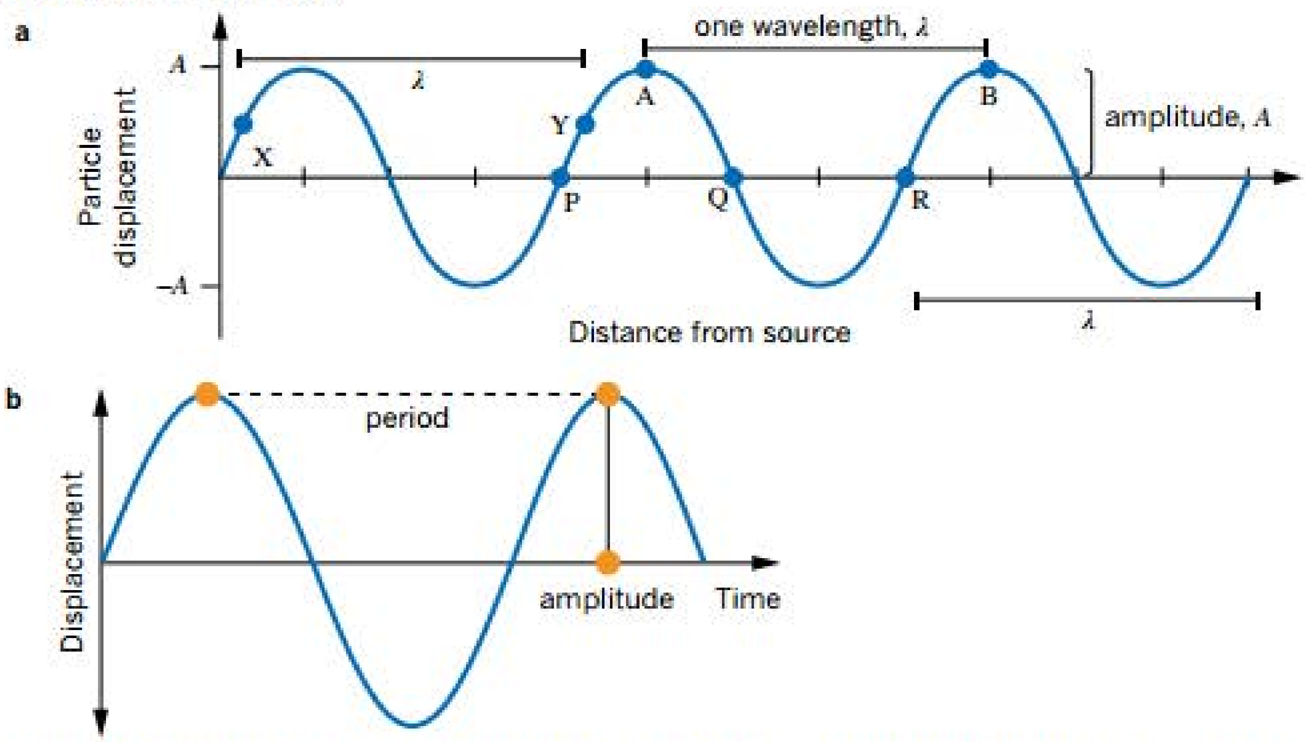


FIGURE 7.1.4 (a) A transverse wave plotted against distance from the source where one complete cycle or oscillation is the wavelength (λ). (b) The wave plotted against time where one complete cycle is the period (T).

The frequency, f, is the inverse of the period: $f = \frac{1}{T}$ (s^{-1}).

The velocity of the wave, v, is related by $v = f\lambda$ ($m\,s^{-1}$).

Two waves can add together to form a composite wave; this is called superposition or interference. Two important cases of this are complete constructive interference, where two identical, and hence coherent, waves that are in phase add together to give double the amplitude, as shown in Figure 7.1.5(a), and complete destructive interference, where the second wave is $\frac{\lambda}{2}$ out of phase and the total amplitude adds to zero, as shown in Figure 7.1.5(b). Two wave sources are coherent if they have the same wavelength and are out of phase by a constant amount, as in these examples.

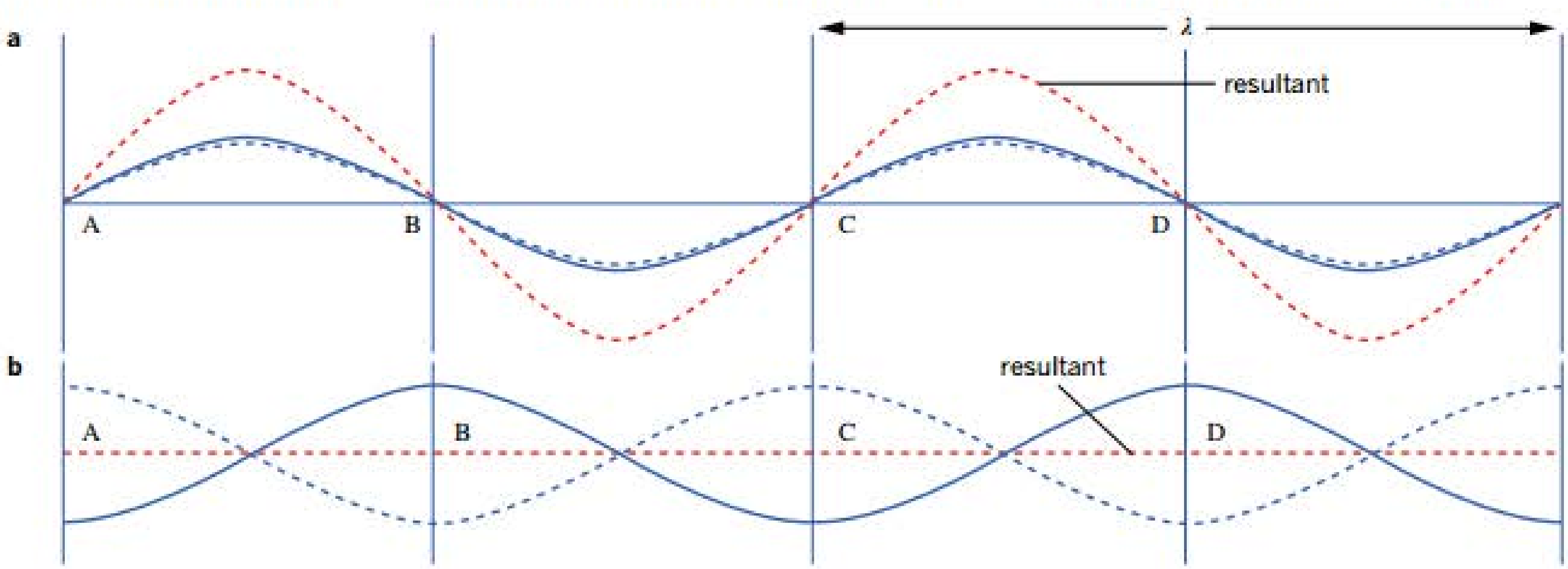


FIGURE 7.1.5 (a) Constructive interference: Two identical waves (blue lines) superpose (or add) together to form a wave with double the amplitude (red line). (b) Destructive interference: The two waves are $\frac{\lambda}{2}$ out of phase, and add together to give zero amplitude.

REFLECTION

Reflection of light can be explained by using both a ray model and a wave model, similar to that of sound waves (covered in *Pearson Physics 11 Western Australia*), as shown in Figure 7.1.6.

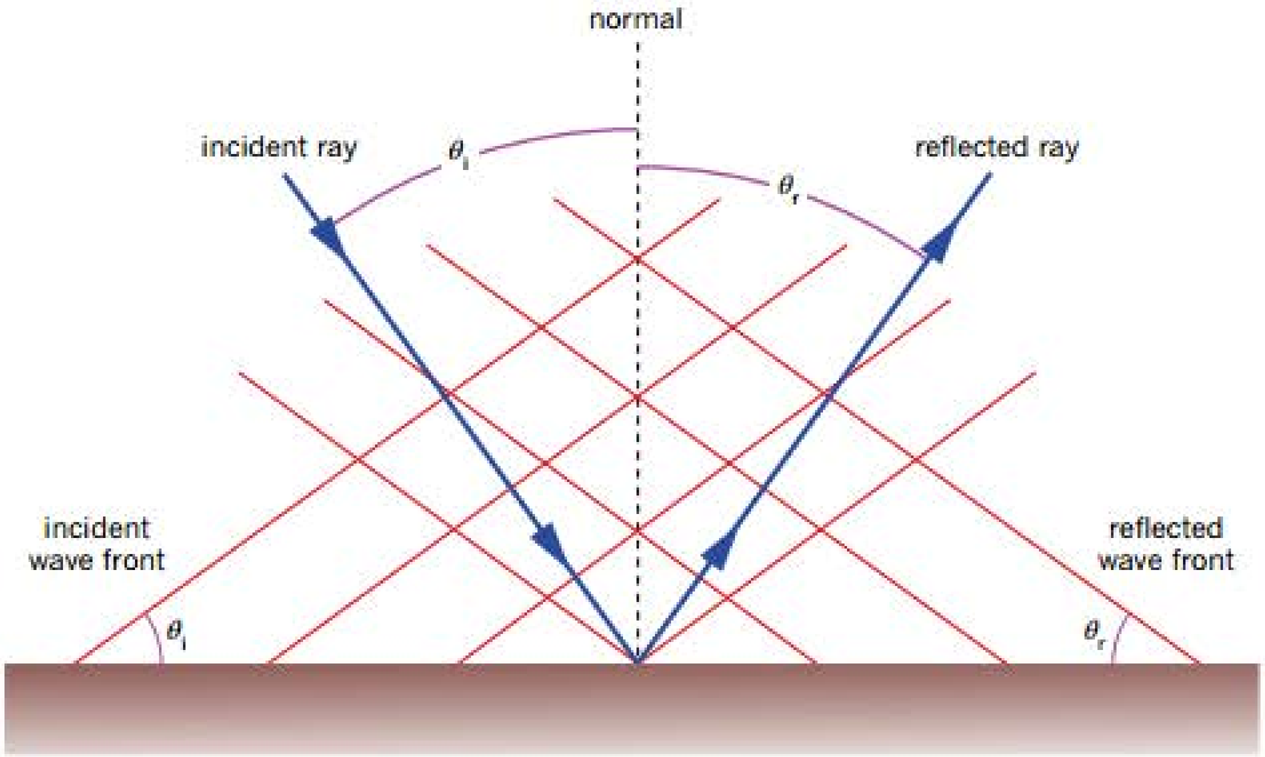


FIGURE 7.1.6 The law of reflection. The angle between the incident ray and the normal (θ_i) is the same as the angle between the normal and the reflected ray (θ_r).

The blue ray illustrates the direction of motion where:

> the angle of incidence, θ_i, is equal to the angle of reflection, θ_r.
>
> $$\theta_i = \theta_r$$

The red lines represent the wavefronts.

REFRACTION

Refraction is a change in the direction of light caused by changes in its speed. Changes in the speed of light occur when light passes from one medium (substance) into another. In Figure 7.1.7, the light changes direction as it leaves the glass prism and re-enters the air.

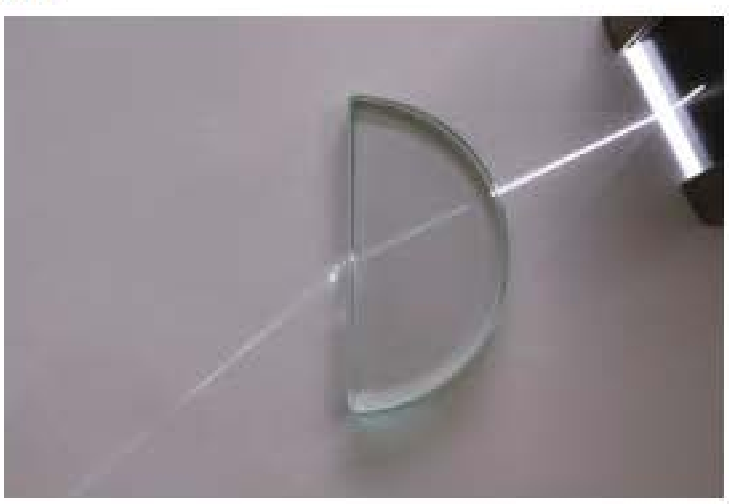


FIGURE 7.1.7 Light refracts as it moves from one medium (i.e. the semicircular glass prism) into another (i.e. air), causing a change in direction.

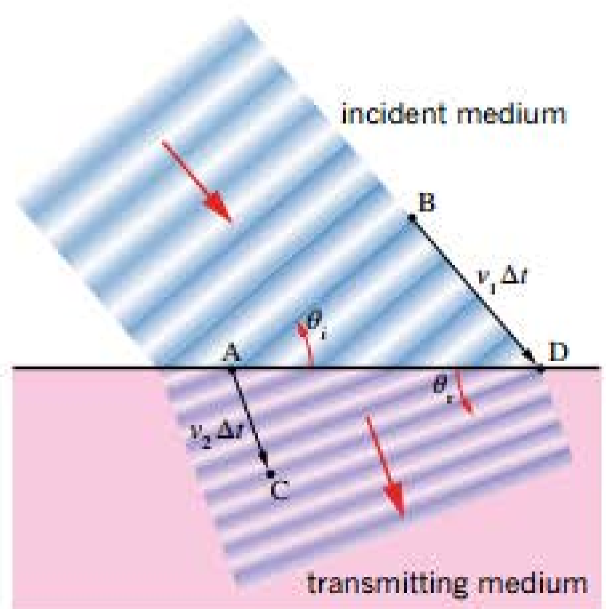


FIGURE 7.1.8 The speed of travel in the transmitting medium is slower than in the incident medium. Wave refraction occurs because for the same time interval the distance A–C travelled by the wave in the transmitting medium is shorter than the distance B–D that it travels in the incident medium.

Consider Figure 7.1.8, where light waves are moving from an incident medium where they have high speed, v_1, into a transmitting medium where they have a lower speed, v_2. For the same time interval that the wave travels a distance $v_1\Delta t$ (B–D) in the incident medium, it travels a shorter distance $v_2\Delta t$ (A–C) in the transmitting medium. In order to do this, the wavefronts must change direction or 'refract' as shown.

Light waves behave in a similar way when they move from a medium such as air into water. The direction of the refraction depends on whether the waves speed up or slow down when they move into the new medium. In Figure 7.1.9, the light waves slow down as they move from air into glass, so the direction of propagation of the wave is refracted towards the normal. The angle of incidence, i, which is defined as the angle between the direction of propagation and the normal, is greater than the angle of refraction, r.

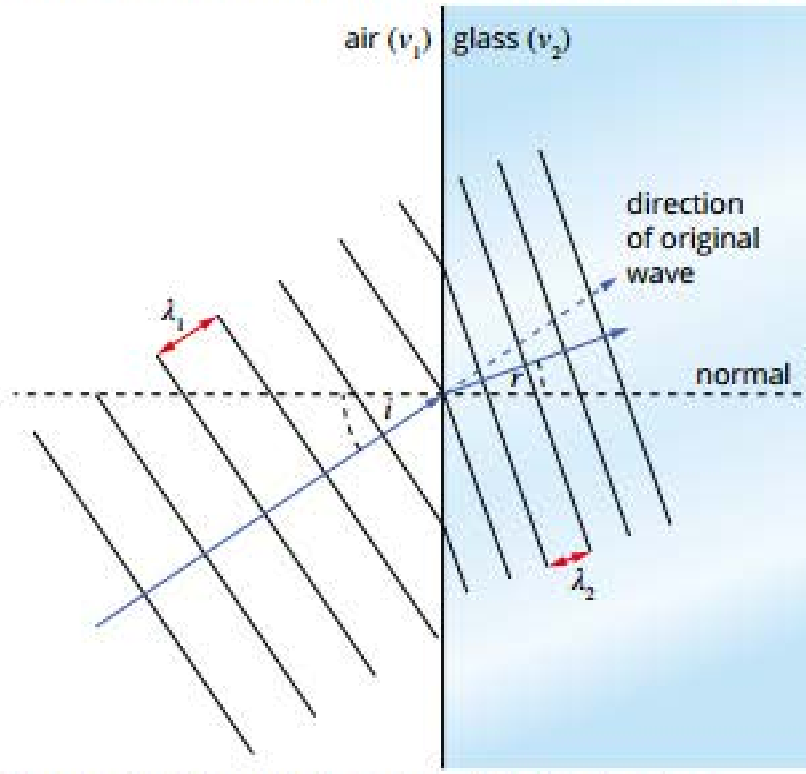


FIGURE 7.1.9 Light waves refract towards the normal when they slow down.

Conversely, when a light wave moves from glass where it has slower speed into air where it travels more quickly, it is refracted away from the normal, as shown in Figure 7.1.10. In other words, the angle of incidence, i, is less than the angle of refraction, r.

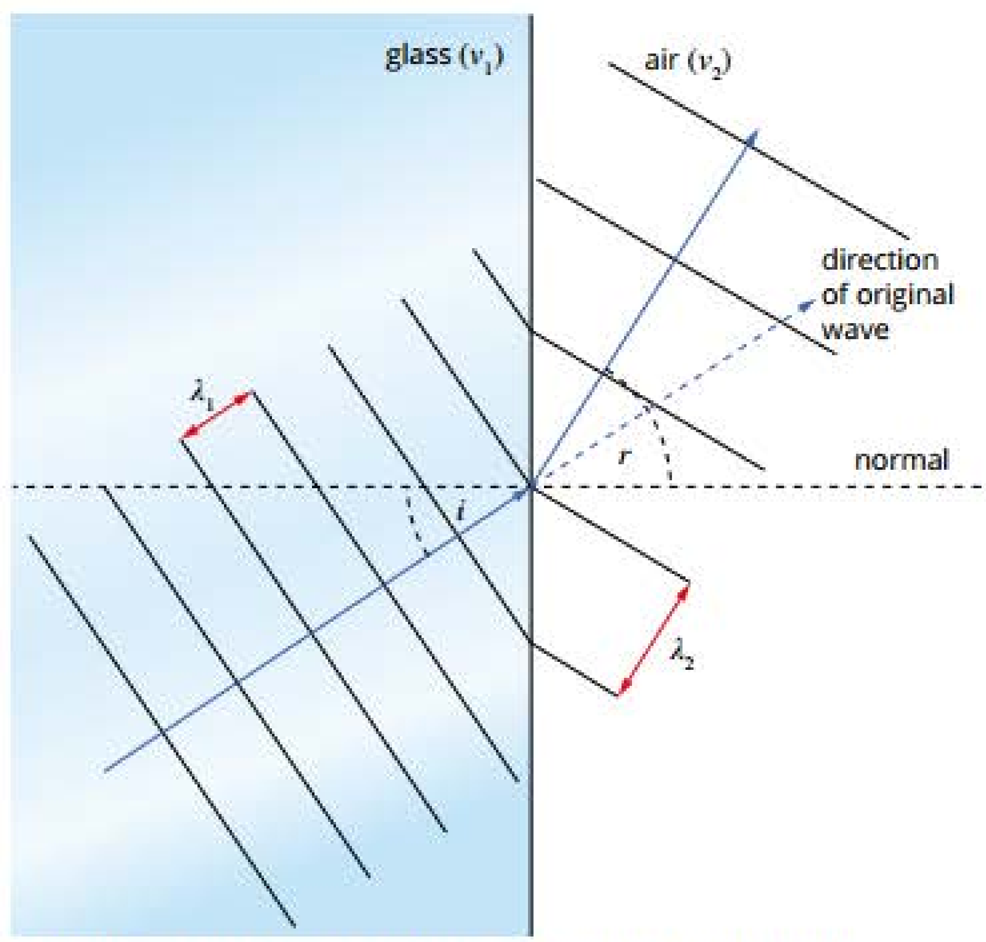


FIGURE 7.1.10 Light waves refract away from the normal when they speed up.

Note that when a wave changes its speed, its wavelength also changes correspondingly, but its frequency does not change as there is still the same number of waves per second; waves cannot be gained or lost.

Refractive index

The amount of refraction that occurs depends on how much the speed of light changes as light moves from one medium to another—when light slows down greatly, it will undergo significant refraction.

The speed of light in a number of different materials is shown in Table 7.1.1.

TABLE 7.1.1 The speed of light in various materials, correct to three significant figures

Material	Speed of light ($\times 10^8 \, m \, s^{-1}$)
vacuum	3.00
air	3.00
ice	2.29
water	2.25
quartz	2.05
crown glass	1.97
flint glass	1.85
diamond	1.24

Scientists find it convenient to describe the change in speed of a wave using a property called the **refractive index**. The refractive index of a material, n, is defined as the ratio of the speed of light in a vacuum, c, to the speed of light in the medium, v:

$$n = \frac{c}{v}$$

Note that n is dimensionless, i.e. it has no units; it is just a ratio.

The refractive index for various materials is given in Table 7.1.2.

TABLE 7.1.2 Refractive indices of various materials

Material	Refractive index, n
vacuum	1.00
air	1.00
ice	1.31
water	1.33
quartz	1.46
crown glass	1.52
flint glass	1.62
diamond	2.42

Worked example 7.1.1

CALCULATING REFRACTIVE INDEX

The speed of light in water is $2.25 \times 10^8 \, m \, s^{-1}$. Given that the speed of light in a vacuum is $3.00 \times 10^8 \, m \, s^{-1}$, calculate the refractive index of water.

Thinking	Working
Recall the definition of refractive index.	$n = \frac{c}{v}$
Substitute the appropriate values into the formula and solve.	$n = \frac{3.00 \times 10^8}{2.25 \times 10^8} = \frac{3.00}{2.25} = 1.33$

Worked example: Try yourself 7.1.1

CALCULATING REFRACTIVE INDEX

The speed of light in crown glass (a type of glass used in optics) is $1.97 \times 10^8\,\mathrm{m\,s^{-1}}$. Given that the speed of light in a vacuum is $3.00 \times 10^8\,\mathrm{m\,s^{-1}}$, calculate the refractive index of crown glass.

By definition, the refractive index of a vacuum is exactly 1, since $n = \frac{c}{c} = 1$. Similarly, the refractive index of air is effectively equal to 1 because the speed of light in air is practically the same as its speed in a vacuum.

The definition of refractive index allows you to determine changes in the speed of light as it moves from one medium to another.

Since $n = \frac{c}{v}$, therefore $c = nv$. This applies for any material, therefore:

$$n_1 v_1 = n_2 v_2$$

where n_1 is the refractive index of the first material
v_1 is the speed of light in the first material
n_2 is the refractive index of the second material
v_2 is the speed of light in the second material

Worked example 7.1.2

SPEED OF LIGHT CHANGES

A ray of light travels from crown glass ($n = 1.52$), where it has a speed of $1.97 \times 10^8\,\mathrm{m\,s^{-1}}$, into water ($n = 1.33$). Calculate the speed of light in water.

Thinking	Working
Recall the formula.	$n_1 v_1 = n_2 v_2$
Substitute the appropriate values into the formula and solve.	$1.52 \times 1.97 \times 10^8 = 1.33 \times v_2$ $\frac{1.52 \times 1.97 \times 10^8}{1.33} = v_2$ $v_2 = 2.25 \times 10^8\,\mathrm{m\,s^{-1}}$

Worked example: Try yourself 7.1.2

SPEED OF LIGHT CHANGES

A ray of light travels from water ($n = 1.33$), where it has a speed of $2.25 \times 10^8\,\mathrm{m\,s^{-1}}$, into glass ($n = 1.85$). Calculate the speed of light in glass.

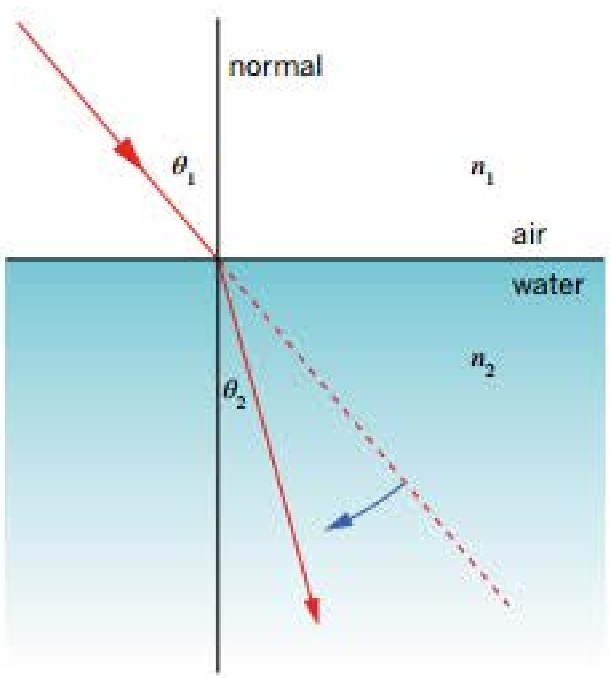


FIGURE 7.1.11 Light refracts as it moves from air into water.

Snell's law

Refractive indices can also be used to determine how much a light ray will refract as it moves from one medium to another. Consider the situation shown in Figure 7.1.11, where light refracts as it moves from air into water.

In 1621, the Dutch mathematician Willebrord Snell described the geometry of this situation with a formula now known as **Snell's law**:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

Worked example 7.1.3

USING SNELL'S LAW

A ray of light in air strikes the surface of a pool of water ($n = 1.33$) at an angle of 30° to the normal. Calculate the angle of refraction of the light in water.	
Thinking	**Working**
Recall Snell's law.	$n_1 \sin\theta_1 = n_2 \sin\theta_2$
Recall the refractive index of air.	$n_1 = 1.00$
Substitute the appropriate values into the formula to find a value for $\sin\theta_2$.	$1.00 \times \sin 30° = 1.33 \times \sin\theta_2$ $\sin\theta_2 = \frac{1.00 \times \sin 30°}{1.33} = 0.3759$
Calculate the angle of refraction.	$\theta_2 = \sin^{-1} 0.3759 = 22.1°$

Worked example: Try yourself 7.1.3

USING SNELL'S LAW

A ray of light in air strikes a piece of flint glass ($n = 1.62$) at an angle of incidence of 50° to the normal. Calculate the angle of refraction of the light in the glass.

These relationships can be used as a guide when you are asked to draw the angle of refraction, when light travels from one medium to another. From the equations $n_1 \sin\theta_1 = n_2 \sin\theta_2$ and $n_1 v_1 = n_2 v_2$, it can be seen that velocity is proportional to the angle.

Therefore, if $v_1 > v_2$ then $\theta_1 > \theta_2$ and $\lambda_1 > \lambda_2$.

Conversely, if $v_1 < v_2$ then $\theta_1 < \theta_2$ and $\lambda_1 < \lambda_2$.

Worked example 7.1.4

APPLYING SNELL'S LAW

Light travels from water into crown glass. Describe the changes in angle and wavelength when the light enters the crown glass. Use Table 7.1.1. Calculations are not necessary.	
Thinking	**Working**
Use Table 7.1.1 on page 237 to find the velocity of light for the two substances.	$v_{water} = v_1 = 2.25 \times 10^8\,m\,s^{-1}$ $v_{crown\ glass} = v_2 = 1.97 \times 10^8\,m\,s^{-1}$
Recall that velocity is proportional to the angle and the wavelength, and compare the two velocities.	$v_1 > v_2$ then $\theta_1 > \theta_2$ $\lambda_1 > \lambda_2$ The light speed is slowing down.
Describe the effect on the refracted wave.	The light is refracted towards the normal and the wavelengths are closer together, as in Figure 7.1.9 on page 236.

Worked example: Try yourself 7.1.4

APPLYING SNELL'S LAW

Light travels from diamond into air. Describe the changes in the angle and the wavelength when the light enters the air. Use Table 7.1.1. Calculations are not necessary.

Total internal reflection

When light passes from a medium with low refractive index (high velocity) to one with higher refractive index (lower velocity), it is refracted *towards* the normal. Conversely, as shown in Figure 7.1.12, when light passes from a medium with a high refractive index to one with a lower refractive index, it is refracted *away from* the normal (Figure 7.1.12(a)). In this case, as the angle of incidence increases, the angle of refraction gets closer to 90° (Figure 7.1.12(b)). Eventually, at an angle of incidence known as the **critical angle**, the angle of refraction becomes 90° and the light is refracted along the interface between the two mediums (Figure 7.1.12(c)). If the angle of incidence is increased above this value, the light ray does not undergo refraction; instead it is reflected back into the original medium, as if it was striking a perfect mirror (Figure 7.1.12(d)). This phenomenon is known as **total internal reflection** and is seen in action in Figure 7.1.13.

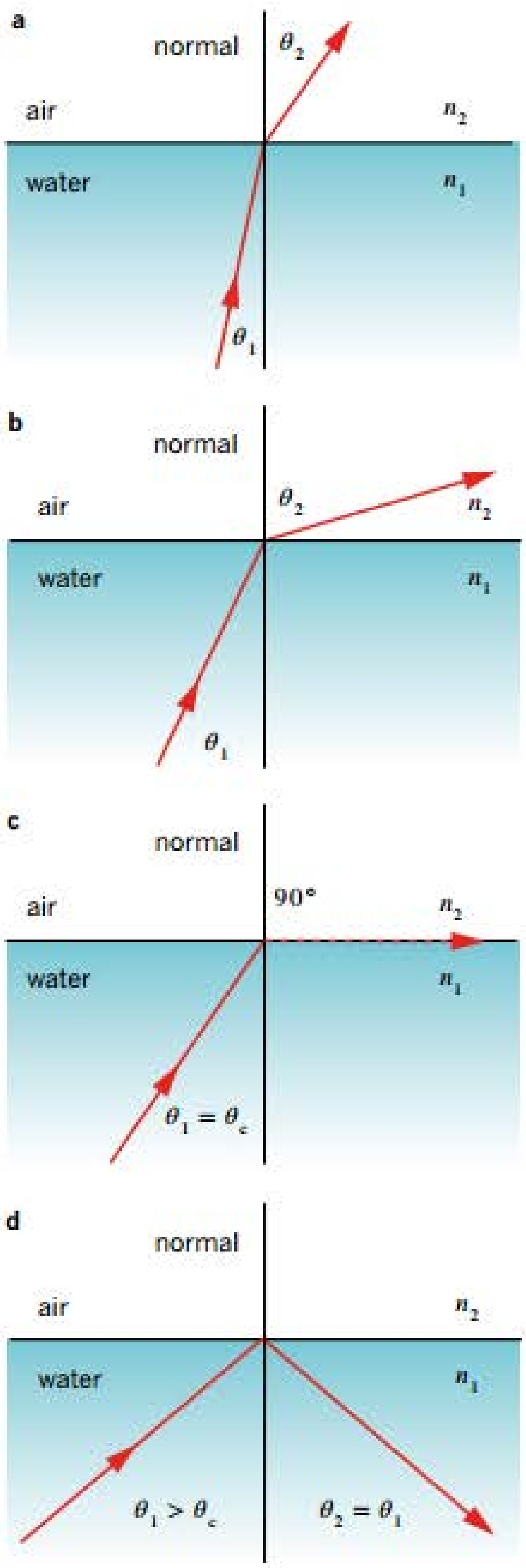


FIGURE 7.1.12 Light refracts as it moves from water into air, as shown in diagrams (a) and (b). In diagram (c) the incident angle is at the critical angle. In diagram (d) the incident angle is greater than the critical angle and the light undergoes total internal reflection.

FIGURE 7.1.13 Optical fibres transmit light using total internal reflection.

As the angle of refraction for the critical angle is 90°, the critical angle is defined by the formula:

$n_1 \sin \theta_c = n_2 \sin 90°$

As $\sin 90° = 1$, then $n_1 \sin \theta_c = n_2$, or $\sin\theta_c = \frac{n_2}{n_1}$.

Worked example 7.1.5

CALCULATING CRITICAL ANGLE

Calculate the critical angle for light passing from water into air.	
Thinking	**Working**
Recall the equation for the critical angle.	$\sin\theta_c = \frac{n_2}{n_1}$
Substitute the refractive indices of water and air into the formula. (Unless otherwise stated, assume that the second medium is air with $n_2 = 1$.)	$\sin\theta_c = \frac{1.00}{1.33} = 0.7519$
Solve for θ_c.	$\theta_c = \sin^{-1} 0.7519 = 48.8°$

Worked example: Try yourself 7.1.5

CALCULATING CRITICAL ANGLE

Calculate the critical angle for light passing from diamond into air.

DISPERSION

When white light passes through a triangular glass prism (as shown in Figure 7.1.15) it undergoes **dispersion**. This is a result of refraction.

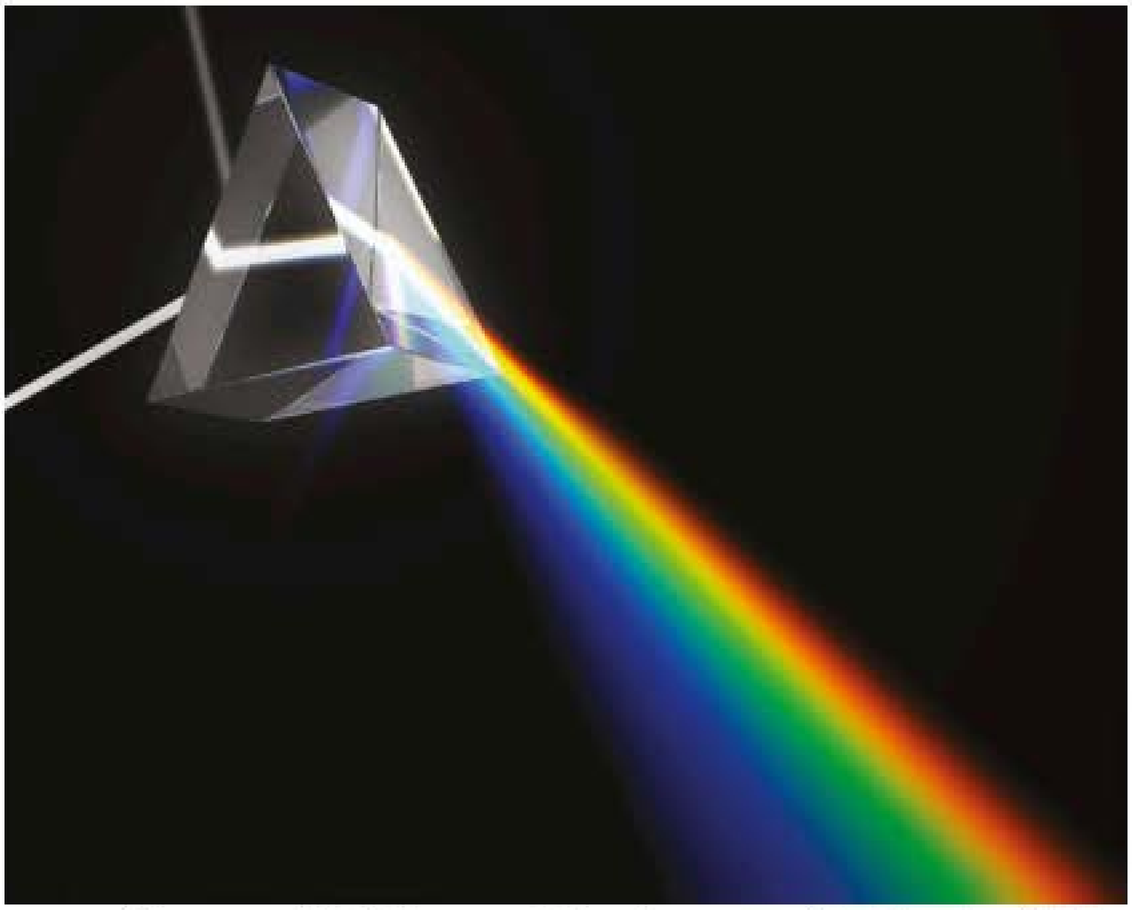


FIGURE 7.1.15 The wave model of light can explain the phenomenon of the dispersion of light, the splitting of white light into its component colours.

According to the wave model of light, each different colour represents a wave of a different wavelength (see Table 7.1.3). White light is a mixture of light waves with many different wavelengths.

TABLE 7.1.3 Approximate wavelength ranges for the colours in the visible spectrum; 1 nm = 10^{-9} m

Colour	Wavelength (nm)
red	780–622
orange	622–597
yellow	597–577
green	577–492
blue	492–455
violet	455–390

When white light passes from one material to another and the light waves slow down, the wavelength shortens as the waves bunch up and the wavelengths of each colour change by different amounts. This means that each colour travels at a slightly different speed in the new medium and therefore each colour is refracted by a slightly different amount.

Longer wavelengths, such as those in red light, travel the fastest in the new material so they are refracted the least. Shorter wavelengths, such as those in violet light, are slower so they are refracted the most.

So, in effect, each colour of light has a different refractive index in a material.

PHYSICSFILE

Refractive index of diamonds

Since diamond has a very high refractive index, it has a small critical angle. This means that a light ray that enters a diamond will often bounce around inside the diamond many times before leaving the diamond. A jeweller can cut a diamond to take advantage of this property; this causes the diamond to 'sparkle' (Figure 7.1.14) as it appears to reflect more light than is falling on it.

FIGURE 7.1.14 The refractive properties of diamonds mean they appear to sparkle.

Colour dispersion in lenses

As each colour of light effectively has a different refractive index in glass, light passing through a glass lens always undergoes some dispersion. This means that coloured images formed by optical instruments such as microscopes and telescopes can suffer from a type of distortion known as *chromatic aberration* (Figure 7.1.17).

FIGURE 7.1.17 Chromatic aberration causes the coloured fringes that can be seen in the circled regions in this image.

Scientists have developed a number of techniques to deal with this problem, including:

- using lenses with very long focal lengths
- using 'achromatic' lenses. These are compound lenses that are made of different types of glass with different refractive properties
- taking separate images using coloured filters and then combining these images to form a single multi-coloured image.

PHYSICSFILE

Where does colour come from?

In the 17th century, many people believed that white light was 'stained' by its interaction with earthly materials. Newton very neatly disproved this with a simple experiment using two prisms (Figure 7.1.16)—one to split light into its component colours and the other to turn it back into white light. This showed that the various colours were intrinsic components of white light since, if colour was a result of 'staining', the second prism should have added more colour rather than producing white light.

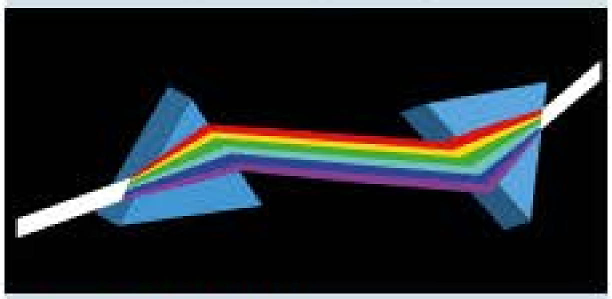


FIGURE 7.1.16 Newton's double prism experiment showed that white light is made up of its component colours.

Newton was the first to identify the colours of the spectrum—red, orange, yellow, green, blue, indigo and violet. He chose seven colours by inventing the colour 'indigo' because seven was considered a sacred number.

DIFFRACTION

When a plane (straight) wave passes through a narrow opening, it bends. Waves will also bend as they travel around obstacles (Figure 7.1.18). This kind of 'bending' phenomenon is known as **diffraction**.

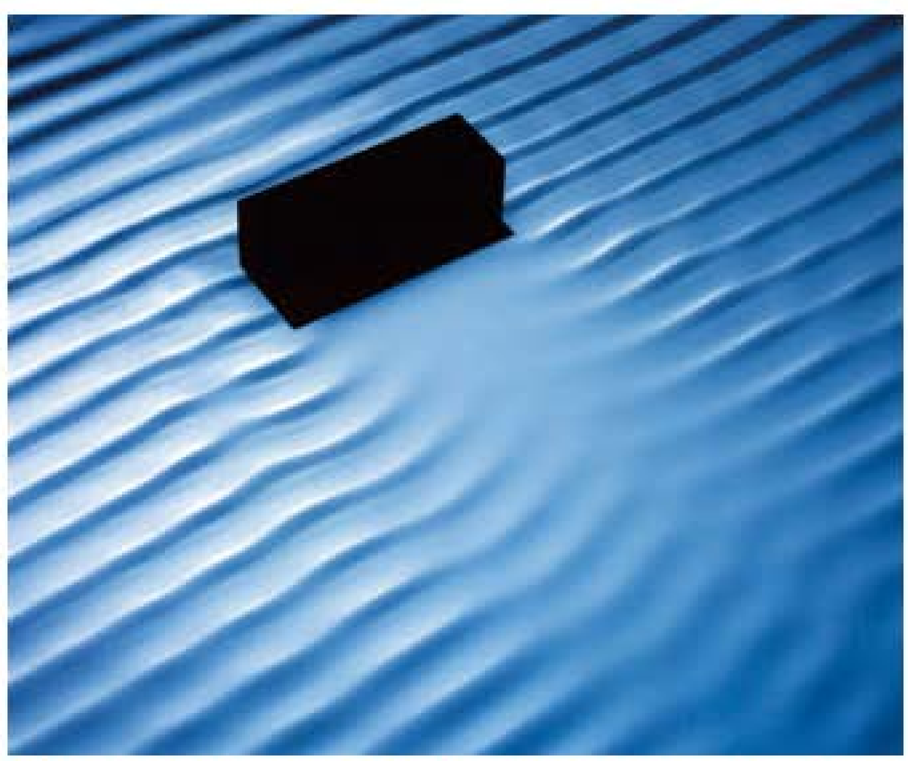


FIGURE 7.1.18 Water waves will bend around an obstacle. Sound waves diffract as well, allowing you to hear around corners.

Diffraction is significant when the size of the opening or obstacle (w) is similar to or smaller than the wavelength of the wave (λ). Light waves range in wavelength from around 700 nm for red light to about 400 nm for violet light. Recall that 1 nm is equal to 10^{-9} m or a one-millionth of a millimetre, therefore the wavelengths of light waves are all less than one-thousandth of a millimetre in length. This means that the diffraction of light is difficult to observe because the wavelength of light is very small. Diffraction can be observed from natural objects such as a human hair or a cotton thread, as shown in Figure 7.1.19.

Usually diffraction occurs with artificially constructed materials such as CDs (Figure 7.1.20) or commercially produced diffraction gratings.

FIGURE 7.1.19 The interference pattern obtained from a cotton thread.

FIGURE 7.1.20 Information on a CD or DVD is stored using a tight spiral of bumps or pits. These small structures are created by a laser and are of the size of a wavelength of light. This means these structures are small enough to cause light to diffract.

Diffraction and slit width

In the diffraction of waves, if the wavelength is much smaller than the gap or obstacle, the degree of diffraction is less. For example, Figure 7.1.21 shows the diffraction of water waves in a ripple tank. In Figure 7.1.21(a), the gap is similar in size to the wavelength, so there is significant diffraction and the waves emerge as circular waves. In Figure 7.1.21(b), the gap is much bigger than the wavelength, so diffraction only occurs at the edges.

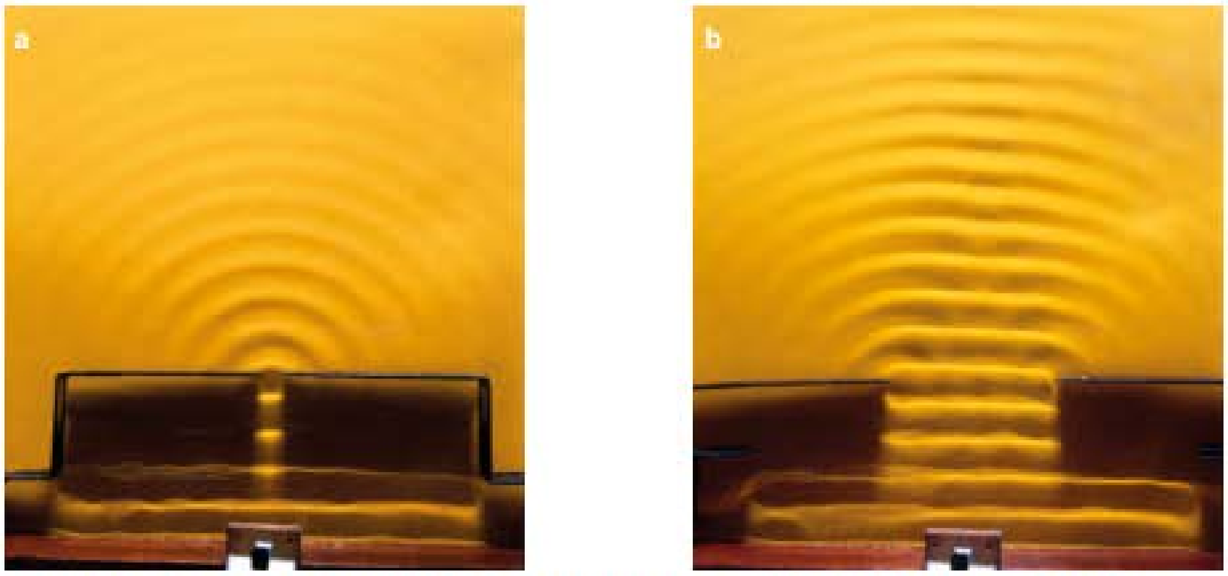


FIGURE 7.1.21 The diffraction of water waves in a ripple tank. (a) Significant diffraction occurs when the wavelength approximates the slit width, i.e. $\lambda \approx w$. (b) As the gap increases, i.e. $\lambda << w$, diffraction becomes less obvious, but is still present.

Wavelengths comparable to or larger than the diameter of the obstacle or gap will produce significant diffraction. This can be expressed as the ratio $\frac{\lambda}{w} \geq 1$, where λ is the wavelength of the wave and w is the width of the gap.

Diffraction and imaging

Diffraction can be a problem for scientists using microscopes and telescopes because it can result in blurred images. For example, a significant problem is that the light from two tiny objects or two distant objects very close together can be diffracted so much that the two objects appear as one blurred object. When this happens, we say that the objects are unresolved. Essentially, the ratio $\frac{\lambda}{w}$ dictates how small an object can be to be clearly imaged by a particular instrument.

This means that, as a general rule, optical microscopes cannot create images of objects that are smaller than the wavelength of the light they use; otherwise, diffraction effects are too significant.

Diffraction also places a theoretical limit on the resolution of optical telescopes. However, atmospheric distortion usually has a much larger effect on telescope images than diffraction. The Hubble Space Telescope, which sits above Earth's atmosphere, is not affected by atmospheric distortion. It can resolve images right down to its diffraction limit, i.e. where the apparent separation of the stars is approximately equal to the wavelength of the light.

Diffraction gratings

As you have already seen, light diffracts as it passes through a very small gap. As the light passes through the gap, some of the wavelets making up the wavefront will diffract at the barriers that form the edges of the gap and some will pass through the centre of the gap. As a result of this the light waves that emerge from the gap will interact. In some places the interactions will be constructive and in others the interactions will be destructive. When these light waves are made to shine on a screen, the areas of **constructive interference** will appear as bright bands and areas of **destructive interference** will appear as dark bands. The pattern of dark and light bands that is seen when light passes through a single small gap is called a **diffraction pattern**.

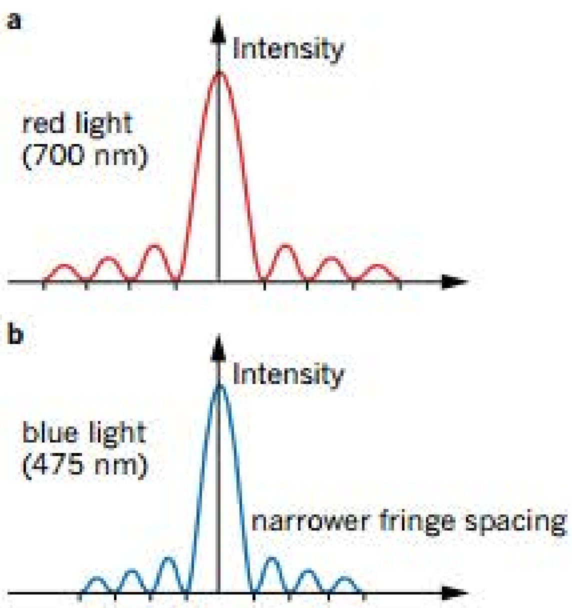


FIGURE 7.1.22 Red light (a) is diffracted to a greater extent than blue light (b). The longer wavelength of red light results in more-widely spaced fringes and a wider overall pattern.

As stated earlier, the extent of diffraction of light waves is proportional to the ratio $\frac{\lambda}{w}$. This ratio also describes the spacing of dark and light bands in a diffraction pattern, and therefore the width of the overall diffraction pattern. According to this relationship, if the wavelength is held constant and the gap made smaller, greater diffraction is seen. If different wavelengths enter the same gap, those with a smaller wavelength will undergo less diffraction than those with a longer wavelength. This is shown in Figure 7.1.22. Note that Figure 7.1.22 shows intensity. High intensity is where bright bands will appear on a screen; zero intensity corresponds to dark bands.

Some diffraction patterns can be observed using natural materials but, in practice, much clearer diffraction patterns can be generated by passing light through a *diffraction grating*. A diffraction grating is a piece of material that contains a large number of very closely spaced parallel gaps or slits. The diffraction pattern from one slit is superimposed on the pattern from the adjacent slit, producing a strong, clear image on the screen.

Diffraction experiments usually use **monochromatic** light (i.e. light of only one colour). When white light, which contains a number of different colours, shines through a diffraction grating, each different colour is diffracted by a different amount and forms its own set of coloured fringes. This results in the light being dispersed into its component colours, as seen in Figure 7.1.23.

FIGURE 7.1.23 A diffraction grating disperses white light into a series of coloured spectra.

This property is used in scientific spectrometers; for example, an emission spectrum shone onto a grating produces a diffraction image and a slit is used to select the wavelength of interest. By rotating the grating, the emission spectrum can be measured as a function of wavelength.

POLARISATION

One of the most convincing pieces of evidence for the wave nature of light is the phenomenon of **polarisation**. Polarisation occurs when a transverse wave is allowed to vibrate in only one direction. For example, the light wave in Figure 7.1.24 is vertically polarised—the wave oscillations occur in the vertical plane only. This also means this wave is unaffected by a polarising filter that is oriented in the vertical plane.

The wave in the Figure 7.1.25 is horizontally polarised. It is completely blocked by the vertical polarising filter.

In Figure 7.1.26, the incoming wave is polarised at 45° to the horizontal and vertical planes. The horizontal component of this wave is blocked by the vertical filter, so the ongoing wave is vertically polarised and has a smaller amplitude than the original wave.

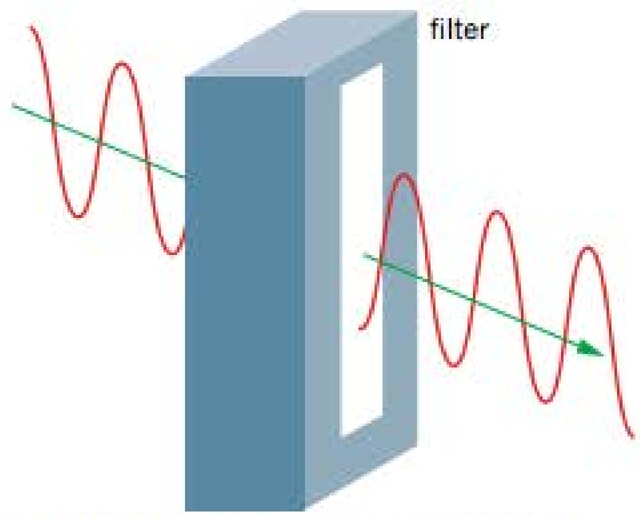


FIGURE 7.1.24 A vertically polarised wave can pass through a vertically oriented polarising filter.

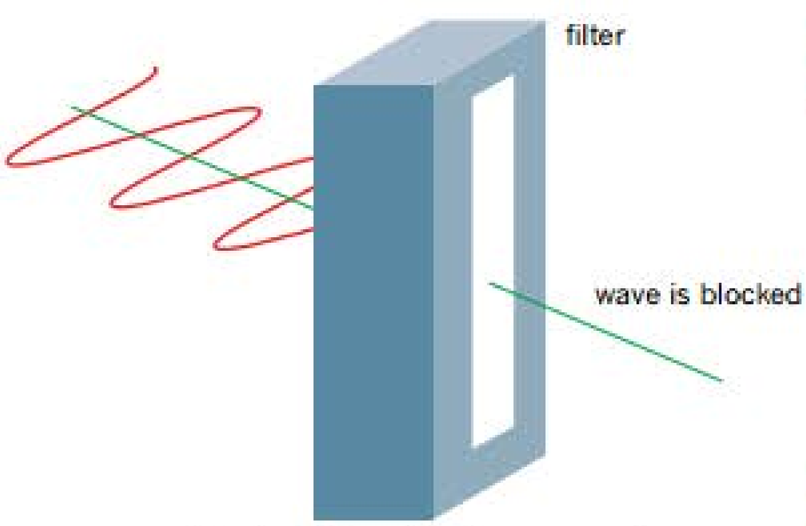


FIGURE 7.1.25 A horizontally polarised wave cannot pass through a vertically oriented polarising filter.

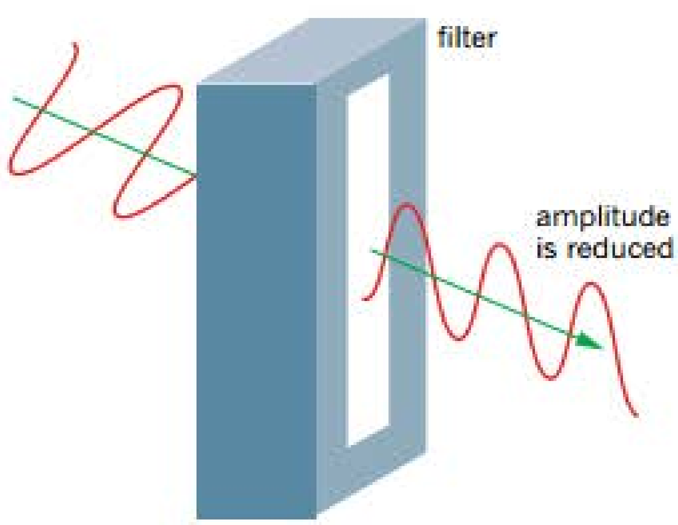


FIGURE 7.1.26 A diagonally polarised wave has its horizontal component suppressed by the vertically oriented polarising filter. A vertically polarised wave of reduced amplitude passes through it.

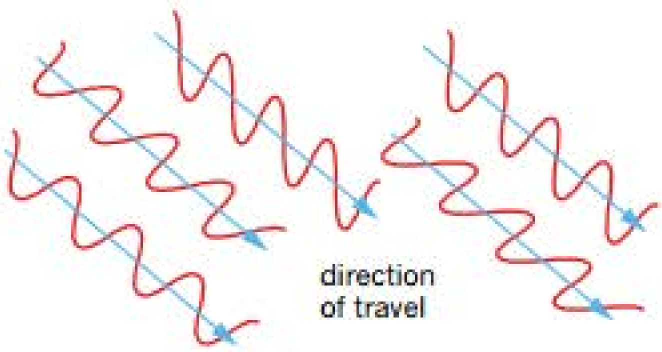


FIGURE 7.1.27 Unpolarised light consists of a collection of waves that are each polarised in a different direction.

Light produced by sources such as a light globe or the Sun is unpolarised, which means that it can be thought of as a collection of waves, each with a different plane of polarisation, as shown in Figure 7.1.27.

Certain materials can act as polarising filters for light. These only transmit the waves or components of waves that are polarised in a particular direction and absorb the rest. Polarising sunglasses work by absorbing the light polarised in a particular direction, thus reducing glare. Photographers also use polarised filters to reduce the glare in photographs or achieve specific effects (Figure 7.1.28).

FIGURE 7.1.28 These are photographs taken of the same tree, one without a polarising filter (left) and one with a polarising filter (right).

PHYSICSFILE

Polarised sunglasses

Light that is reflected from the surface of water or snow is partially polarised (Figure 7.1.29). The polarising plane of polarised sunglasses is selected to absorb this reflected light. This makes polarised sunglasses particularly effective for people involved in outdoor activities such as boating, fishing or skiing.

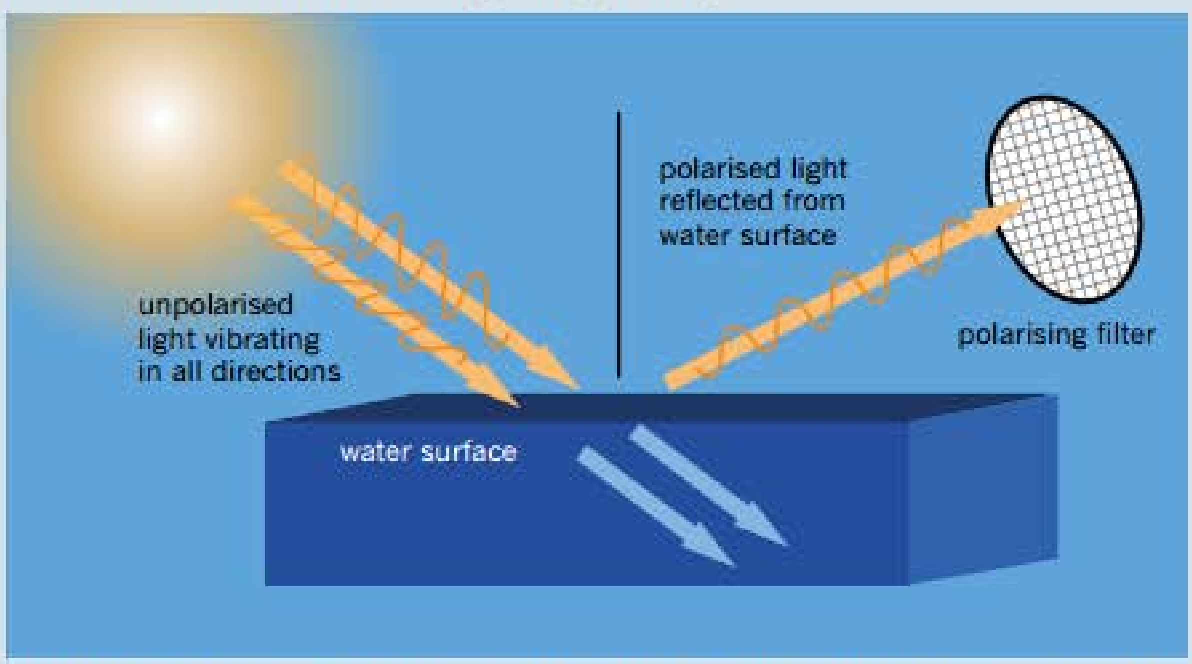


FIGURE 7.1.29 The polarising filter in a pair of sunglasses is designed to block the polarised light reflected from the surface of the water and transmit the unpolarised light from under the water.

7.1 Review

SUMMARY

- A wave model explains a wide range of light-related phenomena, including reflection, refraction, dispersion, diffraction and interference.
- For reflection, the angle of reflection = angle of refraction.
- Refraction is the change in the direction of light that occurs when light moves from one medium to another.
- Refraction is caused by changes in the speed of light waves.
- The refractive index, n, of a material is given by the formula $n = \frac{c}{v}$, where c is the speed of light in a vacuum and v is the speed of light in the material.
- When light moves from one material to another, the changes in speed can be calculated using $n_1 v_1 = n_2 v_2$.
- The amount of refraction of a ray of light can be calculated using Snell's law: $n_1 \sin\theta_1 = n_2 \sin\theta_2$.
- If $v_1 > v_2$ then $\theta_1 > \theta_2$, and in addition $\lambda_1 > \lambda_2$, and conversely if $v_1 < v_2$ then $\theta_1 < \theta_2$ and $\lambda_1 < \lambda_2$.
- The critical angle is the angle the incident wave makes when the refracted angle equals 90°. It can be calculated using $n_1 \sin\theta_c = n_2 \sin 90°$ or $\sin\theta_c = \frac{n_2}{n_1}$
- Total internal reflection occurs when the incident angle exceeds the critical angle and the angle of refraction exceeds a right angle.
- Different colours of light have different wavelengths.
- Dispersion occurs because different colours of light travel at different speeds and hence refract at slightly different angles.
- When a plane (straight) wave passes through a narrow opening or meets a sharp object, it experiences diffraction.
- Significant diffraction occurs when the wavelength of the wave is similar to, or larger than, the size of the diffracting object.
- A transverse wave model is required to explain polarisation.

KEY QUESTIONS

1 Name the model of light supported by each of the following scientists.
  a Hooke
  b Huygens
  c Newton

2 Why did most 18th century scientists support Newton's particle model of light?
  A Newton had better evidence to support his theory.
  B The speed of light in glass had been shown to be faster than in air.
  C Newton had a better reputation as a scientist than either Hooke or Huygens.
  D Newton was English, and Hooke and Huygens were from other parts of Europe.

3 Choose the correct response from those given in bold to complete the sentences about the refractive indices of types of water.
  Although pure water has a refractive index of 1.33, the salt content of seawater means its refractive index is a little higher at 1.38. Therefore, the speed of light in seawater will be **greater than/less than/the same as** in pure water.

4 Calculate the speed of light in seawater, which has a refractive index of 1.38.

5 Light travels at $2.25 \times 10^8\,m\,s^{-1}$ in water and $2.29 \times 10^8\,m\,s^{-1}$ in ice. If water has a refractive index of 1.33, use this information to calculate the refractive index of ice.

6 Light travels from water ($n = 1.33$) into glass ($n = 1.60$). The incident angle is 44°. Calculate the angle of refraction.

7 For which of the following situations can total internal reflection occur?

	Incident medium	Refracting medium
A	air ($n = 1.00$)	glass ($n = 1.55$)
B	glass ($n = 1.55$)	air ($n = 1.00$)
C	glass ($n = 1.55$)	water ($n = 1.33$)
D	glass ($n = 1.55$)	glass ($n = 1.58$)

8 In order to produce significant diffraction of red light (wavelength of approximately 700 nm), a diffraction experiment would need to use an opening with a width of approximately:
  A 0.7 mm
  B 0.07 mm
  C 0.007 mm
  D 0.0007 mm

9 Explain how polarisation supports a transverse wave model for light.

7.2 Interference: Further evidence for the wave model of light

Thomas Young's observation of the interference patterns of light (Figure 7.2.1) was a pivotal moment in the history of science. It tipped the scales in a long-running dispute between scientists about the nature of light, and paved the way for a series of discoveries and inventions that would fundamentally change scientists' understanding of energy and matter.

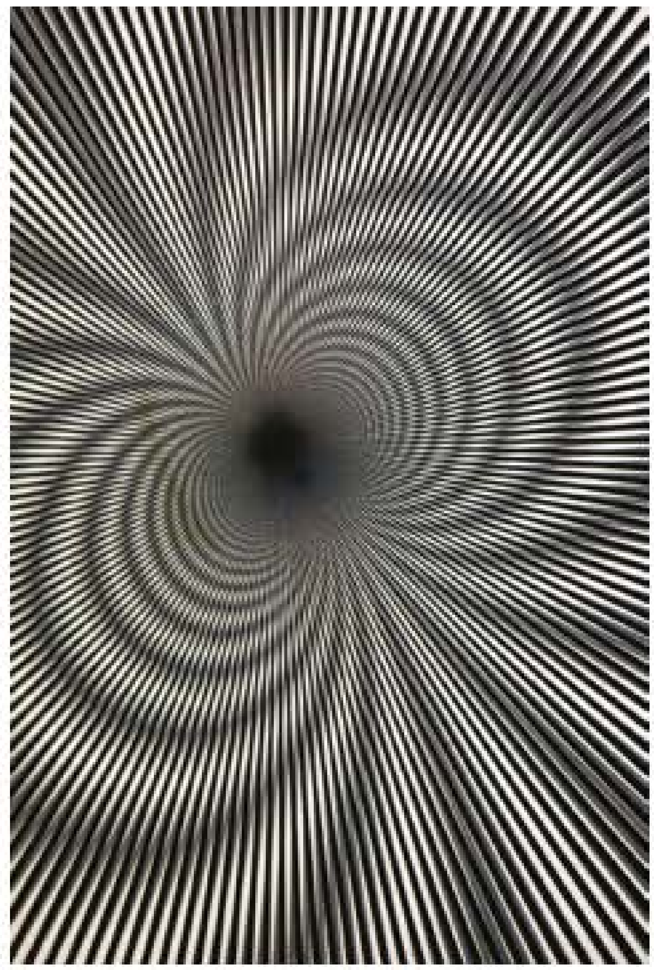


FIGURE 7.2.1 Optical interference can produce spectacular patterns.

YOUNG'S DOUBLE-SLIT EXPERIMENT

Between the 17th and 19th centuries, most scientists considered light to be a stream of particles. This idea was based on the 'corpuscular' theory proposed by Sir Isaac Newton.

In 1801, an English scientist called Thomas Young performed a now-famous experiment in which he shone monochromatic light on a screen containing two very tiny slits. On the far side of the double slits he placed another screen, on which he observed the pattern produced by the light passing through the slits (Figure 7.2.2).

According to the particle theory, light should have passed directly through the slits to produce two bright lines or bands on the screen (Figure 7.2.2(a)). Instead, Young observed a series of bright and dark bands or 'fringes' (Figure 7.2.2(b)).

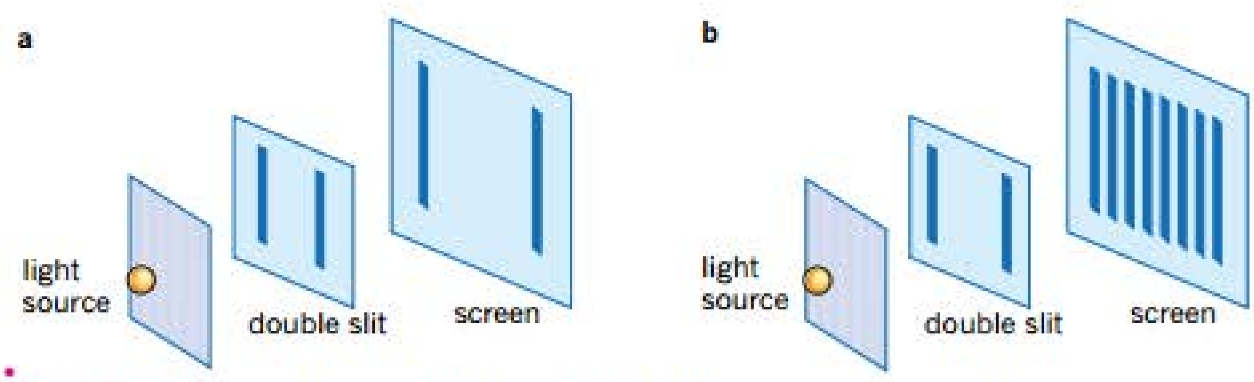


FIGURE 7.2.2 The particle theory of light predicted that Young's experiment should produce two bright bands (a). But his actual experiment (b) produced a series of bright and dark bands or 'fringes'.

Young was able to explain this bright and dark pattern by treating light as a wave. He assumed that the monochromatic light was like plane waves and that as they passed through the narrow slits these plane waves were diffracted into **coherent** (in phase) circular waves, as shown in Figure 7.2.3. The circular waves would interact causing **interference**. The interference pattern produced by these two waves would result in lines of constructive (antinodal) and destructive (nodal) interference that would match the bright and dark fringes respectively.

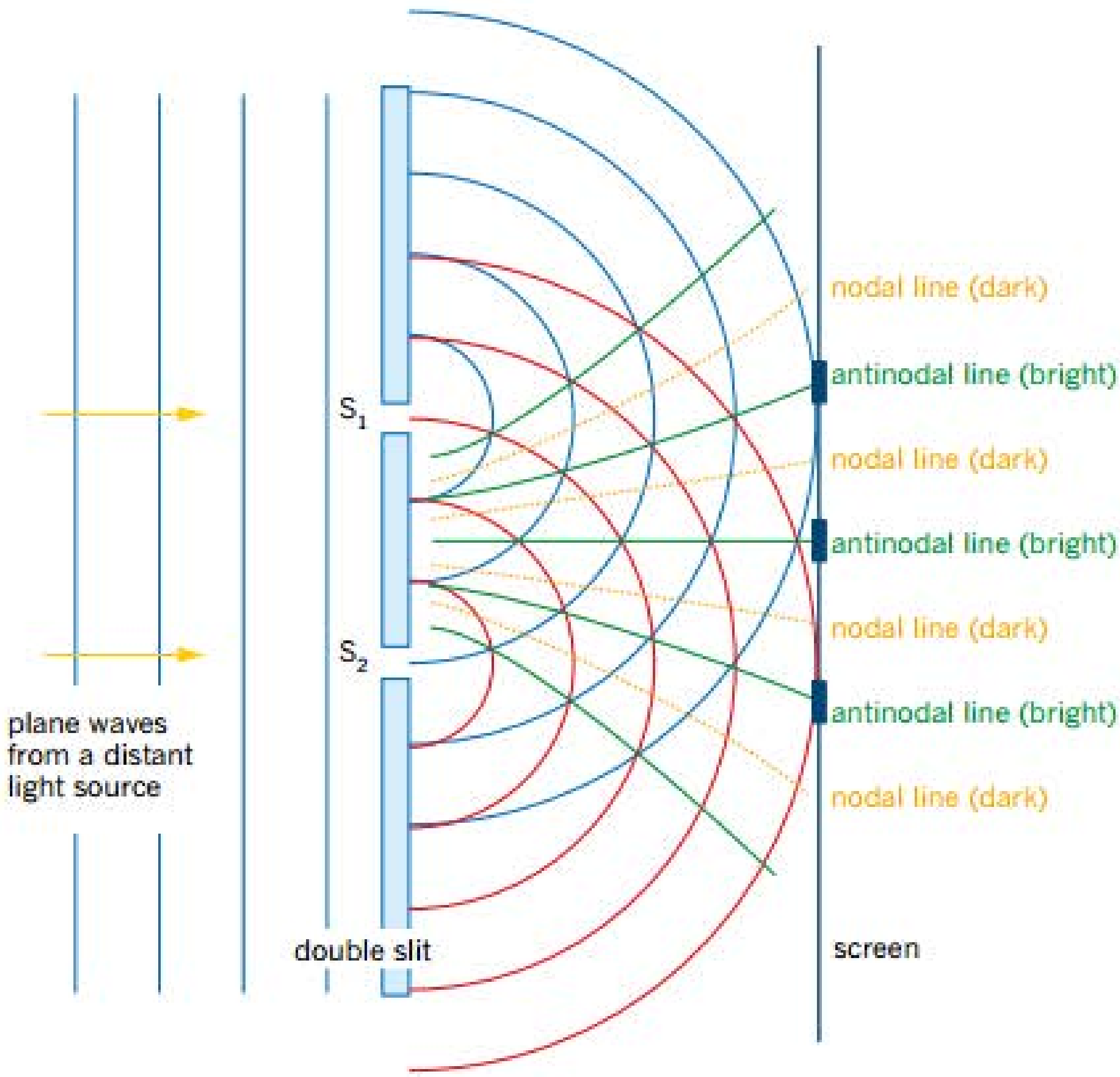


FIGURE 7.2.3 The interaction of two circular waves can produce a pattern of antinodal (constructive interference) and nodal (destructive interference) lines.

Earlier in his scientific career, Young had observed similar interference patterns in water waves (Figure 7.2.4). This gave greater credibility to the wave model for light proposed by Christiaan Huygens and Robert Hooke many years earlier.

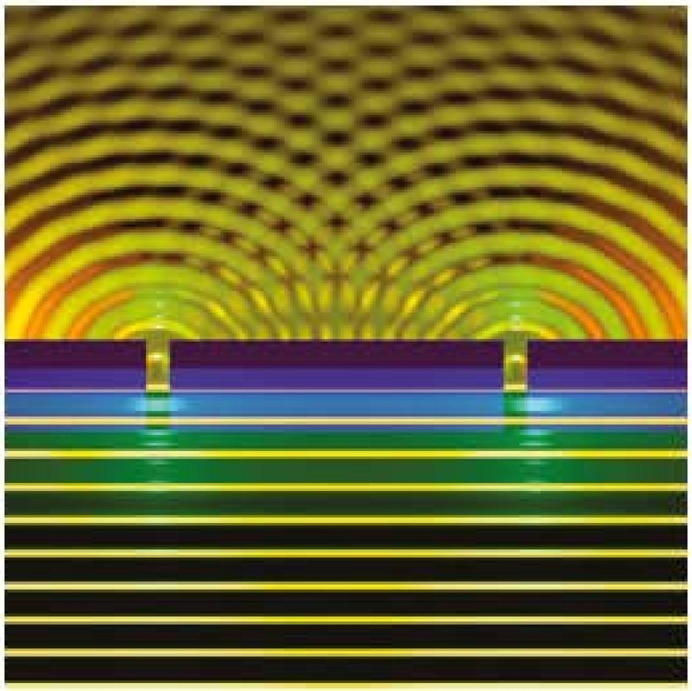


FIGURE 7.2.4 Interference patterns can be observed in water waves (lit here in yellow).

When Young used his data to calculate the wavelength of light, it became clear why no one had ever noticed the wave properties of light before—light waves are tiny, with typical wavelengths of less than 1 micrometre (1 µm = 0.001 mm).

Path difference

To understand Young's experiment more fully, you have to consider how the waves produced by the two slits interact with each other when they hit the screen. Refer to Figure 7.2.5. At a particular point, P, on the screen, the distance travelled by the wave train from slit 1 (S_1) will be different from the distance travelled by the wave train from slit 2 (S_2), i.e. the distance S_1P is different from S_2P. The difference in the distance travelled by each wave train to a point P on the screen is called the **path difference** for the waves (pd).

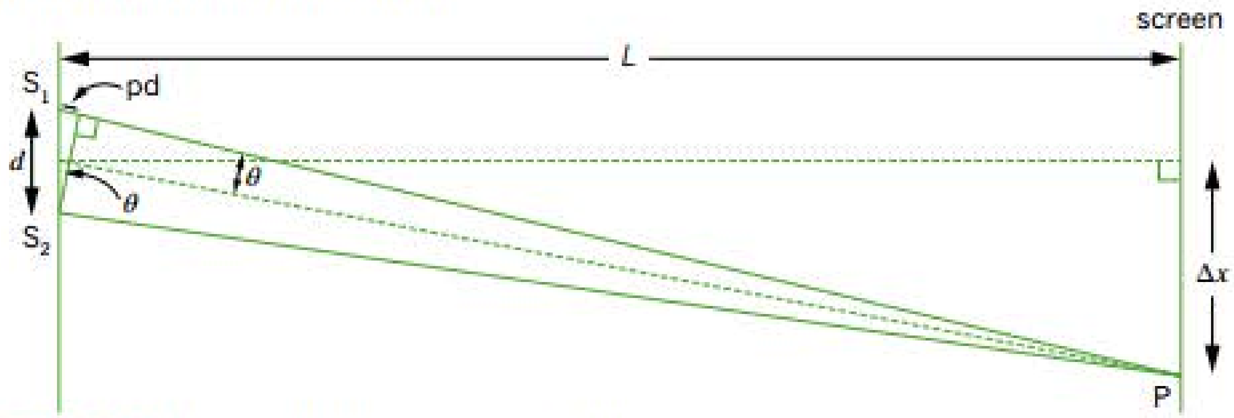


FIGURE 7.2.5 The geometry of two-point source interference.

> The path difference to point P from wave source S_1 and from wave source S_2 is given by:
>
> $pd = |S_1P - S_2P|$

Path difference can be measured in metres, but it is far more useful to measure it in wavelengths in order to determine the light intensity on the screen.

As shown in Figure 7.2.6, at a point, M, at the centre of the screen, equidistant from each slit, each wave train will have travelled through the same distance and so there is no path difference (i.e. $S_1M = S_2M$). The light waves arrive in phase with each other. These light waves reinforce to produce an antinode. A fringe of bright light is seen, known as the 'central maximum'. Recall that phenomenon is called constructive interference.

Constructive interference will occur whenever the path difference between the two wave trains is zero or differs by a whole number of wavelengths, i.e. pd = 0λ, 2λ, 3λ … For example, in Figure 7.2.6, the path difference $S_2R - S_1R$ is equal to λ.

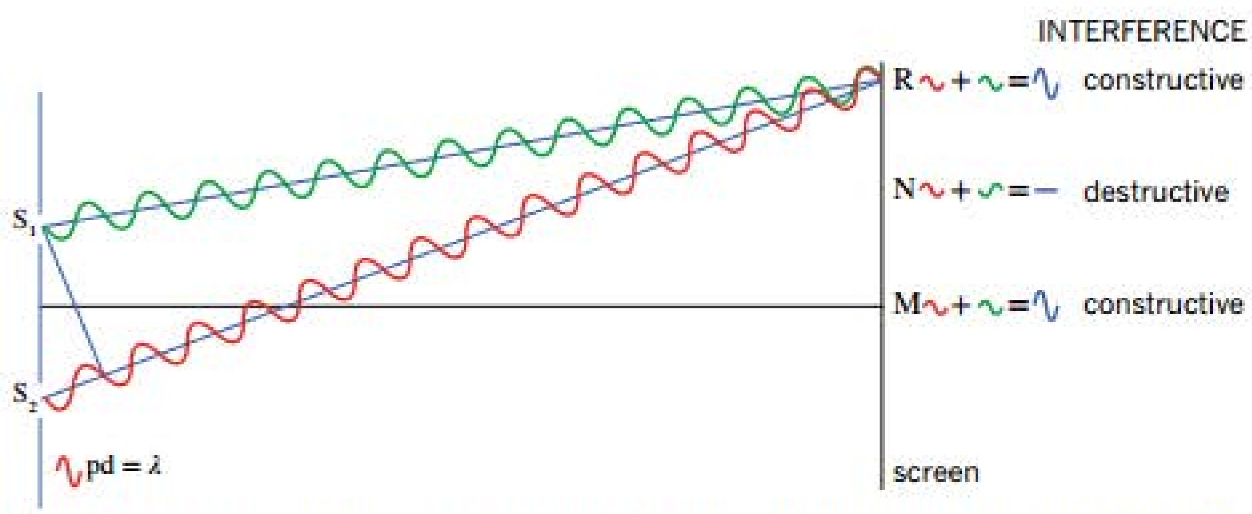


FIGURE 7.2.6 Waves meeting from each slit at R, where the path difference is λ. A bright fringe will be seen as the wave trains arrive at this point in phase again.

There will be points on the screen at which the path difference is $\frac{\lambda}{2}$; for example, at point N in Figure 7.2.6. The two wave trains that meet at this point are completely out of phase and cancel each other to produce a nodal point. Destructive interference occurs at this point, and no light is seen. This creates the dark lines or fringes that appear in between the bright antinodal fringes. Destructive interference occurs when the path difference between the waves is $\frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \frac{7\lambda}{2}$, etc.

In summary:

- constructive interference of coherent waves occurs when the path difference pd = $n\lambda$, where $n = 0, 1, 2, 3\ldots$
- destructive interference of coherent waves occurs when the path difference equals an odd number of half wavelengths; that is, pd = $\left(n - \frac{1}{2}\right)\lambda$, where $n = 1, 2, 3\ldots$

The sequence of constructive and destructive interference effects produces an interference pattern of regularly spaced vertical bands or fringes on the screen that can be represented graphically as shown in Figure 7.2.7.

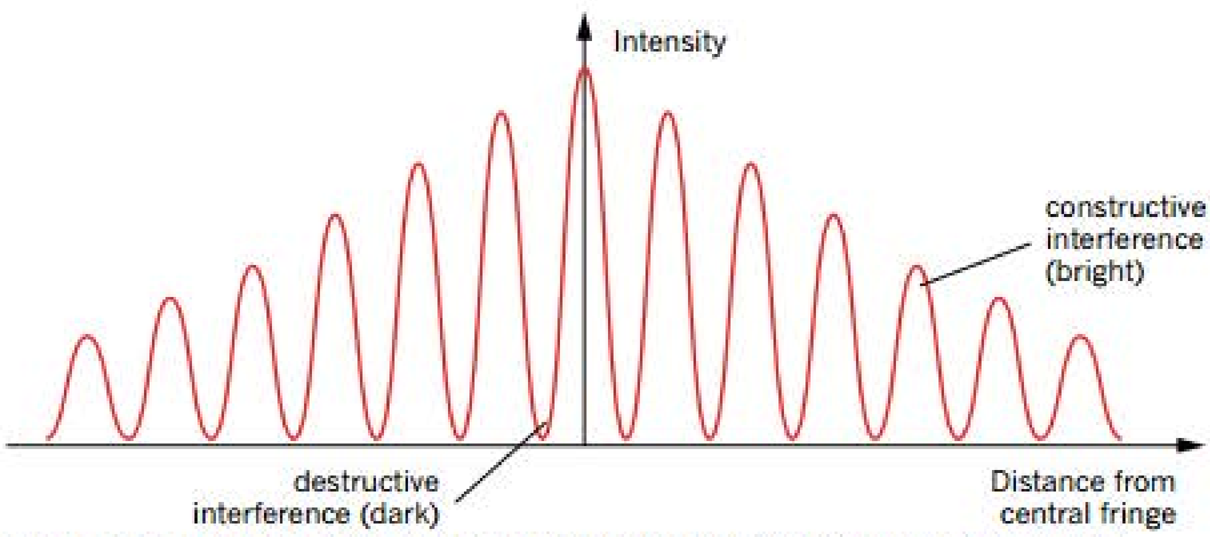


FIGURE 7.2.7 The double-slit interference pattern can be considered in terms of an intensity distribution graph. The horizontal axis represents a line drawn across the screen. The centre of the distribution pattern corresponds to the centre of the brightest central fringe, the central maximum.

Calculating fringe separation for Young's experiment

In Young's experiment, the distance between adjacent bright bands on the screen is known as the fringe spacing (Δx). This distance depends on the wavelength of light (λ), the separation between the two slits (d) and the distance to the screen (L).

- If the viewing screen is moved further from the two slits, the fringes will appear further apart from each other, i.e. $\Delta x \propto L$.
- Conversely, reducing the separation of the slits increases the spacing of the fringes, i.e. $\Delta x \propto \frac{1}{d}$.
- Using light of a longer wavelength will also result in increased fringe spacing, i.e. $\Delta x \propto \lambda$ (Figure 7.2.8).

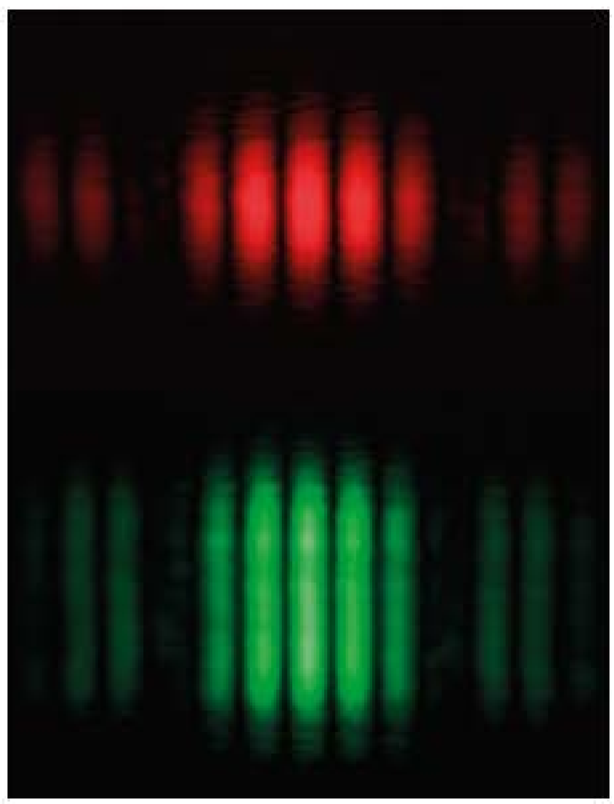


FIGURE 7.2.8 If the separation of the slits and the distance to the screen are kept the same, then the fringes produced by longer wavelength red light are further apart than for shorter wavelength green light.

Fringe separation parameters

These relationships can be combined to develop an overall equation for the fringe separation:

$$\Delta x = \frac{\lambda L}{d}$$

where Δx is the fringe separation

λ is the wavelength of the light waves

L is the distance from the slits to the screen

d is the slit separation

Worked example 7.2.1

CALCULATING WAVELENGTH FROM FRINGE SEPARATION

Light of an unknown wavelength emitted by a laser is directed through a pair of thin slits 50 μm apart. The slits are 2.0 m from a screen on which bright fringes are 2.5 cm apart. Calculate the wavelength of the laser light in nm.	
Thinking	**Working**
Recall the equation for fringe separation.	$\Delta x = \frac{\lambda L}{d}$
Transpose the equation to make λ the subject.	$\lambda = \frac{\Delta x d}{L}$
Substitute values into the equation and solve. (Note: 1 μm = 1×10^{-6} m)	$\lambda = \frac{0.025 \times 50 \times 10^{-6}}{2.0} = 6.25 \times 10^{-7}$ m
Express your answer using convenient units, in this case nm, where 1 nm = 1×10^{-9} m	The wavelength of the laser light is 625 nm.

Worked example: Try yourself 7.2.1

CALCULATING WAVELENGTH FROM FRINGE SEPARATION

Green laser light is directed through a pair of thin slits that are 25 μm apart. The slits are 1.5 m from a screen on which bright fringes are 3.3 cm apart. Use this information to calculate the wavelength of green light in nm.

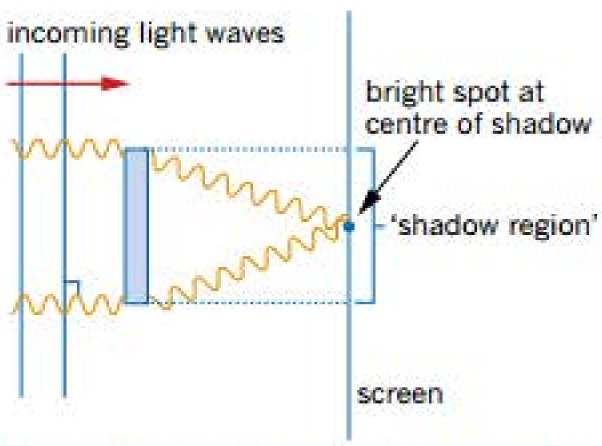


FIGURE 7.2.9 Waves of light incident on a solid disc diffract to give a point of light in the centre of the shadow zone. This is convincing evidence for the wave nature of light.

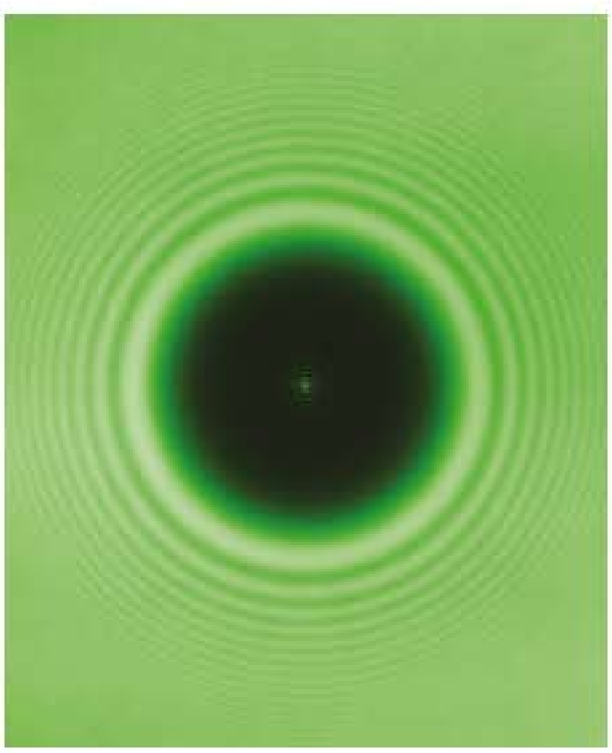


FIGURE 7.2.10 The bright spot inside the shadow region of this image is caused by the diffraction and interference of light waves. The image also shows diffraction and interference patterns surrounding the shadow.

RESISTANCE TO THE WAVE MODEL

Young's wave explanation for his experiment was not immediately accepted by the scientific community. Many scientists were reluctant to abandon the corpuscular theory that had been accepted for over a century.

In 1818, the French scientist Augustin-Jean Fresnel was able to provide a mathematical explanation for Young's double-slit experiment based on Huygens' principle.

Another French scientist, Simeon Poisson, who was a passionate supporter of Newton's particle theory, argued that if the same mathematics was applied to the light shining around a round disc, then there should be a bright spot in the middle of the shadow created by the disc (Figure 7.2.9). Since nobody had ever observed a bright spot in the middle of a shadow, Poisson believed this proved that the wave model was incorrect.

However, one of Poisson's colleagues decided to test these ideas by performing an experiment with a very small bright light source and a round disc, and observed the bright spot predicted by Poisson's calculations (Figure 7.2.10). As a consequence, for the remainder of the 19th century, the wave theory became the almost universally accepted model for light.

This now famous diffraction pattern has come to be known as the 'Poisson bright spot', which means it is named after the person who predicted that it would not exist!

7.2 Review

SUMMARY

- Young's double-slit interference experiment provided evidence to support the wave model of light.
- Path difference (pd) is the difference in the distance travelled by each wave train from a pair of slits to the same point on the screen.
- Constructive interference of coherent waves occurs when the path difference pd = $n\lambda$, where $n = 0, 1, 2, 3...$
- Destructive interference of coherent waves occurs when the path difference equals an odd number of half wavelengths or pd = $\left(n - \frac{1}{2}\right)\lambda$, where $n = 1, 2, 3...$
- The distance between the interference fringes produced in Young's experiment is given by $\Delta x = \frac{\lambda L}{d}$

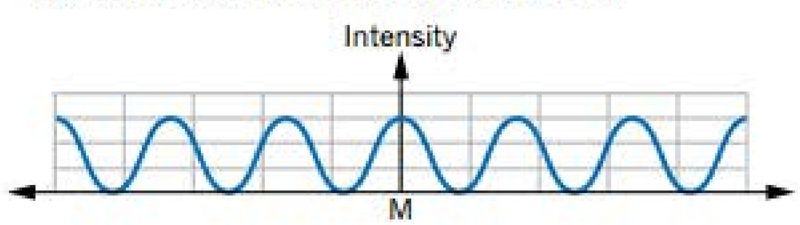
KEY QUESTIONS

1 According to the particle model of light, Young's double-slit experiment should have produced two bright lines on the screen. Instead, what was observed on the screen?
   - **A** It was completely dark.
   - **B** It was completely light.
   - **C** It contained three bright lines.
   - **D** It contained a pattern of alternating bright and dark lines.

2 Two students are trying to replicate Young's double-slit experiment. One uses torch light and the other uses light from a laser. The student using the laser light is more likely to obtain the expected interference pattern because of which one or more of the following statements?
   - **A** Torch light is monochromatic.
   - **B** Torch light is coherent.
   - **C** Laser light is monochromatic.
   - **D** Laser light is coherent.

3 If Thomas Young's double-slit experiment was modelled using circular water waves in a ripple tank, which one or more of the following events would correspond to nodal lines?
   - **A** Crests meet troughs.
   - **B** Troughs meet troughs.
   - **C** Crests meet crests.
   - **D** Troughs meet crests.

4 The following diagram shows the resulting (simplified) intensity pattern when light from two slits reaches the screen in a Young's interference experiment. Copy the diagram into your workbook and circle the points at which the path difference is equal to 1λ.



5 Explain why Young's double-slit experiment led to a significant change in scientists' understanding of the nature of light.

6 A version of Young's double-slit experiment is set up by directing the light from a red laser through a pair of thin slits. An interference pattern appears on the screen behind the slits. The following changes are made to the apparatus. Identify whether the distance between the interference fringes seen on the screen would increase, decrease or stay the same.
   - **a** The screen is moved further away from the slits.
   - **b** A green (i.e. shorter wavelength) laser is used.
   - **c** The slits are moved closer together.

7 A 580 nm yellow light is directed through a pair of thin slits to produce an interference pattern on a screen. Determine the path difference of the fifth dark fringe.

8 Identify the type of interference (constructive or destructive) that corresponds to the following path differences:
   - **a** $\frac{\lambda}{2}$
   - **b** λ
   - **c** $\frac{3\lambda}{2}$

9 A 700 nm red light is directed through a pair of thin slits to produce an interference pattern on a screen. Determine the path difference of the second bright fringe.

10 A blue laser is directed through a pair of thin slits that are 40 μm apart. The slits are 3.25 m from a screen on which bright fringes are 3.7 cm apart. Use this information to calculate the wavelength of blue light in nm.

7.3 Electromagnetic waves

FIGURE 7.3.1 Light cannot be a simple mechanical wave because it can travel through empty space.

The establishment of the wave model for light raised an important question; scientists now wanted to know what type of waves light waves were.

Experiments on polarisation provided the important information that light must be a type of transverse wave, since polarisation does not occur for longitudinal waves. However, light is obviously different from other types of mechanical waves because it can pass through the vacuum of space between the Earth and the Sun (Figure 7.3.1).

ELECTROMAGNETIC WAVES

In the middle of the 19th century, the Scottish physicist James Clerk Maxwell gained a key insight into the nature of light waves. In his mathematical study of electric and magnetic effects, he realised that some of the constants in his equations combined to closely match the current estimates of the speed of light. Maxwell went on to develop a comprehensive theory of electromagnetism in which light is a form of **electromagnetic radiation** (EMR).

The electromagnetic nature of light

As shown in earlier chapters, electric current can be used to produce a magnetic field, and a changing magnetic field can be used to generate an electromotive force (emf) or voltage. Maxwell put these two ideas together. He proposed that if a changing electric field is produced, for example by a charged particle moving backwards and forwards, then this changing electric field will produce a changing magnetic field at right angles to it, as shown in Figure 7.3.2.

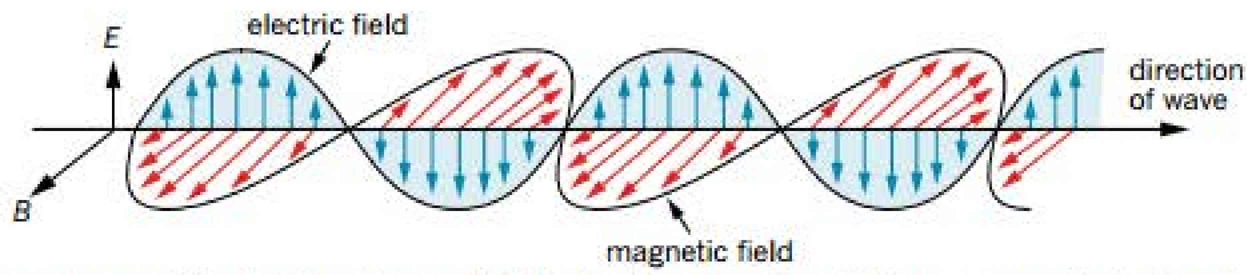


FIGURE 7.3.2 The electric and magnetic fields in electromagnetic radiation are perpendicular to each other and are both perpendicular to the direction of propagation of the radiation.

The changing magnetic field would, in turn, produce a changing electric field and the cycle would be repeated. In effect, this would produce two mutually propagating fields and the EMR would be self-propagating, i.e. it could extend outwards into space. Both the electric and magnetic fields would oscillate at the same frequency: the frequency of the light wave.

Maxwell's theoretical calculations provided a value for the speed at which EMR should propagate through empty space. This matched the experimental value for the speed of light measured by the French physicist Hippolyte Fizeau in 1849. The accepted value for the speed of light today is $299\,792\,458\,\text{m s}^{-1}$. This is such an important constant that it has been designated its own symbol, c. In calculations, the speed of light is usually approximated as $c = 3.00 \times 10^8\,\text{m s}^{-1}$.

For light and other forms of EMR, the familiar wave equation $v = f\lambda$ is usually written as:

$$c = f\lambda$$

where f is the frequency of the wave (Hz)
λ is the wavelength of the wave (m)
$c = 3.00 \times 10^8\,\text{m s}^{-1}$ is the speed of light in air or a vacuum

Maxwell's work represents a pivotal moment in the history of physics. Not only did he provide an explanation of the nature of light, he also brought together a number of formerly distinct areas of study—optics (the study of light), electricity and magnetism. As shown in the next section, Maxwell's work also encompasses other areas of physics.

Worked example 7.3.1

USING THE WAVE EQUATION FOR LIGHT

Calculate the frequency of violet light with a wavelength of 400 nm (400×10^{-9} m).	
Thinking	**Working**
Recall the wave equation for light.	$c = f\lambda$
Transpose the equation to make frequency the subject.	$f = \frac{c}{\lambda}$
Substitute in values to determine the frequency of this wavelength of light.	$f = \frac{3.0 \times 10^8}{400 \times 10^{-9}}$ $= 7.5 \times 10^{14}$ Hz

Worked example: Try yourself 7.3.1

USING THE WAVE EQUATION FOR LIGHT

A particular colour of red light has a wavelength of 600 nm. Calculate the frequency of this colour.

Searching for the aether

One of the characteristics of mechanical waves is that they require a physical medium through which to propagate. For example, sound waves usually propagate through air and water waves propagate through water. For many years, scientists searched for the physical medium through which electromagnetic waves propagate. They even went so far as to give this medium a name: the 'luminiferous ether' or 'aether'.

In a famous experiment, the scientists Albert Michelson and Edward Morley attempted to measure the presence of the aether. In this experiment a Michelson interferometer was used (Figure 7.3.3). One of the arms of the interferometer is aligned in the direction of the Earth's motion through space. An interference pattern is observed when light from the two arms recombines. When the equipment is rotated 90°, the presence of an aether should result in a slight shift in the interference pattern. This shift was never observed, despite many attempts, and thus scientists were forced to conclude that electromagnetic waves are able to propagate through a vacuum.

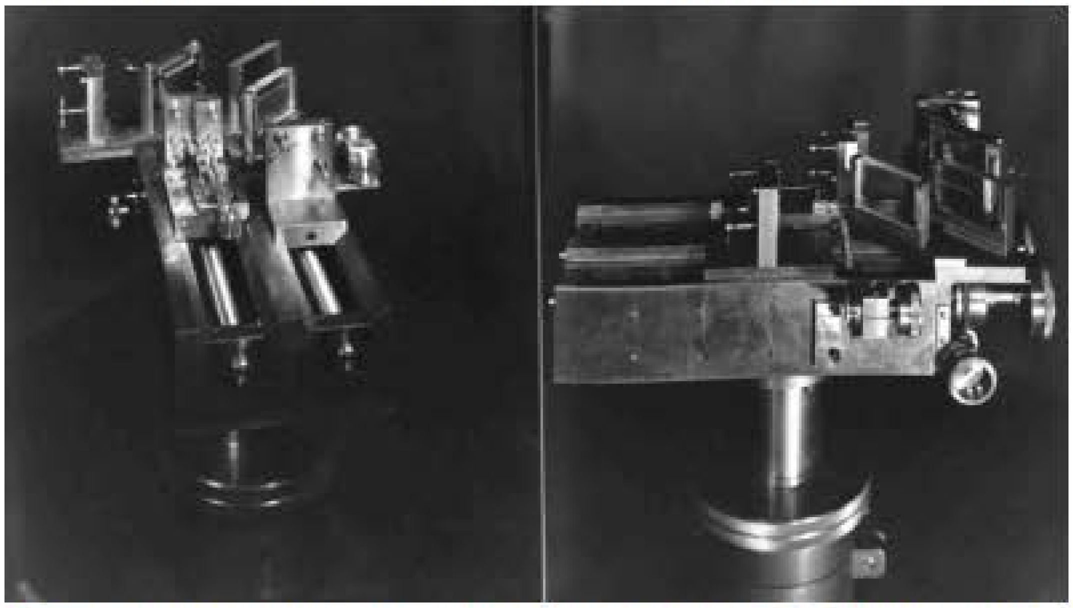


FIGURE 7.3.3 The Michelson–Morley experiment. Light is split along two arms of the interferometer, and reflected back to create an interference pattern.

The electromagnetic spectrum

The wavelengths of all the different colours of visible light fall between approximately 390 nm and 780 nm. Naturally, physicists were bound to inquire about other wavelengths of EMR. It is now understood that the visible spectrum is just one small part of a much broader set of possible wavelengths known as the **electromagnetic spectrum** (Figure 7.3.4).

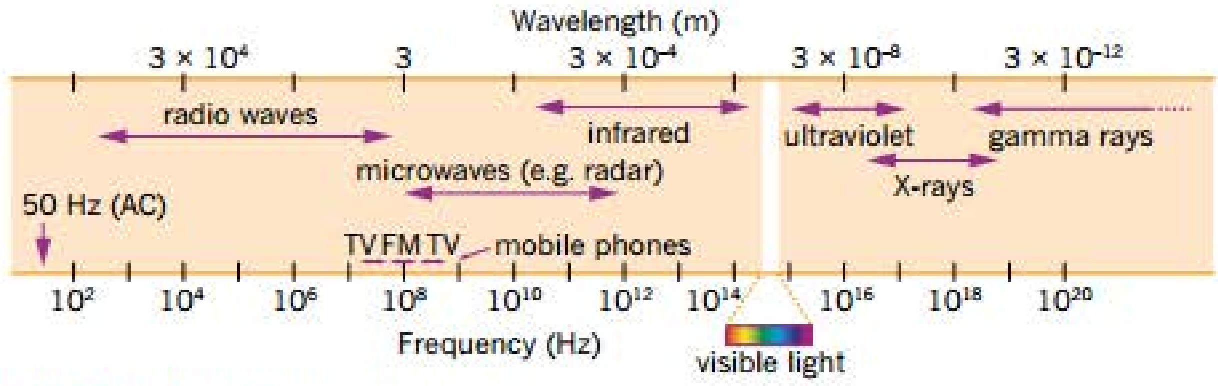


FIGURE 7.3.4 The electromagnetic spectrum.

Changing the frequency and wavelength of the waves changes the properties of the EMR, and so the electromagnetic spectrum is divided into 'bands' according to how the particular types of EMR are used. The shorter the wavelength of the electromagnetic wave, the greater is its penetrating power. This means that waves with extremely short wavelength, such as X-rays, can pass through some materials (e.g. skin), revealing the structures inside (e.g. bone).

Long wavelength waves, such as AM radio waves, have such low penetrating power that they cannot even escape Earth's atmosphere, and can be used to 'bounce' radio signals around to the other side of the world. Table 7.3.1 compares the characteristics of different waves in the electromagnetic spectrum.

TABLE 7.3.1 Comparison of the different waves in the electromagnetic spectrum

Type of wave	Typical wavelength (m)	Typical frequency (Hz)	Comparable object
AM radio wave	100	3×10^6	sports oval
FM radio or TV wave	3	1×10^8	small car
microwaves	0.03	1×10^{10}	50c coin
infrared	10^{-5}	3×10^{13}	white blood cell
visible light	10^{-7}	3×10^{15}	small cell
ultraviolet	10^{-8}	3×10^{16}	large molecule
X-ray	10^{-10}	3×10^{18}	atom
gamma ray	10^{-15}	3×10^{23}	atomic nucleus

Radio waves

One of the most revolutionary applications of EMR is the use of radio waves to transmit information from one point to another over long distances. Radio waves are the longest type of EMR, with wavelengths ranging from 1 mm to hundreds of kilometres. The principle of radio transmission is relatively simple and neatly illustrates the nature of electromagnetic waves, as illustrated in Figure 7.3.5.

The radio transmitter converts the signal (e.g. radio announcer's voice, music or stream of data) into an alternating current. When this alternating current flows in the transmission antenna, the electrons in the antenna oscillate backwards and forwards. This oscillation of charges in the antenna produces a corresponding electromagnetic wave that radiates outwards in all directions from the antenna.

When the radio wave hits the antenna of a radio receiver, the electrons in the receiver's antenna start to oscillate in exactly the same way as in the transmitting antenna.

The radio receiver then reverses the process of the transmitter, converting the alternating current from the reception antenna back into the original signal, as seen in Figure 7.3.5.

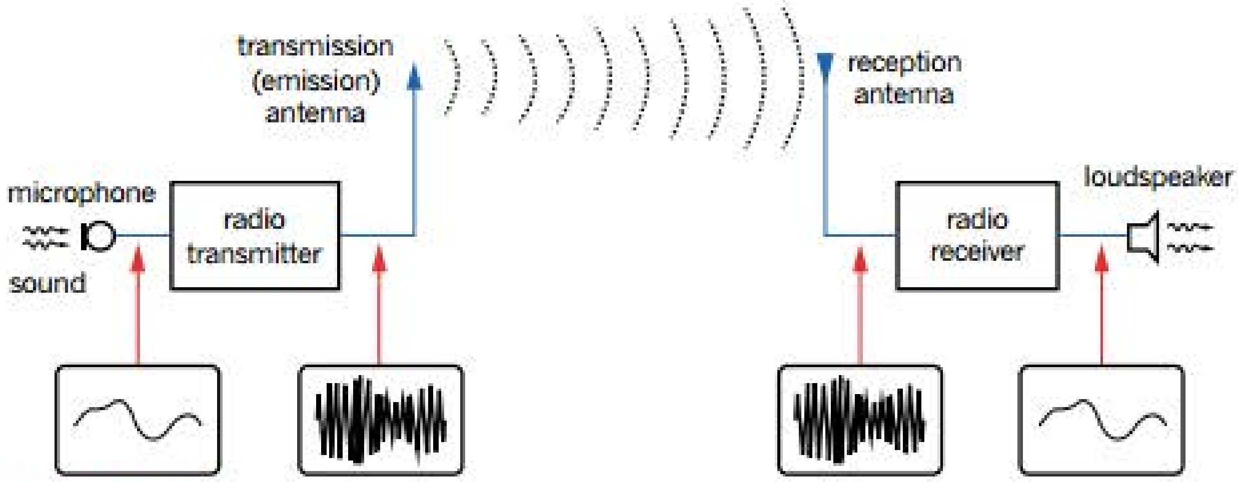


FIGURE 7.3.5 A typical radio transmission system.

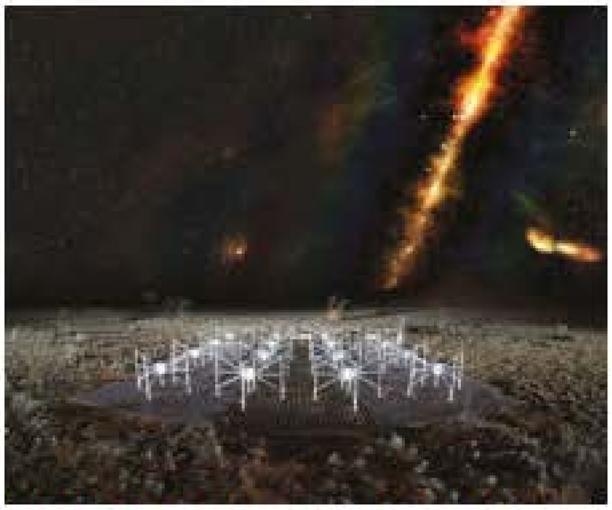


FIGURE 7.3.6 Part of the Murchison Widefield Array radio telescope in outback Western Australia superimposed on a radio image of the sky.

Radio wave telescopes are also used to investigate the universe. Figure 7.3.6 shows the Murchison Widefield Array in remote Western Australia superimposed on a radio image of the Milky Way. This radio telescope is a trial that will eventually be replaced by the Square Kilometre Array (SKA). Once the SKA is established, thousands of radio antennas linked by fibre optics will operate in unison to act as one antenna with an area of one square kilometre.

Microwaves

Microwaves have shorter wavelengths and therefore greater penetrating power than radio waves. They can be produced by devices with short antennas and hence are useful in personal communication applications such as mobile phones and wireless internet transmission. They are also particularly useful in heating and cooking food.

EXTENSION

Microwave ovens

A microwave oven is 'tuned' to produce a particular frequency of electromagnetic radiation: 2.45 GHz (i.e. 2.45×10^9 Hz). This is the resonant frequency of water molecules (Figure 7.3.7).

All solid objects have a frequency at which they will naturally vibrate. Musical instruments such as guitars and violins make use of the resonant frequencies of strings under tension, which you might recall from Chapter 10 of *Pearson Physics 11 Western Australia*.

When water molecules are bombarded with radiation with a frequency of 2.45 GHz, they start to vibrate quickly. Since this increases the average kinetic energy of the water molecules, the temperature of the water in the substance increases. Effectively, the microwaves cause the water to heat up.

FIGURE 7.3.7 Microwave ovens produce electromagnetic radiation with a frequency of 2.45 GHz, which is the resonant frequency of water molecules.

This heat then transfers by conduction and convection to the rest of the food. This is why food sometimes becomes soggy when heated in the microwave: the water molecules heat up faster than the food molecules around them. It also explains why recipes that do not contain much water cannot be cooked well in a microwave oven.

PHYSICSFILE

AM and FM

A radio wave pattern is produced using a 'carrier wave' of fixed frequency. This frequency is the 'channel' that the radio 'tunes into'. Many radio stations use the carrier wave frequency as part of their name, e.g. Nova 100.3 transmits using a 100.3 MHz carrier wave.

The carrier wave is altered or 'modulated' by the signal containing the information to be transmitted. An AM radio system uses 'amplitude modulation', which means that the amplitude of the carrier wave is modulated to match the signal. In comparison, FM stands for 'frequency modulation', in which the frequency of the carrier wave is changed to represent the signal. In terms of circuitry, AM systems are much simpler than FM systems, although FM radio waves tend to transmit signals more clearly.

Infrared

The infrared section of the electromagnetic spectrum lies between microwaves and visible light. Infrared waves are longer than the red waves of the visible spectrum, hence their name.

Infrared light waves become useful because they are emitted by objects, to varying degrees, due to their temperature (see the Physics file 'Night vision'). The warmth that you feel standing next an electric bar heater or fire is due to infrared radiation (Figure 7.3.8). The radiant heat the Earth receives from the Sun is transmitted in the form of infrared waves; life on Earth would not be possible without this important form of EMR.

FIGURE 7.3.8 The coals of a fire appear red because they release red light along with infrared radiation, which you experience as heat.

Ultraviolet light

As their name suggests, ultraviolet (UV) waves have wavelengths that are shorter than those of violet light and therefore cannot be detected by the human eye. Their shorter wavelengths give UV rays stronger penetrating power than visible light. In fact, UV rays can actually penetrate human skin and damage the DNA of skin cells, producing harmful skin cancers.

Scientists can make use of UV light to take images. Figure 7.3.10 is a UV image of the surface of the Sun taken after a solar flare has occurred. The image has been re-coloured so that it highlights areas of different temperature. Here, areas that are coloured white are the hottest. Images like this help scientists to learn about the temperatures of very hot objects. Taking an image of the Sun using visible light would not allow this same distinction.

PHYSICSFILE

Night vision

Night-vision goggles enhance visibility in low light conditions by greatly amplifying the visible light available and also by detecting a small part of the infrared radiation that is emitted by objects due to their temperature.

FIGURE 7.3.9 Night-vision goggles enhance visibility.

FIGURE 7.3.10 Re-coloured UV image of the surface of the Sun. The white areas reveal the hottest parts.

X-rays and gamma rays

X-rays and gamma rays have much shorter wavelengths than visible light. This means that these forms of EMR have very high penetrating powers. For example, some X-rays can pass through different types of human tissues, which means that they are very useful in medical imaging (Figure 7.3.11).

FIGURE 7.3.11 This X-ray image of a human hand could be formed because X-rays can pass through human tissue.

Unfortunately, this useful penetrating property of X-rays comes with inherent dangers. As X-rays pass through a human cell, they can do damage to the tissue, sometimes killing the cells or damaging the DNA in the cell nucleus, leading to harmful cancers. For this reason, a person's exposure to X-rays has to be carefully monitored to minimise harmful side effects.

Similarly, exposure to gamma rays can be very dangerous to human beings. The main natural sources of gamma radiation exposure are the Sun and radioactive isotopes. Fortunately, Earth's atmosphere protects people from most of the Sun's harmful gamma rays, and radioactive isotopes are not commonly found in sufficient quantities to produce harmful doses of radiation.

7.3 Review

SUMMARY

- Although light exhibits many wave properties, it cannot solely be modelled as a mechanical wave because it can travel through a vacuum.
- Light is a form of electromagnetic radiation.
- Electromagnetic waves are transverse waves made up of mutually perpendicular, oscillating electric and magnetic fields.
- Electromagnetic radiation can be used for a variety of purposes depending on the properties of the waves, which are determined by their frequency and wavelength.
- Oscillating charges produce electromagnetic waves of the same frequency as the oscillation. Electromagnetic waves cause charges to oscillate at the frequency of the wave.
- Light (i.e. all electromagnetic radiation) travels through a vacuum at approximately $c = 3.0 \times 10^8 \text{ m s}^{-1}$.
- The wave equation $c = f\lambda$ can be used to calculate the frequency and wavelength of electromagnetic waves.

KEY QUESTIONS

1 What is a key difference between light waves and mechanical waves?
- **A** Light waves do not have a measurable wavelength.
- **B** Light waves can travel through a vacuum.
- **C** The speed of light is too fast to be accurately measured.
- **D** Light waves do not undergo diffraction.

2 In an electromagnetic wave, what is the orientation of the changing electric and magnetic fields?
- **A** at 45° to each other
- **B** parallel to each other
- **C** parallel but in opposite directions
- **D** perpendicular to each other

3 What type of electromagnetic radiation would have a wavelength of 200 nm?
- **A** radio waves
- **B** microwaves
- **C** visible light
- **D** ultraviolet light

4 Arrange the following types of electromagnetic radiation in order of increasing wavelength.
FM radio waves, visible light, infrared radiation, X-rays

5 Calculate the frequencies of the following wavelengths of light.
- **a** red light of wavelength 656 nm
- **b** yellow light of wavelength 589 nm
- **c** blue light of wavelength 486 nm
- **d** violet light of wavelength 397 nm

6 Although the currently accepted value for the speed of light is 299 792 458 m s^{-1}, this is often approximated as $c = 3.00 \times 10^8 \text{ m s}^{-1}$. Calculate the percentage error introduced by this approximation.

7 Calculate the wavelength (in nm) of light with a frequency of 6.0×10^{14} Hz.

8 Calculate the wavelength of a UHF (ultra-high frequency) television signal with a frequency of 7.0×10^7 Hz.

9 Calculate the frequency of an X-ray with a wavelength of 200 pm.

10 Calculate the wavelength of the electromagnetic waves produced by a microwave oven, if the frequency of electromagnetic radiation produced is 2.45 GHz.

7.4 Light quanta: Blackbody radiation and the photoelectric effect

FIGURE 7.4.1 Albert Einstein helped revolutionise our understanding of the nature of light.

At the turn of the 20th century, a number of scientists turned their attention to light phenomena that could not be readily explained using Maxwell's electromagnetic wave model. The study of these phenomena required the development of much more sophisticated models for light, and eventually led to a revolution in the scientific understanding of the nature of energy and matter. One of the scientists who made a significant contribution to this new way of understanding light was Albert Einstein (Figure 7.4.1).

RADIATION

All objects that have temperatures above absolute zero emit electromagnetic radiation. For objects at room temperature, practically all of this radiation is emitted in the infrared part of the electromagnetic spectrum, which is not visible to the human eye (Figure 7.4.2).

FIGURE 7.4.2 This infrared image (a thermogram) shows that the woman's face is much hotter than the clothes—and that her nose must be cold

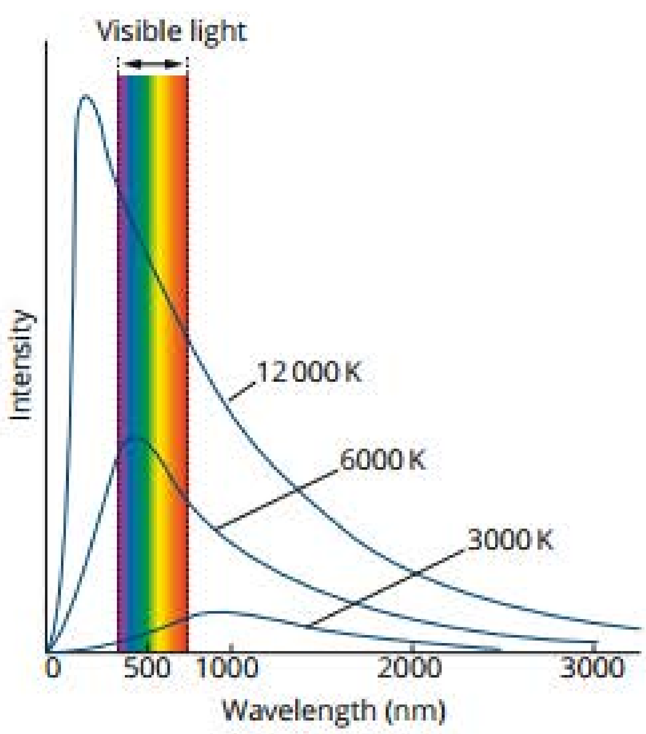


FIGURE 7.4.3 Objects at temperatures of 3000 K, 6000 K and 12 000 K have different emission spectra. As the temperature increases, the amount of energy radiated (shown by the area under these graphs) increases, and the peak wavelength decreases (it becomes more blue).

At higher temperatures, not only is the total amount of radiation emitted larger, but also a higher proportion of it is in the visible part of the spectrum. At temperatures of about 3000 K, a significant amount of energy is emitted as red wavelengths of the visible spectrum and the object appears 'red hot' (Figure 7.3.8 on page 258).

As the temperature of the object increases further, the proportion of the electromagnetic radiation emitted in the visible spectrum increases and more of the energy is radiated at shorter wavelengths (Figure 7.4.3). At about 6000 K, energy is radiated roughly evenly across different wavelengths of the visible spectrum, so the object appears 'white hot'. At 12 000 K, more of the energy is being radiated in the ultraviolet and violet parts of the spectrum, so the object appears 'blue hot'.

The shape of the emission spectrum generally depends on the type and shape of the material as well as its temperature. However, an ideal blackbody emits a radiation spectrum that only depends on its temperature.

PHYSICSFILE

Measuring the temperatures of stars

With an understanding of the radiation spectra of different objects, a great deal of information is gained about a star by assessing the colour of the light it emits. Colour is so important to astronomers that they refer to it in their classification system for stars, using names such as red giant, white dwarf, blue supergiant.

A graph that plots the absolute magnitude (luminosity relative to the Sun) of a star against its surface temperature is called a Hertzsprung–Russell diagram (Figure 7.4.4). Low-mass stars are cool and red. Medium-mass stars such as the Sun (circled in Figure 7.4.4) are yellow. High-mass stars are hot and blue.

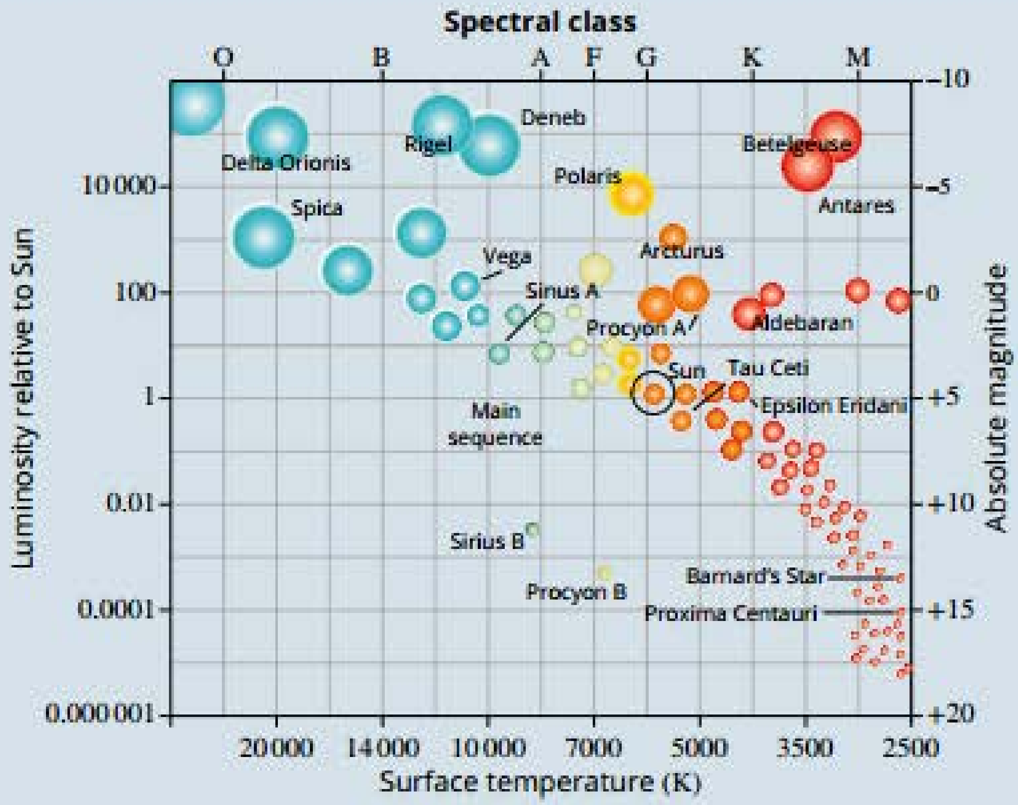


FIGURE 7.4.4 Stellar classification of stars using a Hertzsprung–Russell diagram.

BLACKBODY RADIATION

A **blackbody** is an ideal absorber and absorbs all incident radiation; it is in equilibrium with its surroundings, as it radiates and absorbs energy at the same rate, and so its temperature remains constant. A blackbody can be approximated by the inside of a cavity with a small entrance hole, because all light entering the hole is trapped. The shape of the radiation spectrum then only depends on temperature, as shown in Figure 7.4.5.

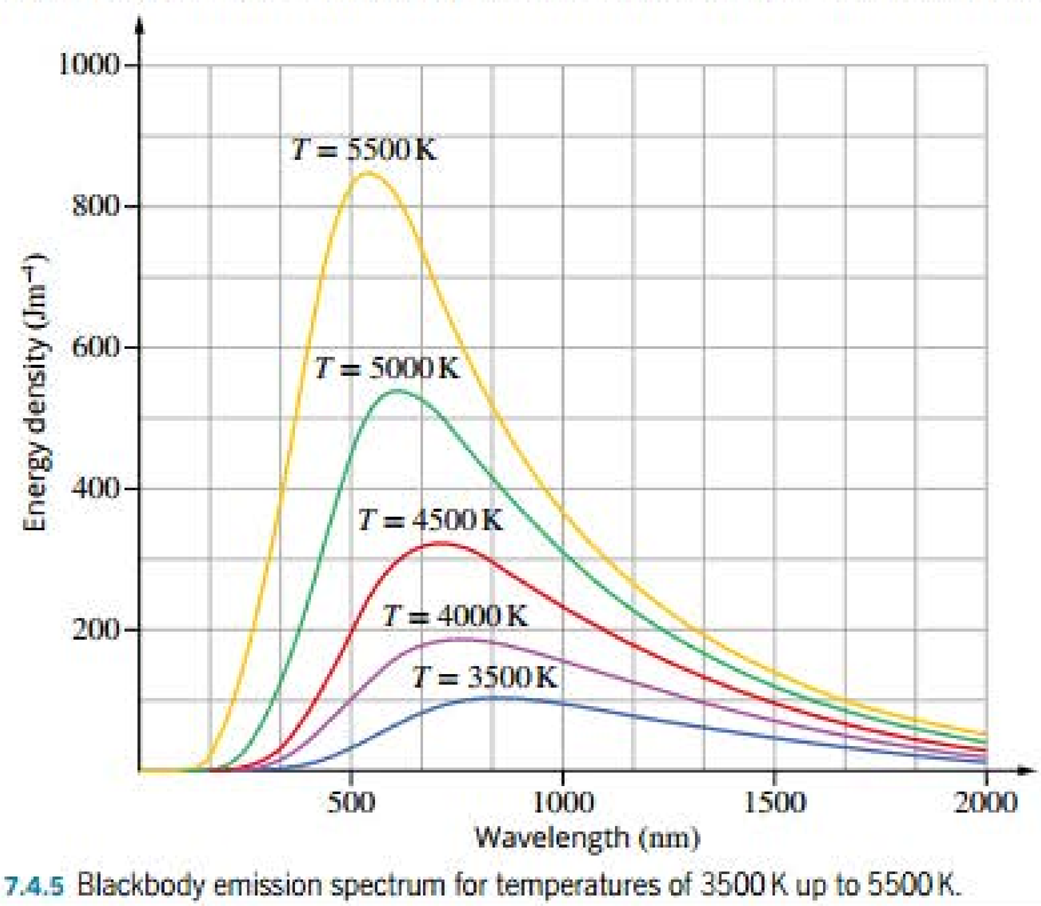


FIGURE 7.4.5 Blackbody emission spectrum for temperatures of 3500 K up to 5500 K.

As the temperature increases, the peak in energy density increases and the total energy emitted (represented by the area under the curve) by the object increases. In addition, the peak wavelength (λ_{max}) decreases; this corresponds to an increase in peak frequency and energy. This shift to shorter wavelengths was found to have the following relationship, known as Wien's displacement law.

$\lambda_{max} T = 0.2898 \times 10^{-2}$ (Units m K)

where T is the temperature in kelvin and λ_{max} is in metres.

Classically, thermal radiation is emitted by accelerating charges near the surface of a material. The charges have a distribution of accelerations that leads to a range of thermal energies. The classical expression, known as the Rayleigh–Jeans law, gives the average energy per oscillating or vibrating charge proportional to temperature. The radiated energy can be considered to be produced by standing waves or resonant modes within the cavity. At long wavelength, λ, there is a reasonable agreement between the experimental data and the calculation, as shown in Figure 7.4.6.

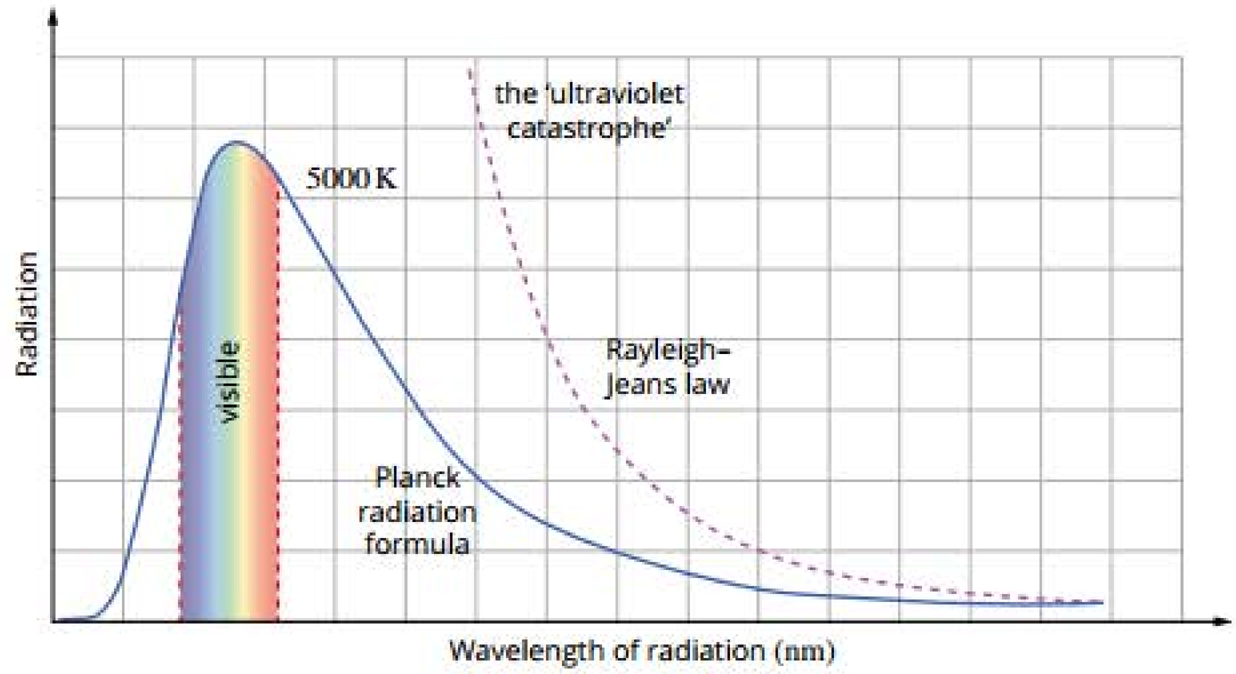


FIGURE 7.4.6 Comparison of the experimentally observed blackbody radiation intensity and the Rayleigh–Jeans classical calculation.

However, as the wavelength gets shorter (or the frequency gets higher) the number of modes continues to get larger and larger, so the energy density increases and approaches infinity; this is known as the ultraviolet catastrophe.

In order for the theory to fit experimental observations Planck incorporated two fundamental postulates:

1. Molecules vibrate at discrete energies or frequencies, as shown in Figure 7.4.7. These energies are given by

   $E_n = nhf$

   E_n is the energy of vibration

   h is Planck's constant and $h = 6.63 \times 10^{-34}$ J s

   $n = 1, 2, 3 \ldots$

2. Molecules emit and absorb radiation in discrete packets called **photons**, and the energy of an individual photon is $E = hf$. To gain energy a molecule absorbs a photon and to lose energy it emits a photon, as shown in Figure 7.4.7.

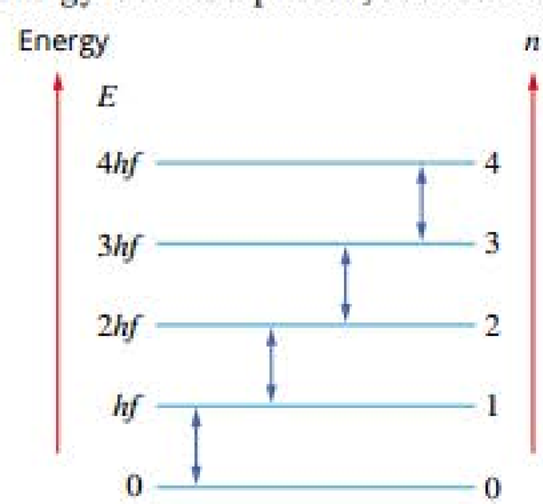


FIGURE 7.4.7 Molecules vibrate at discrete energies, $E_n = nhf$. Molecules gain energy by absorbing a photon (upwards arrow) and lose energy by emitting a photon (downwards arrow). The photons emitted are typically low energy, in the infrared at room temperature.

PHYSICSFILE

Max Karl Ernst Ludwig Planck

Max Planck (1858–1947) was a German physicist. At the age of 21 he obtained a PhD in physics, and in 1889 was appointed professor at Berlin University. Planck was an author of numerous works about physics, and about quantum theory in particular. On 14 December 1900 he presented the correct version of the Wien formula and introduced a new constant—Planck's constant. This date is now recognised as the beginning of the era of quantum mechanics. In 1918, Planck was awarded the Nobel Prize in Physics for the discovery of the quantum nature of energy.

FIGURE 7.4.8 Max Planck (1858–1947).

PLANCK'S EQUATION

As discussed above, Max Planck proposed that light existed in discrete packets of energy. He called the discrete packets of energy 'quanta', and developed an equation for the energy, E, of each **quantum**:

> $E = hf$
>
> where E is the energy of a quantum of light (J)
>
> f is the frequency of the electromagnetic radiation (Hz)
>
> h is the constant 6.63×10^{-34} J s, now known as Planck's constant

Electromagnetic radiation is often described according to its wavelength, so Planck's equation can be combined with the wave equation for light $c = f\lambda$ as follows:

$$E = hf \text{ and } f = \frac{c}{\lambda}$$

so

> $$E = \frac{hc}{\lambda}$$

At the time, most scientists disregarded Planck's work, as the particle model it suggested was very much at odds with the wave model that had become the widely accepted explanation for light.

Worked example 7.4.1

USING PLANCK'S EQUATION

Calculate the energy in joules of a quantum of ultraviolet light that has a frequency of 2.00×10^{15} Hz.	
Thinking	**Working**
Recall Planck's equation.	$E = hf$
Substitute in the appropriate values to solve.	$E = 6.63 \times 10^{-34} \times 2.00 \times 10^{15}$ $= 1.33 \times 10^{-18}$ J

Worked example: Try yourself 7.4.1

USING PLANCK'S EQUATION

> Calculate the energy in joules of a quantum of infrared radiation that has a frequency of 3.6×10^{14} Hz.

THE ELECTRON-VOLT

When studying light, the quantities of energy considered are usually very small. These are often so small that the joule (J) is no longer a convenient unit to use. Scientists have therefore adopted a unit called an **electron-volt** (eV). An electron-volt is the amount of energy an electron gains when it moves through a potential difference of 1V. Recall that work done or energy gained is given by $W = qV$, and that the charge, q, on an electron is -1.6×10^{-19} C. Then:

$$\begin{aligned} 1\,\text{eV} &= 1\text{e} \times 1\text{V} \\ &= 1.6 \times 10^{-19}\,\text{C} \times 1\,\text{J}\,\text{C}^{-1} \\ &= 1.6 \times 10^{-19}\,\text{J} \end{aligned}$$

Here is a simple way to convert between the units for energy:

> $1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$
> To convert a value expressed in J into eV, divide it by $1.6 \times 10^{-19}\,\text{J eV}^{-1}$.
> To convert a value expressed in eV into J, multiply it by $1.6 \times 10^{-19}\,\text{J eV}^{-1}$.

Worked example 7.4.2

CONVERTING TO ELECTRON-VOLTS

A quantum of light has 1.33×10^{-18} J of energy. Convert this energy to electron-volts.	
Thinking	**Working**
Recall the conversion rate for joules to electron-volts.	$1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$
Divide the value expressed in joules by $1.6 \times 10^{-19}\,\text{J eV}^{-1}$ to convert to electron-volts.	$\frac{1.33\times10^{-18}}{1.6\times10^{-19}} = 8.3\,\text{eV}$

Worked example: Try yourself 7.4.2

CONVERTING TO ELECTRON-VOLTS

A quantum of light has 2.4×10^{-19} J of energy. Convert this energy to electron-volts.

As seen from Worked examples 7.4.1 and 7.4.2, it is easier to compare the relative energies of quanta when they are expressed in eV.

For convenience, Planck's constant can also be given in terms of electron-volts:

$$h = 6.63 \times 10^{-34}\,\text{J s}$$
$$= \frac{6.63\times10^{-34}}{1.6\times10^{-19}}$$
$$= 4.14 \times 10^{-15}\,\text{eV s}$$

THE PHOTOELECTRIC EFFECT

At the start of the 20th century, another phenomenon that could not be explained using the wave model for light was being observed.

Scientists noticed that when some types of electromagnetic radiation are incident on a piece of metal, the metal becomes positively charged. This positive charge is due to electrons being ejected from the surface of the metal. The electrons became known as **photoelectrons** because they were released due to light or other forms of electromagnetic radiation. The phenomenon is known as the **photoelectric effect**.

A common apparatus used to observe the photoelectric effect consists of a clean metal surface (the cathode) illuminated with light from an external source. If the light causes photoelectrons to be emitted, they are detected at the anode. The flow of electrons is called the **photocurrent** and is registered by a sensitive ammeter. A typical circuit used to investigate the photoelectric effect is shown in Figure 7.4.9 and includes a variable voltage supply.

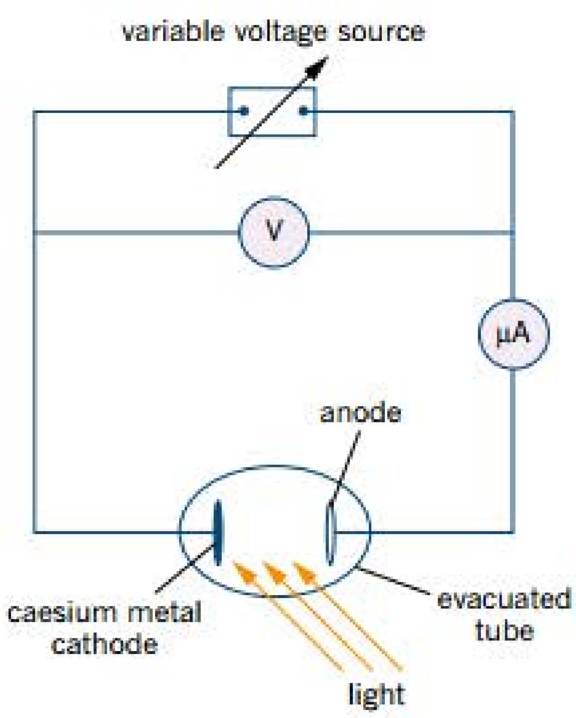


FIGURE 7.4.9 Circuit diagram of an experimental investigation of the photoelectric effect.

A **forward voltage** or forward bias is created when the cathode is negative and the anode positive. In this case the negatively charged photoelectrons will be accelerated or helped across the gap to the positively charged anode by the resulting electric field, thus creating a photocurrent in the circuit.

A **reverse voltage** or reverse bias is created when the cathode is positive and the anode is negative. The electrons emitted are repelled by the anode, and this slows them down. As the anode voltage is increased, the photoelectrons are repelled more and more until the photocurrent drops to zero.

Using the apparatus shown in Figure 7.4.9, the German physicist Philipp Lenard made a number of surprising discoveries about the photoelectric effect, for which he won the Nobel Prize in Physics in 1905.

Lenard used a filter to vary the frequency of the incident light. He discovered that for a particular cathode metal there is a certain frequency below which no photoelectrons are observed. This is called the **threshold frequency**, f_0. For frequencies of light greater than the threshold frequency (i.e. $f > f_0$), photoelectrons will be collected at the anode and registered as a photocurrent. For frequencies below the threshold frequency (i.e. $f < f_0$), no photoelectrons will be detected.

Lenard also discovered that, for light that has a frequency greater than the threshold frequency (i.e. $f > f_0$), the rate at which the photoelectrons are produced varies in proportion with the intensity (brightness) of the incident light, as shown in Figure 7.4.10.

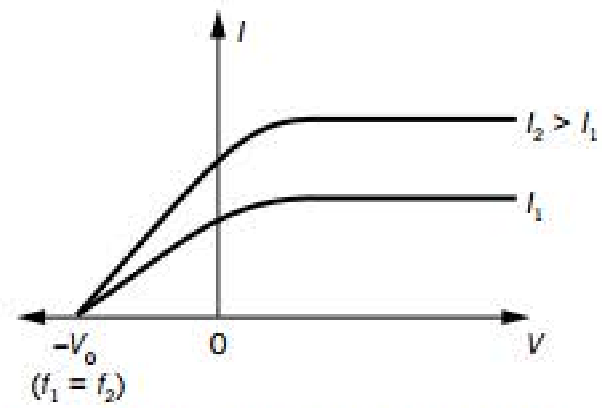


FIGURE 7.4.10 Photocurrent (I) plotted as a function of the voltage (V) applied between the cathode and the anode for different light intensities. For brighter light ($I_2 > I_1$) of the same frequency ($f_1 = f_2$), there is a higher photocurrent, but the same stopping voltage, V_0.

This graph shows a number of important properties of the photoelectric effect.

- As the light intensity increases, the photocurrent increases.
- At zero applied voltage there is still a photocurrent due to the kinetic energy of the photoelectrons.
- When the applied voltage is positive, photoelectrons are attracted to the collector electrode (anode). A small positive voltage is enough to ensure that every available photoelectron is collected. The current therefore reaches a maximum value and remains there even if the voltage is increased.
- When the applied voltage is negative, the cathode is positive. Therefore the photoelectrons are attracted back towards the illuminated cathode, repelled by the collector electrode (anode) and the photocurrent is reduced. As the reverse bias becomes increasingly negative, the photocurrent is reduced because fewer and fewer photoelectrons have the kinetic energy to overcome the opposing electric potential. There is a voltage, V_0, for which no photoelectrons reach the collector. This is known as the **stopping voltage**. For a particular metal, each frequency of light will give a characteristic stopping voltage. This value is independent of light intensity.

Recall that the work done on a charge by an applied voltage is given by $W = qV$. In this case, the voltage used is designated the stopping voltage, V_0, and q_e is the charge on an electron: $q_e = -1.6 \times 10^{-19}$ C. Hence the work done on the electron is given by $W = q_e V_0$. Since the stopping voltage is large enough to stop even the fastest-moving electrons from reaching the anode, this expression gives the value of the maximum possible kinetic energy of the emitted photoelectrons. For example, when the stopping voltage is −2.5 V, the maximum kinetic energy of any photoelectron is 2.5 eV.

$$E_{k\,max} = W = q_e V_0$$

where $E_{k\,max}$ is the maximum kinetic energy of the emitted photoelectrons
W is the work done on the electron
q_e is the electron charge
V_0 is the stopping potential or voltage

When light sources of the same intensity but different *frequencies* are used, they produce the same maximum current. However, the higher frequency light has a higher stopping voltage (Figure 7.4.11).

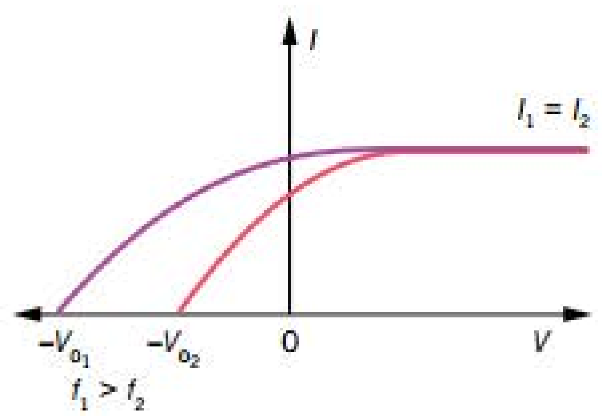


FIGURE 7.4.11 Photocurrent (I) plotted as a function of the voltage (V) applied between the cathode and the anode for different frequencies ($f_1 > f_2$) of incident light with the same intensity ($I_1 = I_2$). Both frequencies produce the same maximum photocurrent; however, light with the higher frequency requires a larger stopping voltage.

Finally, as long as the incident light has a frequency above the threshold frequency of the cathode material, photoelectrons are found to be emitted without any appreciable time delay. This fact holds true regardless of the intensity of the light.

When illuminated with light above the threshold frequency, electrons with maximum kinetic energy are usually emitted from the first layer of atoms at the surface of the metal and are the least tightly bound. Other photoelectrons come from deeper inside the metal and lose some of their kinetic energy due to collisions on their way to the surface. Hence, the emitted photoelectrons have a range of kinetic energies from the maximum value downwards.

EXPLAINING THE PHOTOELECTRIC EFFECT

The characteristics of the photoelectric effect could not be explained using a wave model of light. According to the wave model, the frequency of light should be irrelevant to whether or not photoelectrons are ejected. Since a wave is a form of continuous energy transfer, it would be expected that even low-frequency light should transfer enough energy to emit photoelectrons if left incident on the metal for long enough. Similarly, the wave model predicts that there should be a time delay between the light striking the metal and photoelectrons being emitted, as the energy from the wave builds up in the metal over time.

The dual nature of light

In 1905, Albert Einstein proposed a solution to this problem. Einstein drew on Planck's earlier work by assuming that light exists as particles, or photons (like Planck's 'quanta'), each with an energy of $E = hf$. This assumption made the properties of the photoelectric effect relatively easy to explain.

Einstein's work was actually a significant extension of Planck's ideas. Although Planck had assumed that light was being emitted in quantised packets, he never questioned the assumption that light was fundamentally a wave phenomenon.

Einstein's work went further, challenging scientists' understanding of the nature of light itself.

Einstein and the photoelectric effect

Einstein identified that, for a particular metal, the amount of energy required to eject a photoelectron is a constant value that depends on the strength of the bonding within the metal. This energy was called the **work function**, ϕ, of the metal. For example, the work function of lead is 4.14 eV, which means that 4.14 eV of energy is needed to just release one electron from the surface of a piece of lead.

According to Einstein's model, shining light on the surface of a piece of metal is equivalent to bombarding it with photons. When a photon strikes the metal, it can transfer its energy to an electron. Importantly, a single photon interacts with a single electron, transferring all of its energy at once to the electron.

If the energy of the photon is less than the work function, then photoelectrons will not be released, as the electrons will not gain enough energy to let them break free of the lead atoms. For example, the photons of violet light ($f = 7.50 \times 10^{14}$ Hz) each contain 3.11 eV of energy.

$$
\begin{aligned}
E &= hf \\
&= \frac{6.63\times10^{-34}}{1.6\times10^{-19}}\times7.5\times10^{14} \\
&= 3.11\,\text{eV}
\end{aligned}
$$

This means that violet light shining on lead would not release photoelectrons since the energy of each photon, 3.11 eV, is less than the work function of lead, 4.14 eV.

However, ultraviolet photons of frequency 1.20×10^{15} Hz each contain 4.97 eV of energy.

$$
\begin{aligned}
E &= hf \\
&= \frac{6.63\times10^{-34}}{1.6\times10^{-19}}\times1.2\times10^{15} \\
&= 4.97\,\text{eV}
\end{aligned}
$$

Therefore ultraviolet light of this frequency would release photoelectrons from lead since the energy of each photon, 4.97 eV, is greater than the work function of lead, 4.14 eV.

Each metal has a threshold frequency—this is the frequency at which the photons have an energy equal to the work function of the metal:

$\phi = hf_0$
where ϕ is the work function (J or eV)
h is Planck's constant
f_0 is the threshold frequency for that metal (Hz)

Worked example 7.4.3

CALCULATING THE WORK FUNCTION OF A METAL

Calculate the work function (in J and eV) for aluminium, which has a threshold frequency of 9.8×10^{14} Hz.	
Thinking	**Working**
Recall the formula for work function.	$\phi = hf_0$
Substitute the threshold frequency of the metal into this equation.	$\phi = 6.63 \times 10^{-34} \times 9.8 \times 10^{14}$ $= 6.5 \times 10^{-19}$ J
Convert this energy from J to eV.	$\phi = \frac{6.5 \times 10^{-19}}{1.6 \times 10^{-19}}$ $= 4.1$ eV

Worked example: Try yourself 7.4.3

CALCULATING THE WORK FUNCTION OF A METAL

Calculate the work function (in J and eV) for gold, which has a threshold frequency of 1.2×10^{15} Hz.

THE KINETIC ENERGY OF PHOTOELECTRONS

If the energy of the photon is greater than the work function of the metal, then a photoelectron is released. The remainder of the energy in excess of the work function is transformed into the kinetic energy of the photoelectron.

Einstein described this relationship with his photoelectric equation:

$E_{k\,max} = hf - \phi$
where $E_{k\,max}$ is the maximum kinetic energy of an emitted photoelectron (J or eV)
ϕ is the work function of the metal (J or eV)
h is Planck's constant
f is the frequency of the incident photon (Hz)

Graphing the maximum kinetic energy versus frequency in Einstein's equation results in a linear (straight line) graph, like the one shown in Figure 7.4.12. Such a graph is useful because it clearly shows key information such as the work function and threshold frequency for a particular metal.

Einstein's equation, $E_{k\,max} = hf - \phi$ can be compared with the equation of a straight line, $y = mx + c$. In making this comparison, it can be seen that extrapolating (extending) the graph back to the vertical axis will give the magnitude of the work function, ϕ (Figure 7.4.12). The gradient of the graph is Planck's constant, h. The threshold frequency is shown on the graph by the x-intercept, where $E_{k\,max} = 0$. At the threshold frequency, electrons are no longer bound to the metal, but they have no kinetic energy. From the graph it is also apparent that, as soon as the threshold frequency is exceeded, an electron at the surface is able to be ejected and escape with some kinetic energy. The greater the frequency (and hence energy) of the light, the greater the kinetic energy of the photoelectron.

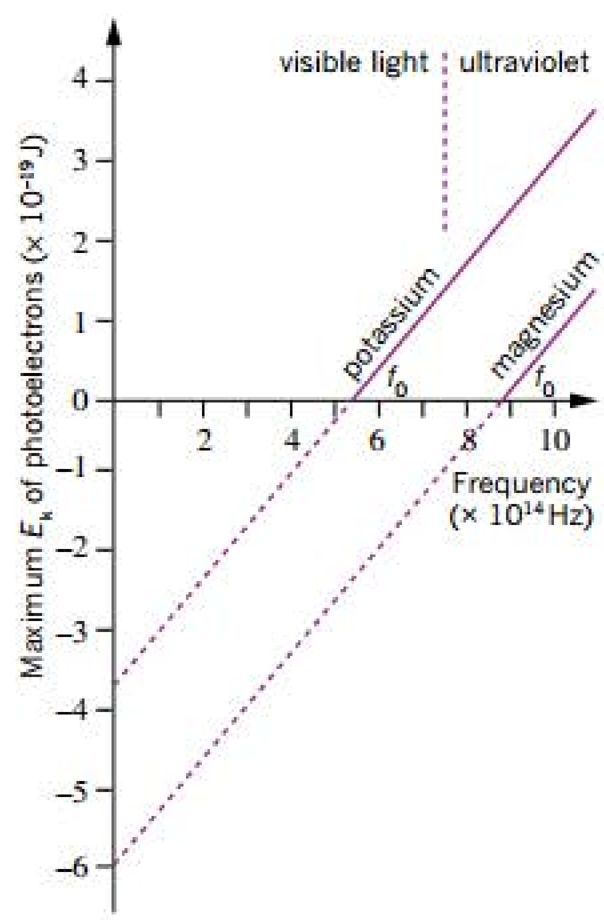


FIGURE 7.4.12 Magnesium has a high threshold frequency that is in the ultraviolet region. The x-intercept gives the threshold frequency, f_0. For potassium, the threshold frequency is in the visible region. The gradient of the graph for each metal is Planck's constant, h. The magnitude of the y-intercept gives the work function, ϕ.

Worked example 7.4.4

CALCULATING THE KINETIC ENERGY OF PHOTOELECTRONS

Calculate the kinetic energy (in eV) of the photoelectrons emitted from lead by ultraviolet light with a frequency of 1.20×10^{15} Hz. The work function of lead is 4.14 eV.	
Thinking	**Working**
Recall Einstein's photoelectric equation.	$E_{k\ max} = hf - \phi$
Substitute values into this equation.	$hf = \frac{6.63 \times 10^{-34}}{1.6 \times 10^{-19}} \times 1.20 \times 10^{15} = 4.97\ \text{eV}$ $E_{k\ max} = 4.97 - 4.14$ $= 0.83\ \text{eV}$

Worked example: Try yourself 7.4.4

CALCULATING THE KINETIC ENERGY OF PHOTOELECTRONS

Calculate the kinetic energy (in eV) of the photoelectrons emitted from lead by ultraviolet light with a frequency of 1.50×10^{15} Hz. The work function of lead is 4.14 eV.

PHYSICSFILE

Einstein's Nobel Prize

Although Albert Einstein is most famous for his work on relativity (and its related equation $E = mc^2$), he gained his Nobel Prize 'for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect'. His work on relativity was never formally recognised with a Nobel Prize.

Resistance to the quantum model of light

This new particle or 'quantum' model of light was not initially well received by the scientific community. It had already been well established that a discrete particle model for light could not explain many of light's properties such as polarisation and the interference patterns produced in Young's experiment.

Most scientists believed instead that wave explanations for the photoelectric effect would be found. However, eventually the quantum model of light was accepted and the Nobel Prize in Physics was awarded to both Planck (1918) and Einstein (1921) for their groundbreaking work in this field.

PHYSICS IN ACTION

Photovoltaic cells

The photovoltaic cells that are used in many solar panels work on the principle of the photoelectric effect (Figure 7.4.13). Sunlight falling on the solar panel provides energy that causes photoelectrons to be emitted as a current that can be used to drive electrical appliances.

However, whereas many photoelectric effect experiments use high-energy photons of ultraviolet light, photovoltaic cells use materials that will produce photoelectrons when exposed to visible light. Most commonly these are semiconducting materials based on silicon 'doped' with small amounts of other elements. Although solar cells are designed to produce the highest current possible from sunlight, most commercially available solar cells have an energy efficiency of less than 20%.

FIGURE 7.4.13 Solar panels are used to convert sunlight into electrical energy using the photoelectric effect.

Scientists hope to improve this in order to make solar cells an economic alternative to fossil fuels for large-scale energy generation.

One approach includes research on other semiconductors that absorb light more efficiently.

7.4 Review

SUMMARY

- A number of phenomena related to the behaviour of light and electromagnetic radiation, such as blackbody radiation and the photoelectric effect, can only be explained using the concept of photons, or light quanta.
- All objects with temperature greater than absolute zero emit electromagnetic radiation.
- A blackbody absorbs all incident radiation. It radiates and absorbs radiation at the same rate and is in equilibrium with its surroundings.
- The blackbody radiation distribution depends only on temperature and cannot be explained classically, particularly with emitted radiation of high energy, low frequency.
- Planck explained the blackbody radiation distribution by stating that molecules vibrate only at discrete or particular frequencies or energies.
- A blackbody absorbs and emits radiation in discrete packets called photons.
- The energy of a photon is proportional to its frequency: $E = hf = \frac{hc}{\lambda}$
- The electron-volt is a more convenient (non-SI) unit of energy: $1\text{ eV} = 1.6 \times 10^{-19}\text{ J}$.
- The photoelectric effect is the emission of photoelectrons from a clean metal surface due to incident light whose frequency is greater than a threshold frequency, f_0.
  - If $f < f_0$, no electrons are released.
  - If $f > f_0$, the *rate* of electron release (the photocurrent) is proportional to the intensity of the light and occurs without any time delay.
- Only a small forward voltage is needed for the maximum photocurrent to be reached. As the forward voltage is increased, the photocurrent remains at this maximum.
- The maximum kinetic energy for the electrons $E_{k\,max}$ (i.e. that of the fastest electron) can be found by using a reverse voltage called the stopping voltage, V_0, where $E_{k\,max} = q_eV_0$.
- The work function, ϕ, for a metal is given by $\phi = hf_0$, and is different for each metal. If the frequency (energy) of the incident light is greater than the threshold frequency (energy), then a photoelectron will be ejected with some kinetic energy up to a maximum value.
- The maximum kinetic energy of the photoelectrons emitted from a metal is the energy of the photons minus the work function, ϕ, of the metal: $E_{k\,max} = hf - \phi$.
- A graph of $E_{k\,max}$ versus frequency will have a gradient equal to Planck's constant, h, and a y-intercept equal to the work function, ϕ. The x-intercept, where $E_{k\,max} = 0$, gives the threshold frequency, f_0.
- The wave approach to light could not explain various features of the photoelectric effect: the existence of a threshold frequency, the absence of a time delay when using very weak light sources, and increased intensity of light resulting in a greater rate of electron release rather than increased electron energy.
- Einstein used Planck's concept of a photon to explain the photoelectric effect, stating that each electron release was due to an interaction with only one photon.
- The photon model of light explained the existence of a threshold frequency for each metal, the absence of a time delay for the photocurrent even for weak light sources, and why brighter light resulted in a higher photocurrent.

7.4 Review *continued*

KEY QUESTIONS

1 Indicate whether each of the following statements is true or false.

a A blackbody is a body that is black in colour.

b A blackbody completely absorbs the incident radiation without emitting an electromagnetic wave.

c A blackbody absorbs and emits electromagnetic radiation.

d A blackbody emits electromagnetic waves of different wavelengths.

e The blackbody radiation spectrum depends on the type of material, shape and temperature.

2 What two fundamental assumptions did Planck have to make to explain blackbody radiation?

3 Calculate the energies (in joules) of the following wavelengths of light:

	Colour	Wavelength (nm)
a	red	656
b	yellow	589
c	blue	486
d	violet	397

4 When light shines on a metal surface, why might the metal become positively charged?

5 Which of the following statements about the photoelectric effect are true and which are false? For those that are false, rewrite them to make them correct.

a When the intensity of light shining on the surface of the metal increases, the photocurrent increases.

b When light sources of the same intensity but different frequencies are used, the higher frequency light has a higher stopping voltage and produces a higher maximum current than the lower frequency light.

c When the applied voltage is positive, photoelectrons are attracted to the anode or collector electrode.

6 Calculate the work functions (in electron-volts) of the following metals:

	Metal	Threshold frequency ($\times 10^{15}$ Hz)
a	lead	1.0
b	iron	1.1
c	platinum	1.5

7 In an experiment on the photoelectric effect, different frequencies of light were shone on a piece of magnesium with a work function of 3.66 eV. Identify which of the frequencies listed would be expected to produce photoelectrons.

A 3.0×10^{14} Hz

B 5.0×10^{14} Hz

C 7.0×10^{14} Hz

D 9.0×10^{14} Hz

8 Light with a frequency of 9.0×10^{14} Hz is shone onto a piece of magnesium with a work function of 3.66 eV. Calculate the maximum kinetic energy, in electron-volts, of the emitted photoelectrons.

9 Blue light with a wavelength of 475 nm is shone on a piece of sodium with a work function of 2.36 eV. Calculate the maximum kinetic energy, in electron-volts, of the emitted photoelectrons.

10 The metal caesium has a work function of 1.81 eV. Which of the following types of electromagnetic radiation would cause photoelectrons to be emitted?

A infrared radiation, $\lambda = 800$ nm

B red light, $\lambda = 700$ nm

C violet light, $\lambda = 400$ nm

D ultraviolet radiation, $\lambda = 300$ nm

11 Which of the following statements are true and which are false with respect to the value of the stopping voltage obtained when light is incident on a metal cathode? For those that are false, rewrite them to make them true.

a The stopping voltage indicates how much work must be done to stop the most energetic photoelectrons.

b The stopping voltage is reached when the photocurrent is reduced to almost zero.

c If only the intensity of the incident light is increased, the stopping voltage will not alter.

d For a given metal, the value of the stopping voltage is affected only by the frequency of the incident light.

12 Yellow-green light of wavelength 500 nm shines on a metal that has a stopping voltage of 0.80 V. Calculate the work function of the metal in electron-volts.

7.5 Atomic spectra

Recall from the previous sections that some light phenomena can be explained using a wave model and some can be explained using a particle (photon) model. This idea of wave–particle duality is counterintuitive and was not immediately accepted by most scientists, even after the groundbreaking work of Einstein, de Broglie and others, discussed fully in Section 7.6 'The quantum nature of light and matter'.

It was the work of Danish physicist Niels Bohr that finally convinced scientists that the particle model was required as part of a complete understanding of the nature of light. Bohr built on the work of Planck and Einstein to explain the emission and absorption spectra of hydrogen (Figure 7.5.1). This led to important discoveries in astronomy and, eventually, a reformulation of the understanding of the nature of energy and matter.

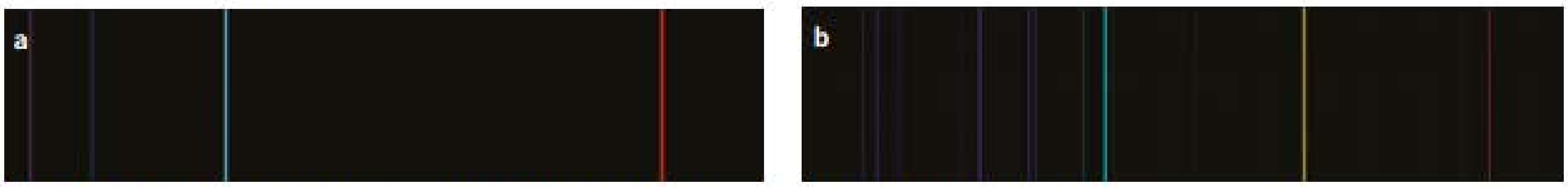


FIGURE 7.5.1 The spectral lines for hydrogen (a) and helium (b). The spectral lines of an element are unique to that element and are produced as electrons transition between different energy levels within the atom. Bohr's explanation of emission and absorption spectra was instrumental in furthering the understanding of the nature of energy and matter.

OBSERVING ABSORPTION SPECTRA

In 1814, the German physicist Joseph von Fraunhofer reported a number of dark lines appearing in the spectrum of sunlight (Figure 7.5.2).

You will recall that a spectrum showing all the colour components of white light can be obtained by passing sunlight through a prism. When Fraunhofer did this, he observed the spectrum (as expected), but he also noticed that there were some colours 'missing' from the spectrum. The missing colours appeared as black lines at various points along the spectrum. These apparently missing colours came to be known as Fraunhofer lines.

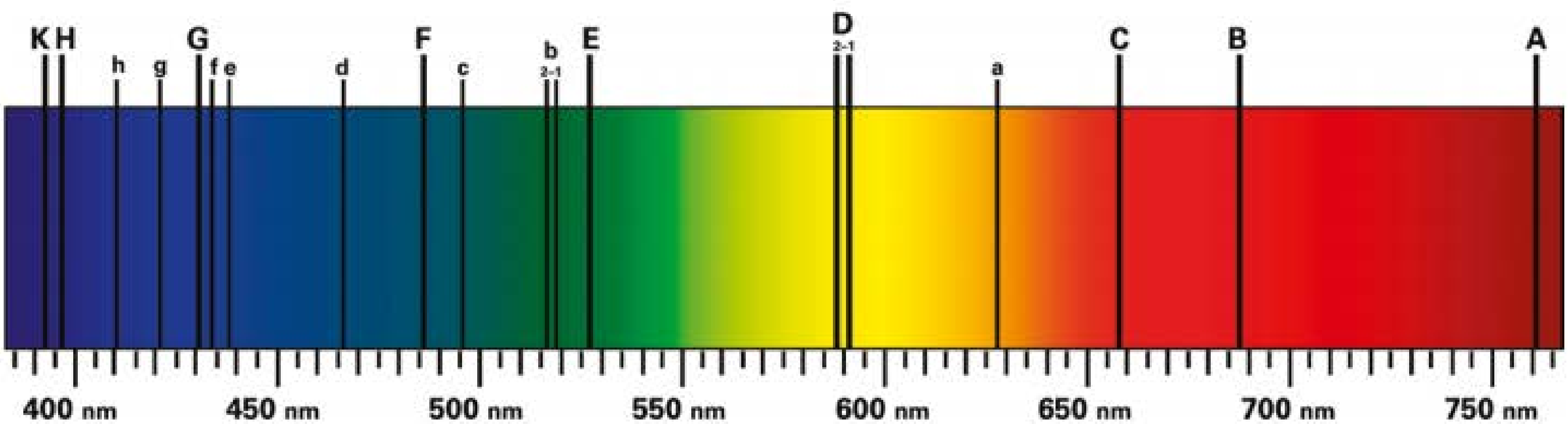


FIGURE 7.5.2 The spectrum of sunlight contains some missing colours known as Fraunhofer lines.

About 50 years later, scientists including Kirchhoff and Bunsen (of Bunsen burner fame) recognised that some of these lines corresponded to the colours emitted when certain gases were heated to high temperatures. They deduced that the dark lines were due to these colours (wavelengths or frequencies) being *absorbed* by gases as light made its way through the outer atmosphere of the Sun. This **absorption spectrum** allowed astronomers to determine that the Sun is largely composed of hydrogen with small quantities of helium and some other heavier elements.

Absorption spectra are valuable for scientists who wish to know what elements are present in a sample of gas or in a solution, so their use is not limited to astronomy. First, light is directed through a cool sample of a gas or through a solution containing an element or compound. Only certain wavelengths (or frequencies) of light will be absorbed by the elements present in the sample, which means that when the spectrum is viewed, this particular wavelength will be 'missing'. The wavelengths that are absorbed are unique to each type of atom. For this reason, by analysing which wavelengths are missing, scientists can determine exactly what elements are present in the sample.

OBSERVING EMISSION SPECTRA

FIGURE 7.5.3 The different metals used in fireworks are responsible for the colours in this display. For example, strontium gives red, sodium gives yellow and copper gives green.

When additional energy is applied to an element, the element produces discrete energies of light. Additional energy or excitation can be in the form of heating to high temperatures, passing an electric current through the element, bombarding it with electrons or photo excitation by light. Atoms within the material absorb energy and become 'excited' (more on what this means later in this section). This makes the atom unstable and eventually the electron will return to its 'unexcited' or **ground state**. When this happens, the energy that had been absorbed is released as a single photon. The colour of this photon will depend on the amount of energy it has.

Since atoms can usually have a number of different **excited states**, they can produce a number of different colours. The combination of colours produced by a particular element is distinctive to that element (Figure 7.5.3) and is known as its **emission spectrum** (shown in Figure 7.5.1, on page 271, for hydrogen and helium).

HYDROGEN'S ABSORPTION SPECTRUM

In the late 19th century, the emission and absorption spectra of hydrogen were of particular interest to scientists, as it had been recognised that lines in the absorption spectrum of hydrogen matched lines in the solar spectrum (Figure 7.5.4).

a

b

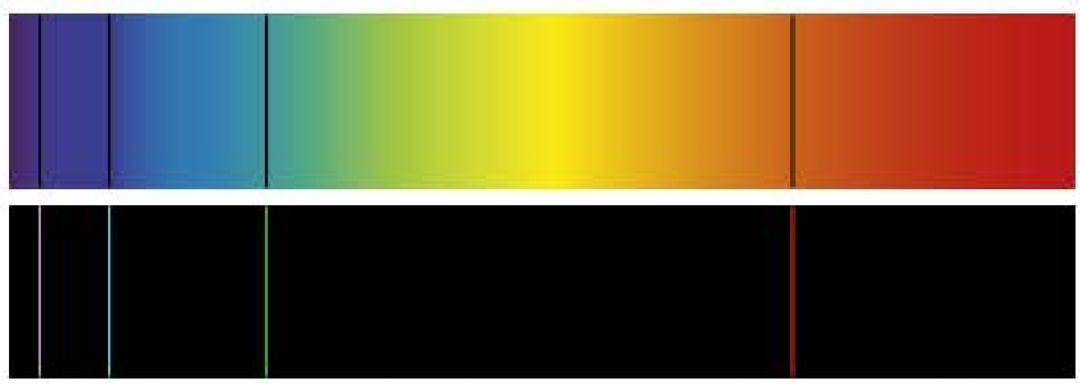


FIGURE 7.5.4 (a) In the absorption spectrum of hydrogen there is a background of continuous white light (broken into a spectrum of colours), with black lines that correspond to wavelengths of the radiation absorbed by the hydrogen atoms. (b) In the emission spectrum of hydrogen there is a black background against which lines corresponding to the wavelengths emitted by the hydrogen atoms can be observed.

Although some scientists were able to come up with an empirical (based on experimental data) formula that predicted the wavelength of the lines in the hydrogen spectra, no one was able to provide a theoretical explanation for the production of these lines using a wave model for light.

BOHR MODEL OF THE ATOM

In 1913, the Danish physicist Niels Bohr proposed an explanation for the absorption and emission spectra of hydrogen that drew on the quantum ideas proposed by Planck and Einstein, including Planck's quantum relation equation, $E = hf$. Bohr realised that:

- The absorption spectrum of hydrogen showed that the hydrogen atom was only capable of absorbing light of very specific frequencies or energies of light; that is, the absorbed energy was quantised.
- The emission spectrum of hydrogen showed that the hydrogen atom was also capable of emitting quanta of the exact energy value that it was able to absorb.
- If the frequency, and hence energy, of the incident light was below a certain value, the light would pass straight through hydrogen gas; no absorption would occur.

- Hydrogen atoms have an ionisation energy of 13.6 eV. Light of this energy or greater can remove an electron from a hydrogen atom, creating a positive ion.
- Photons of light with all energies above the ionisation value for hydrogen are continuously absorbed.

Bohr's explanation relied on a significant refinement of Rutherford's planetary model of the atom. He devised a sophisticated model of electron energy levels for atoms, a development for which he later won the Nobel Prize in Physics (Figure 7.5.5).

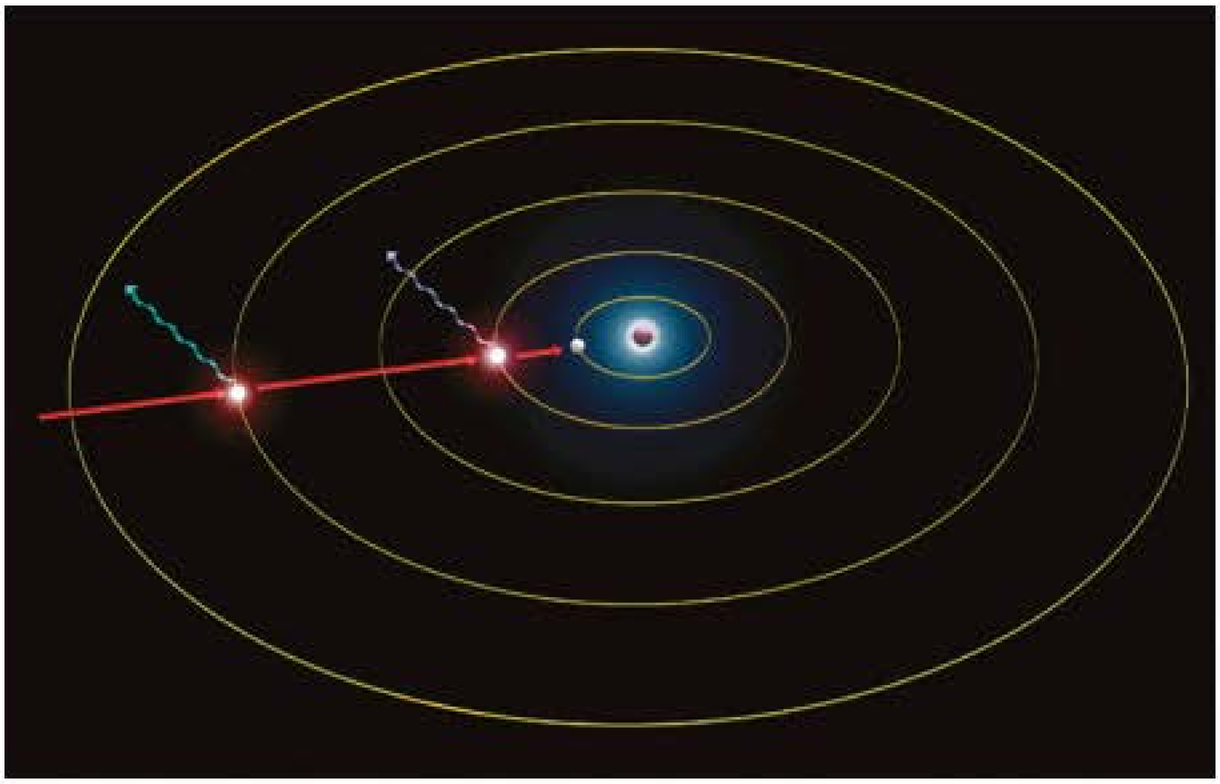


FIGURE 7.5.5 A diagram showing hydrogen spectrum emission levels based on the Bohr model of the atom. Electrons may only orbit in specific energy orbits, shown by the concentric circles. Electrons absorb energy to move to higher levels in their excited states and emit light in specific wavelengths characteristic of the element when returning to the ground state.

Bohr's model of the atom contained the following ideas:

- Electrons move in circular orbits around the nucleus of the hydrogen atom.
- The centripetal force keeping an electron in a circular orbit is provided by the electrostatic force of attraction between the positive nucleus and the negative electron.
- A number of allowable orbits of different radii exist for each atom (labelled $n = 1, 2, 3\ldots$ and known as the principal quantum number). Electrons may only occupy these orbits.
- An electron ordinarily occupies the lowest energy orbit available (i.e. the ground state).
- An electron does not radiate energy while it is in a stable orbit.

Bohr labelled the possible electron orbits for the hydrogen atom with a quantum number (n), and he was able to calculate the energy associated with each quantum number. The energy values are given by the following equation:

$$E_n = -\frac{13.6}{n^2}, \text{ where } E_n \text{ is in eV}$$

Using these energy levels and Planck's equation, Bohr could theoretically predict the wavelengths of all of the lines of the hydrogen emission spectrum:

- Electromagnetic radiation (in the form of photons) can be absorbed by an atom when the photon energy is *exactly equal* to the difference in energies between an occupied orbit and a higher orbit.
- Electromagnetic radiation is emitted by an excited atom when an electron returns from a higher energy orbit to a lower energy orbit. The photon energy will be *exactly equal* to the energy difference between the initial and final levels of the electron.

$$\Delta E = hf = \frac{hc}{\lambda} = E_m - E_n$$

where ΔE is the energy of the photon produced (J or eV)

h is Planck's constant (6.63×10^{-34} J s or 4.14×10^{-15} eV s)

f is the frequency of the photon (Hz)

c is the speed of light (3.00×10^8 m s^{-1})

λ is the wavelength of the photon (m)

E_n is the energy of the initial energy level with quantum number n

E_m is the energy of the final energy level with quantum number m

Figure 7.5.6 shows the energy levels for the hydrogen atom. These energies are expressed in terms of how strongly the electron is bound to the nucleus. Conventionally these levels are expressed as negative values. The ground level ($n = 1$) represents the orbit that is closest to the nucleus (i.e. the unexcited state). An electron in this orbit is the most tightly bound and has an energy of (–)13.6 eV, which means that it would need to gain 13.6 eV of energy for it to escape the atom. Higher energy levels represent orbits that are further from the nucleus. At higher levels the energy levels become closer together. A free electron (at $n = \infty$) must have zero potential energy, as it has escaped the electrostatic attraction of the proton in the nucleus. An electron that has escaped has gained kinetic energy and its energy value is positive.

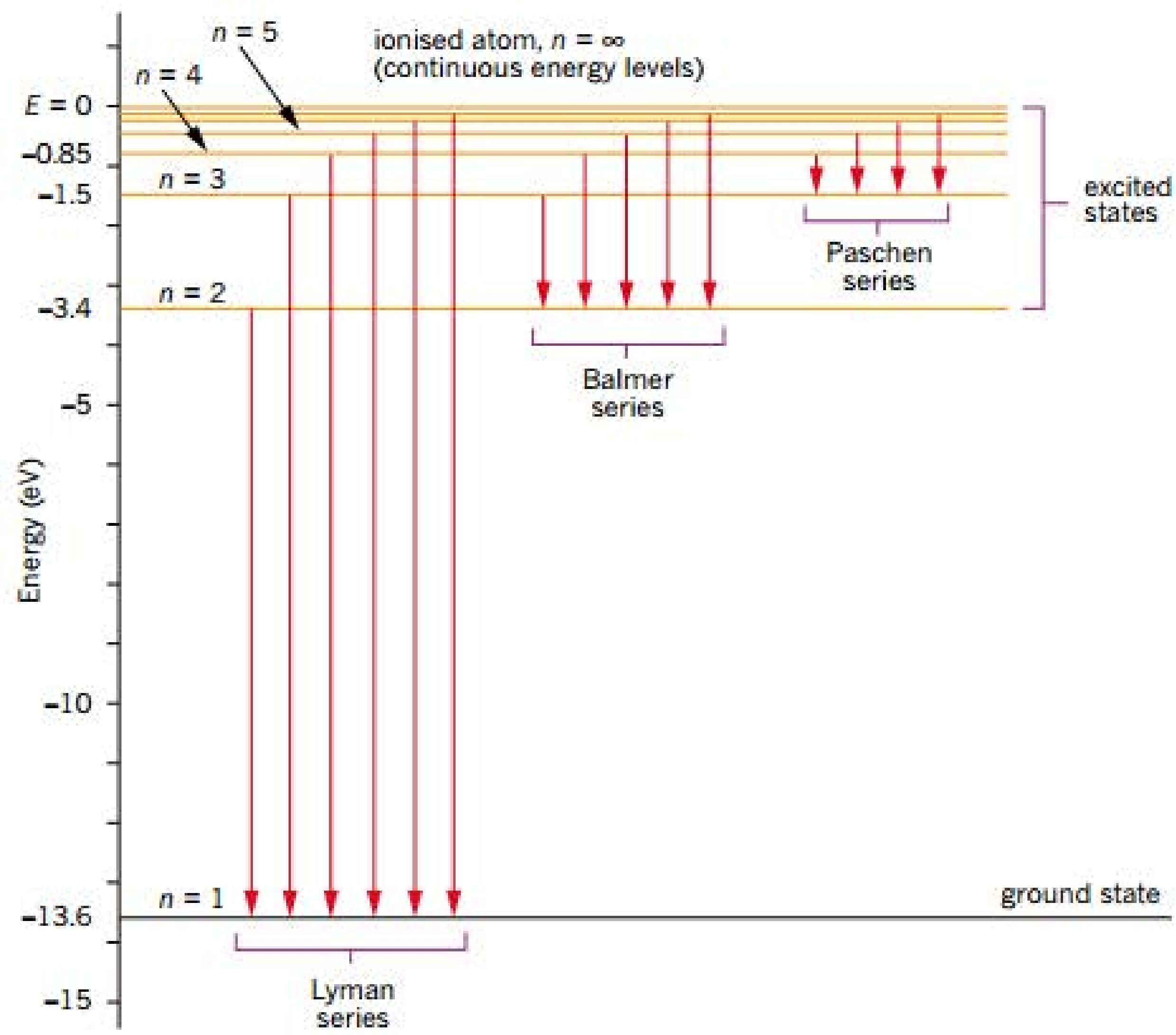


FIGURE 7.5.6 An energy level diagram for hydrogen. An electron in the ground state ($n = 1$) has an energy of (–)13.6 eV. For higher energy levels ($n > 1$), the energy levels can be seen to crowd together.

EMISSION SPECTRA

When a hydrogen atom gains energy, its electron moves from the ground state ($n = 1$) to one of the higher energy levels. This type of atom is described as 'excited'. Eventually, the electron will drop from the higher energy level to one of the lower levels and will emit a photon with energy equal to the *difference* in energy between the levels: $\Delta E = E_m - E_n$. As the electron falls back to its previous energy level, its energy value decreases; that is, it becomes a larger negative number.

Figure 7.5.6 also shows that the spectral lines of hydrogen can be explained in terms of electron transitions. The different spectral series shown on the diagram (Lyman, Balmer, Paschen) represent specific transitions. The Balmer series, for example, shows transitions back to $n = 2$ from various excited energy levels. These transitions represent wavelengths of the visible lines of the hydrogen emission spectrum.

Worked example 7.5.1

USING THE BOHR MODEL OF THE HYDROGEN ATOM

Calculate the wavelength (in nm) of the photon produced when an electron drops from the $n = 4$ energy level of the hydrogen atom to the $n = 2$ energy level. Identify the spectral series to which this line belongs.
Use Figure 7.5.6 to calculate your answer.

Thinking	Working
Identify the energy of the relevant energy levels of the hydrogen atom.	$E_4 = -0.85\,\text{eV}$ $E_2 = -3.40\,\text{eV}$
Calculate the change in energy.	$\Delta E = E_4 - E_2$ $= -0.85 - (-3.40)$ $= 2.55\,\text{eV}$
Calculate the wavelength of the photon with this amount of energy. Either use $h = \frac{6.63 \times 10^{-34}}{1.6 \times 10^{-19}}$ eV or convert ΔE in eV to joules i.e. $\Delta E = 2.55 \times 1.6 \times 10^{-19}$ J.	$\Delta E = \frac{hc}{\lambda}$ $\therefore \lambda = \frac{hc}{\Delta E}$ $= \frac{6.63 \times 10^{-34} \times 3.00 \times 10^{8}}{1.6 \times 10^{-19} \times 2.55}$ $= 4.87 \times 10^{-7}\,\text{m}$ $= 487\,\text{nm}$
Identify the spectral series to which this line belongs.	The electron drops down to the $n = 2$ energy level. Therefore, the photon must be in the Balmer series.

Worked example: Try yourself 7.5.1

USING THE BOHR MODEL OF THE HYDROGEN ATOM

Calculate the wavelength (in nm) of the photon produced when an electron drops from the $n = 3$ energy level of the hydrogen atom to the $n = 1$ energy level. Identify the spectral series to which this line belongs.
Use Figure 7.5.6 to calculate your answer.

EXTENSION

Balmer and Rydberg—empirical equations

In 1885, the Swiss mathematician Johann Balmer found an empirical equation that predicted the wavelength of the visible lines of the hydrogen emission spectrum:

$$\lambda = \frac{hm^2}{m^2 - n^2}$$

where λ is the wavelength of light (nm)

h is a constant with a value of 365 nm (not to be confused with Planck's constant)

$n = 2$

m could take values of 3, 4, 5 or 6.

When Balmer put $m = 7$ into the equation, it gave an answer of 397 nm, which corresponded to a spectral line that had been independently observed by Anders Angstrom. This set of spectral lines in the visible part of the electromagnetic spectrum came to be known as the Balmer series.

In 1888, Johannes Rydberg (Figure 7.5.7) realised that Balmer's formula was a special case of the more general formula:

$$\frac{1}{\lambda} = R_H\left(\frac{1}{n^2} - \frac{1}{m^2}\right)$$

where R_H is the Rydberg constant for hydrogen ($1.097 \times 10^7\,m^{-1}$) and n and m are any two integers, referring to energy levels, where $m > n$.

This equation predicted that there should be spectral lines in other parts of the electromagnetic spectrum. The ultraviolet series was later observed by Theodore Lyman, and two different infrared series were observed by Friedrich Paschen and Frederick Brackett.

FIGURE 7.5.7 Johannes Rydberg developed a general formula predicting the wavelengths of the lines of the emission spectrum of hydrogen.

ABSORPTION OF PHOTONS

The Bohr model also explains the *absorption* spectrum of hydrogen (Figure 7.5.4(a) on page 272).

You have already seen that the missing lines in absorption spectra correspond to the energies of light that a given atom is capable of absorbing. This is due to the energy differences between the atom's electron orbits, ΔE. Only incident light carrying just the right amount of energy to raise an electron to an allowed level can be absorbed.

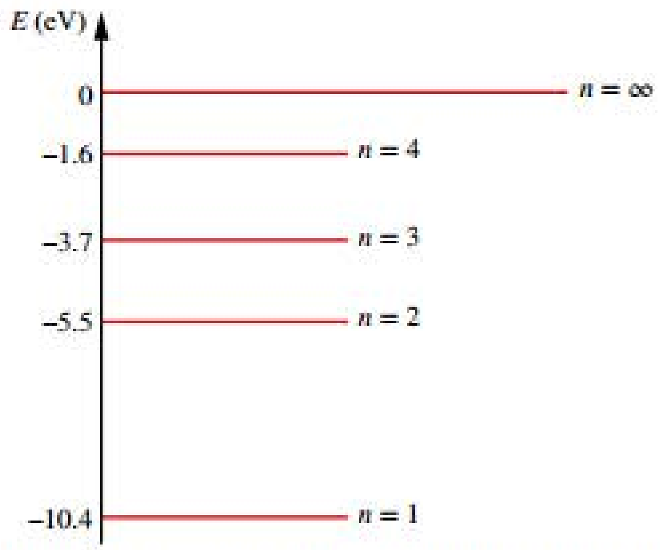
An electron ordinarily occupies the lowest energy orbit. Incident light that does not carry enough energy to raise an electron from this lowest energy level to the next level cannot be absorbed by the atom; the light would simply pass straight through. If the incident light is exactly equal to the energy difference ΔE, the electron will be promoted to the next level.

If the energy of the incident light is greater than the ionisation energy of an atom, then the excess energy provided by the photon will simply translate to extra kinetic energy for the released electron (recall the photoelectric effect from Section 7.4). For example, if a hydrogen atom absorbs a photon with 13.6 eV or more, the hydrogen atom is said to be 'ionised'.

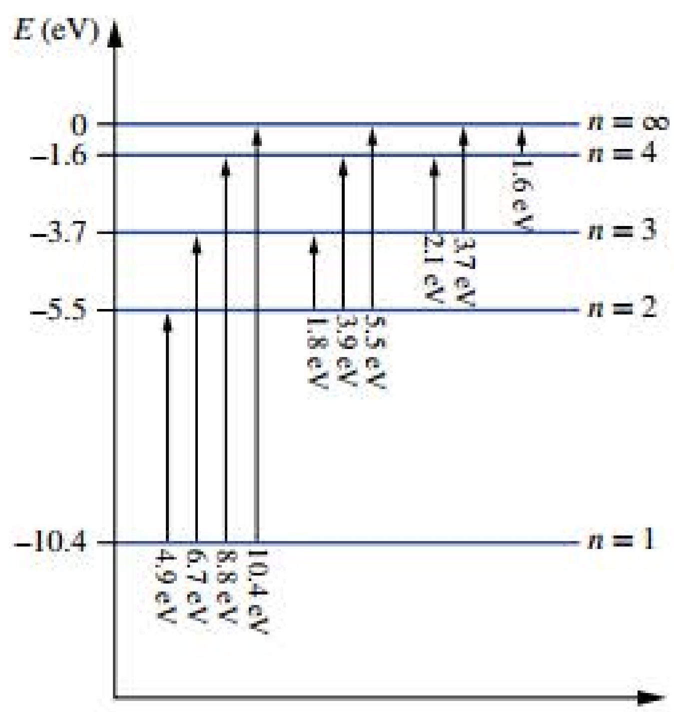
Worked example 7.5.2

ABSORPTION OF PHOTONS

Some of the energy levels for atomic mercury are shown in the diagram below.



Ultraviolet light with photon energies 4.9 eV, 5.0 eV and 10.5 eV is incident on some mercury gas. What could happen as a result of the incident light?

Thinking	Working
Check whether the energy of each photon corresponds to any differences between energy levels by determining the difference in energy between each level.	
Compare the energy of the photons with the energies determined in the previous step. Comment on the possible outcomes.	A photon of 4.9 eV corresponds to the energy required to promote an electron from the ground state to the first excited state ($n = 1$ to $n = 2$). The photon may be absorbed. A photon of 5.0 eV cannot be absorbed since there is no energy level above the ground state that corresponds exactly to 5.0 eV. A photon of 10.5 eV may ionise the mercury atom. The ejected electron will leave the atom with 0.1 eV of kinetic energy.

PHYSICSFILE

The special case of hydrogen

The hydrogen atom was a relatively simple place to begin the development of the field that would come to be known as 'quantum mechanics'. The hydrogen atom contains two charged particles—the positively charged nucleus (which usually contains just a single proton) and the electron (Figure 7.5.8). This means that only one electrical interaction (between the electron and the nucleus) needs to be considered.

In more complex atoms, such as helium, electrical interactions between the electrons are also significant. This makes the construction of mathematical models for these atoms vastly more complicated than for hydrogen.

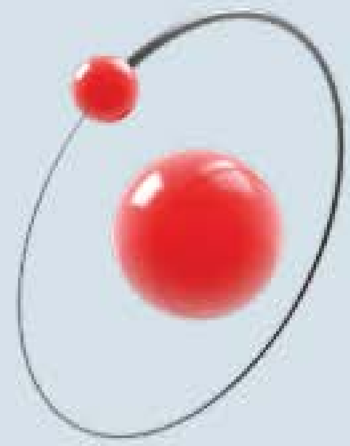
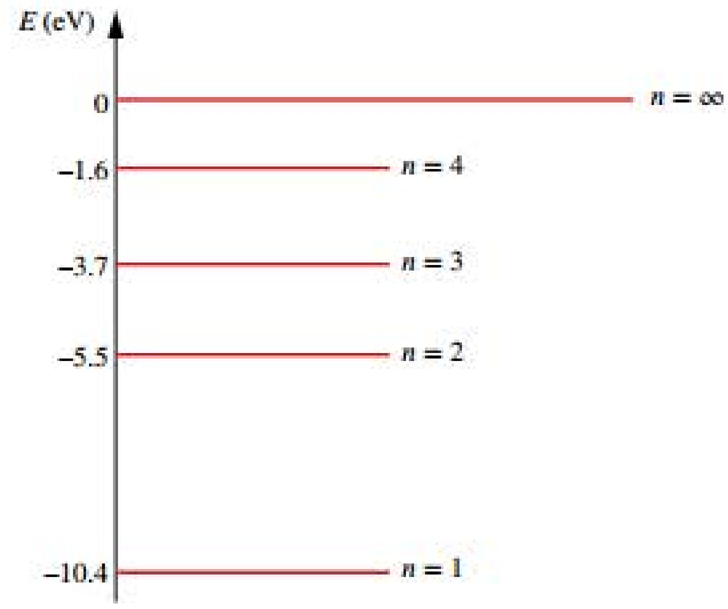


FIGURE 7.5.8 The hydrogen atom contains only two particles: the proton in the nucleus and the electron.

Worked example: Try yourself 7.5.2

ABSORPTION OF PHOTONS

Some of the energy levels for atomic mercury are shown in the diagram below.



Light with photon energies 6.7 eV, 9.0 eV and 11.0 eV is incident on some mercury gas. What could happen as a result of the incident light?

SPECTRAL ANALYSIS

In atomic emission spectroscopy, the chemical composition of a material can be determined by analysing the light that is emitted from the material, for example when it is burned or when an electric current is passed through a gas. The light can be separated into its component wavelengths by a spectroscope or diffraction grating, and the specific wavelengths found are characteristic of each particular element, very much like the way a fingerprint or DNA is used to identify an individual person. The interactions between the electrons alter the energy levels from those of hydrogen.

Metal vapour lamps produce light by exciting electrons in vaporised metal atoms, which then emit a photon as they return to their ground state. The emitted photons have wavelengths characteristic of the metal whose atoms are being excited in the lamp. A common type of metal vapour lamp is the sodium lamp. These are often used in street lighting and emit light of a distinctive yellow colour (Figure 7.5.9).

FIGURE 7.5.9 Sodium vapour lamps are sometimes used as street lights and have a distinctive yellow colour due to the yellow wavelengths of the sodium emission spectrum.

Problems with Bohr's model

Bohr's model of the hydrogen atom applied a quantum approach to the energy levels of atoms to explain a set of important, previously unexplained phenomena—the emission and absorption spectra of hydrogen. In principle, Bohr's work on the hydrogen atom could be extended to other atoms and, in 1914, the German scientists James Franck and Gustav Hertz demonstrated that mercury atoms contained energy levels similar to those of hydrogen atoms. Bohr's model signified an important conceptual breakthrough.

However, Bohr's model was limited in its application. It could only really be accurately applied to single-electron atoms—hydrogen and ionised helium. It modelled inner-shell electrons well but could not predict the higher-energy orbits of multi-electron atoms. Nor could it explain the discovery of the continuous spectrum emitted by solids. Further studies even showed problems with the emission spectrum of hydrogen. Some of the observed emission lines could be resolved into two very close spectral lines, and Bohr's model could not explain this. A more complex quantum approach was required.

Absorption with electron excitation

Recall from the previous sections that for an electron to be promoted to a higher level by a photon (absorption), the energy of the photon must exactly match the difference in energy of the two levels. However, if the atom is bombarded by electrons, then the bombarding electrons can lose energy to promote an electron within the atom to a higher energy level. For example, consider atomic mercury being bombarded by electrons of energy 5.7 eV. In Worked example 7.5.2, it was shown that this does not match any of the energy level differences exactly. However, the 5.7 eV bombarding electron can promote an $n = 2$ electron to $n = 4$ (an energy difference of 5.5 eV) by transferring its energy. The bombarding electron will then leave the atom with an energy of $5.7 - 5.5 = 0.2$ eV.

FIGURE 7.5.10 In an incandescent light globe, electricity is passed through a tungsten filament. As the filament heats up, the free electrons in the tungsten atoms collide, accelerate and emit photons. A wide range of photon wavelengths are emitted due to a wide range of different collisions (some weak, some strong).

COMPARING DIFFERENT LIGHT SOURCES

So far in this section, you have learnt that photons of light are emitted when electrons move from an excited state to ground state. But this is not the only way to produce light. An **incandescent** light bulb produces light by heating a filament to a very high temperature (Figure 7.5.10). The thermal motion of free electrons produces electromagnetic radiation with a range of wavelengths; that is, the light from the incandescent globe is a continuous spectrum. Some of the light produced is actually in the infrared part of the spectrum that is invisible to human beings (Figure 7.5.11).

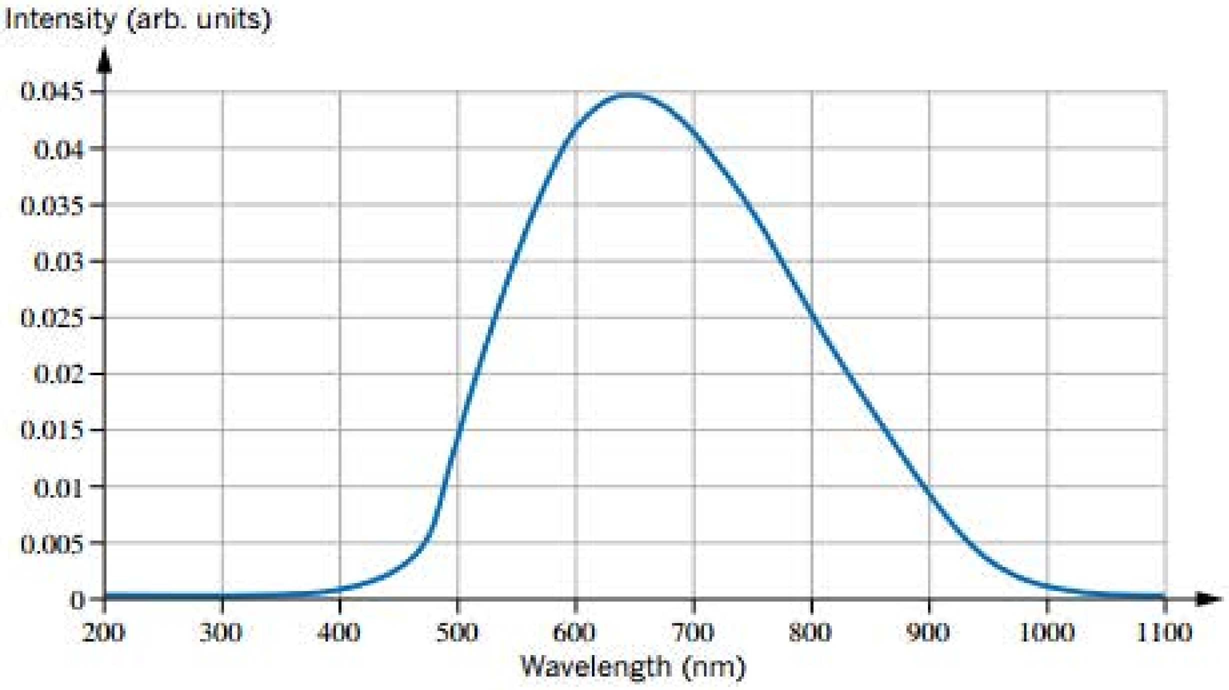


FIGURE 7.5.11 The spectrum of a 230 V, 60 W incandescent light globe. Since the visible spectrum is between about 400 nm and 750 nm, some of the radiation produced is not visible to humans.

PHYSICSFILE

Inefficiency

Incandescent light bulbs are widely seen as one of the most inefficient forms of lighting available. Traditional incandescent bulbs convert less than 10% of the electrical energy into light; the remainder is released as heat and other forms of long wavelength electromagnetic radiation. As a consequence, these types of light bulbs have almost been phased out and are being replaced by compact fluorescent lights (CFLs) and now white light emitting diodes (LEDs).

Other light sources emit light with different properties due to the different methods used to produce it.

DETERMINING THE COLOUR OF LEDS

A **light-emitting diode** (LED) is a semiconducting device that uses the excitation of electrons to produce light (Figure 7.5.12). LEDs are made from a variety of semiconductors, including GaAs, CdS, silicon and GaN-based materials. The addition of a small amount of another material, in a process known as 'doping', can change the conductive properties of semiconductors.

In a solid, the individual electron levels found in atoms merge together to form bands. In a semiconductor diode, most of the electrons sit in an energy band known as the *valence band*. At a slightly higher energy is another energy band known as the *conduction band*. If a small amount of electrical energy is provided as a potential difference, electrons can jump from the valence band to the conduction band and then move through the semiconductor as an electric current. When the electrons eventually drop back into the valence band, their energy is released as photons.

FIGURE 7.5.12 LEDs can be produced in virtually every visible and near-visible colour.

The difference in energy between the conduction band and the valence band level is known as the band gap, E_g. To turn on a basic LED, a potential difference or voltage needs to be applied. The threshold voltage, V_{th}, or potential difference required is approximately equivalent to the band gap, E_g, or difference in the two energy levels of the emitting photons. The colour or wavelength (λ) of the light produced is determined by the energy gap between the valence and conduction bands.

An approximate relationship can be given by

$$eV_{th} = \frac{hc}{\lambda} = hf = E_g$$

Measuring the threshold voltage for a series of LEDs can give an approximate measure of the band gap and emitting wavelength.

If the wavelength or frequency of the LED is known, then V_{th} can be plotted versus f. Comparing this to an equation of a straight line, $y = mx + c$, it can be seen that the slope of the line is $\frac{h}{e}$. Knowing e, the electron charge, a value for h can be determined.

This band gap can be altered by creating semiconductor alloys. For example, GaAs emits in the infrared, and AlGaAs increases the band gap and can be used for red LEDs. In addition, growing very thin layers of semiconductors can also alter the electron energy levels and properties to increase the efficiency of emission. This means that scientists and engineers can effectively 'tune' LEDs to produce photons of a particular wavelength, and hence to produce light of a particular colour.

Worked example 7.5.3

BAND GAP OF LEDS

A semiconductor LED is made of $In_{0.2}Ga_{0.8}N$. A student gradually increases the voltage and determines the threshold voltage for the LED to just turn on to be 2.5 V. Determine the approximate band gap of $In_{0.2}Ga_{0.8}N$ and the wavelength of emission. Use Table 7.1.3 on page 241 to determine what colour would this be.

Thinking	Working
The threshold voltage is related to the band gap by the relationship: $eV_{th} = E_g$	$V_{th} = 2.5\,V$ so $eV_{th} = 2.5\,eV$ and $E_g = 2.5\,eV$
Rearrange $\frac{hc}{\lambda} = E_g$.	$\lambda = \frac{hc}{E_g}$ $= \frac{6.63\times10^{-34}\times3\times10^{8}}{1.6\times10^{-19}\times2.5}$ $= 4.97 \times 10^{-9}$ $= 497\,nm$
Use Table 7.1.3.	497 nm would be green.

Worked example: Try yourself 7.5.3

BAND GAP OF LEDS

A GaN LED has a threshold voltage of 3.40 V. Determine the band gap, the wavelength of emission and the expected colour.

PHYSICS IN ACTION

Modern lighting

LEDs are gradually replacing most common uses of lighting, for example in cars, traffic lights and advertising signs. One of the newest forms of lighting on the market is white light LEDs. These use significantly lower currents than both tungsten filament lamps and incandescent lamps and tubes. In addition, they are very robust and typically have a lifetime of 20 000 hours. Since the cost of production has significantly reduced, these lights are rapidly replacing conventional sources.

At the time of writing, these white light LEDs are made of a semiconducting alloy, indium gallium nitride, which emits light in the blue or ultraviolet range and is coated with a phosphor. The emitted light is absorbed by the phosphor, which then emits white light, similar to the way a fluorescent tube works; a simplified diagram is shown in Figure 7.5.13.

In the future, it is expected that white light LEDs will be made of a composite of three different semiconducting structures that will emit in the red, blue and green and thus mix to produce white light. (Take a ×10 magnifying lens and look at a white part of your computer screen to see light mixing in action.)

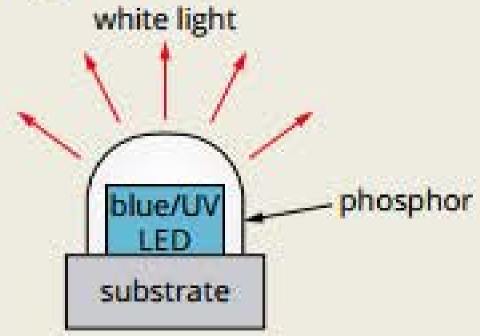


FIGURE 7.5.13 White light LEDs are currently made from an indium gallium nitride semiconductor alloy. The blue or ultraviolet light emitted interacts with a phosphor coating that radiates white light.

Lasers

The term '**laser**' is an acronym that stands for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation (Figure 7.5.14). Common gas lasers are the red HeNe laser and the blue argon ion laser.

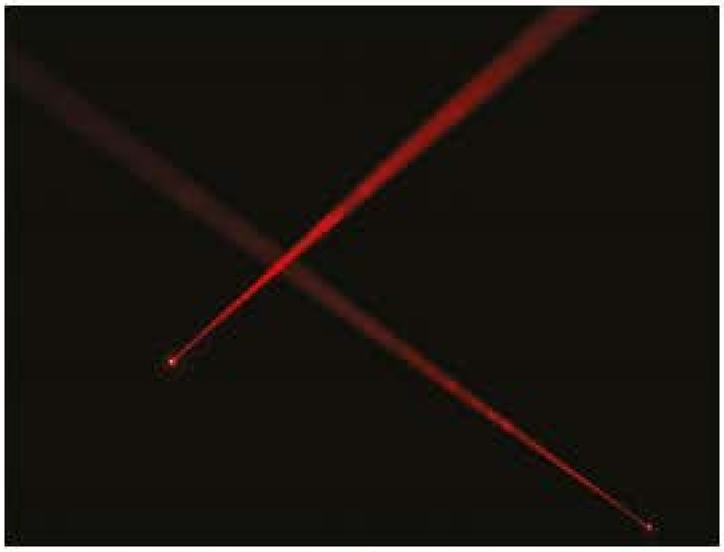


FIGURE 7.5.14 Laser light is coherent, polarised and monochromatic.

For lasing to occur, the gas atom needs to be in an excited state in which the higher level has more electrons than the lower level. This is known as a population inversion. The second condition requires two mirrors on either side of the chamber containing the gas. One of the mirrors must be partially silvered to allow some of the laser light to reflect and some light to escape. Some of the light that is emitted as electrons return from the higher energy level will be reflected back into the gas and stimulate further emission of the same wavelength. These emissions may also be reflected back, stimulating further emissions that reinforce the laser beam. This is known as stimulated emission.

The laser light is coherent, which means that all of the waves are in phase, i.e. their crests and troughs occur in time with each other. In addition, it is usually polarised and monochromatic or limited to a very narrow band of wavelengths.

Lasers the size of a computer chip can also be made from semiconductors, and are known as optoelectronic devices. They are of similar design to an LED but have a small-scale mirror structure incorporated in their fabrication. Lasers have a wide variety of applications in areas as diverse as communications, medicine and weaponry.

PHYSICS IN ACTION

National Broadband Network and optoelectronic communication

The National Broadband Network sends telephone and internet signals over a fibre-optic cable. A fibre-optic cable is usually made of glass of two different refractive indices. The light source is sent down the fibre-optic cable and is reflected off the walls due to total internal reflection. The light source is a semiconductor laser. A schematic of the process is shown in Figure 7.5.15. The laser light carrying the information is sent down the fibre-optic cable. At the end of the fibre-optic cable the light is detected by a semiconductor detector, converted to an electronic signal and then sent to your computer or device. The advantage of an optical signal is that it can carry a significantly larger amount of information, at a faster speed, than a copper electrical cable (e.g. ADSL).

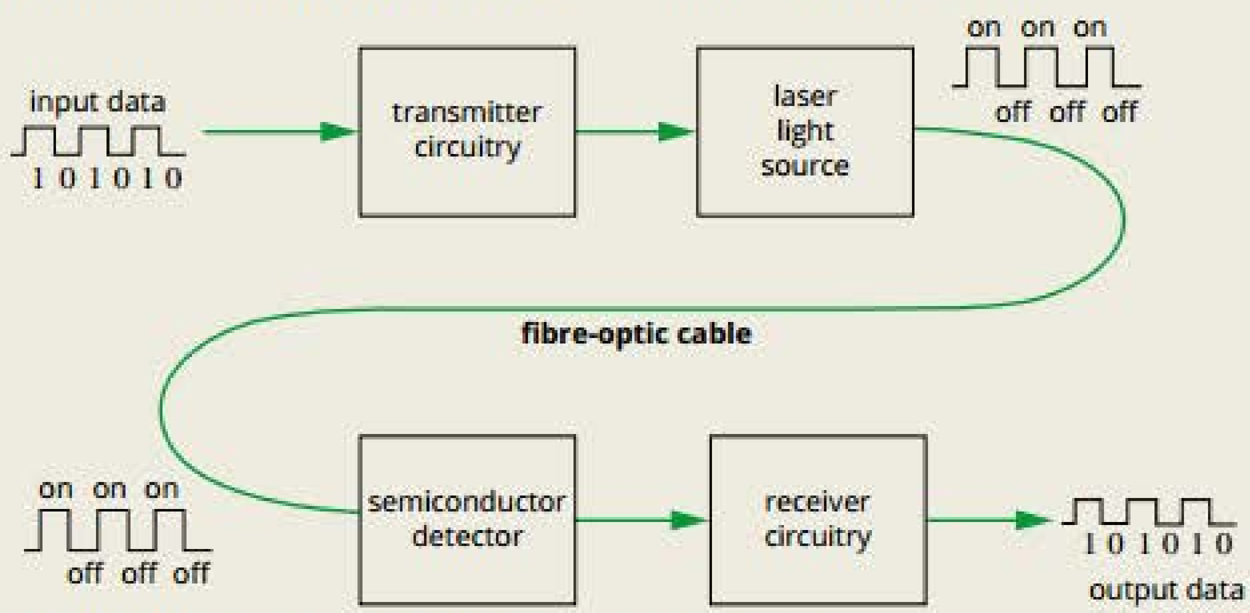


FIGURE 7.5.15 Schematic of a fibre-optic communication system.

SYNCHROTRON LIGHT

A **synchrotron** is a machine that uses powerful electric fields and magnets to accelerate charged particles, usually electrons, to velocities close to the speed of light (Figure 7.5.16). In Chapter 9 you will see that an electron in a magnetic field will undergo a circular motion.

Most synchrotrons are very large so that the electrons can be accelerated to the very high speeds required. For example, the Australian Synchrotron in Clayton, Victoria, has a circumference of over 200 m.

When electrons are accelerated, they produce electromagnetic radiation (light) known as synchrotron light. Because of the extremely high energies involved, the electromagnetic radiation produced by a synchrotron has a number of special properties. It is:

- extremely bright
- highly polarised
- emitted in very short pulses
- produced across a broad range of wavelengths from microwaves to gamma rays.

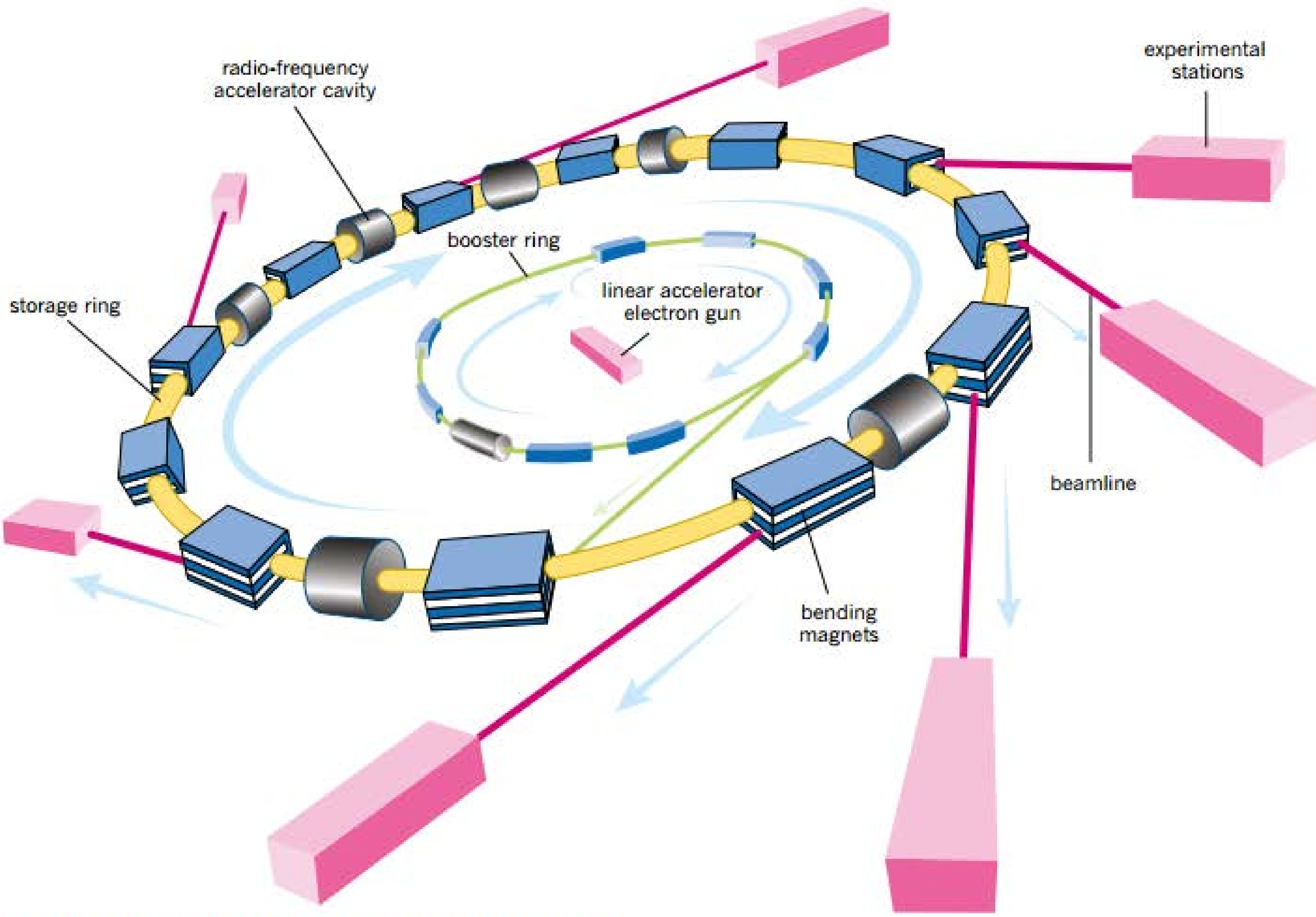


FIGURE 7.5.16 A synchrotron accelerates electrons to near light speed.

Synchrotron light has a wide range of scientific uses in medicine, bioscience, materials science and engineering. For example, synchrotron light is useful for exploring the structure of very small objects, smaller than those that can be seen with visible light. An ordinary light microscope is incapable of resolving many small structures due to the long wavelength of visible light. However, the short-wavelength X-rays in synchrotron light are ideal for examining structures at the cellular or atomic level, as they can resolve images down to the size of individual atoms (see Figures 7.5.17).

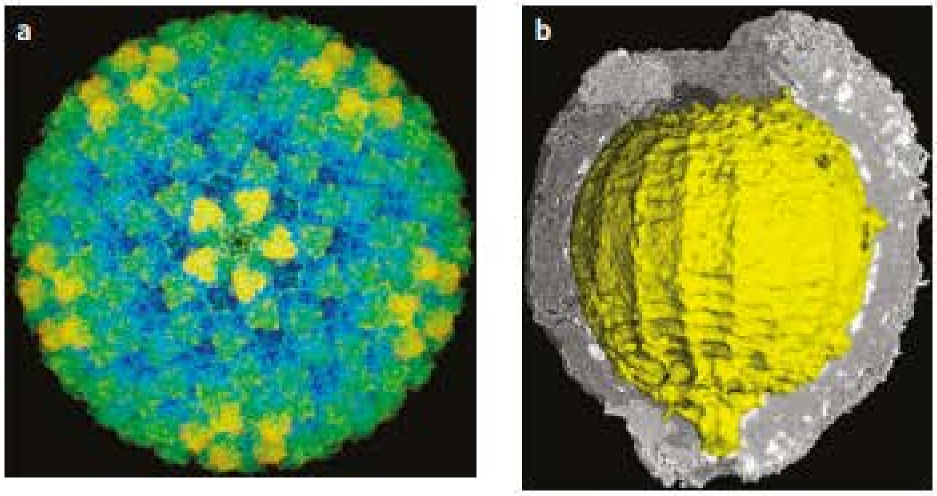


FIGURE 7.5.17 (a) Colour-enhanced image of the bluetongue virus obtained from the diffraction pattern of high-energy X-rays from synchrotron radiation. The virus gets its name for the blue tongue it causes in sheep. (b) Synchrotron light was used to map the internal structure of this fossilised alga without having to cut or break it open.

7.5 Review

SUMMARY

- The production of spectra suggests an internal structure to the atom. A line *emission* spectrum is produced by energised atoms. An *absorption* spectrum can be created when white light passes through a gas.
- The spectrum for any particular element is unique to that element.
- Bohr suggested that electrons in atoms orbit the nucleus in specially defined energy levels. No radiation is emitted or absorbed unless the electron can jump from one energy level to another. Electron energies are said to be *quantised*, since only certain values are allowed.
- The frequency, wavelength and energy of a photon emitted or absorbed by a hydrogen atom can be calculated from the difference between the energy levels involved: $\Delta E = E_2 - E_1 = hf = \frac{hc}{\lambda}$.
- The Bohr model of the atom is limited in its application, but was a significant development at the time, as it took a quantum approach to the energy levels of atoms and incorporated the quantum nature of electromagnetic radiation.
- In an incandescent lamp, the thermal motion of free electrons produces a continuous spectrum. Lasers, LEDs, metal-vapour lamps and synchrotrons are light sources that produce light at discrete frequencies via the emission of photons when excited electrons release energy. The means of excitation varies by source.
- Light-emitting diodes require a threshold voltage to just turn on or emit light. This threshold voltage is approximately equivalent to the band gap or difference in energy levels of the emitting photon.

KEY QUESTIONS

1 When does an element such as sodium produce an emission spectrum?

2 An emission line of frequency 6.0×10^{14} Hz is observed when looking at the emission spectrum of a particular elemental gas. What is the energy, in joules, of photons corresponding to this frequency?

3 Photons of energy 0.42 eV are emitted by a particular atom as it returns from the excited to the ground state. What is the corresponding wavelength of these photons?

4 What do the following acronyms stand for?

a LED

b LASER

5 At what wavelength (in nm) will an LED radiate if it is made from a material with an energy band gap of 1.84 eV?

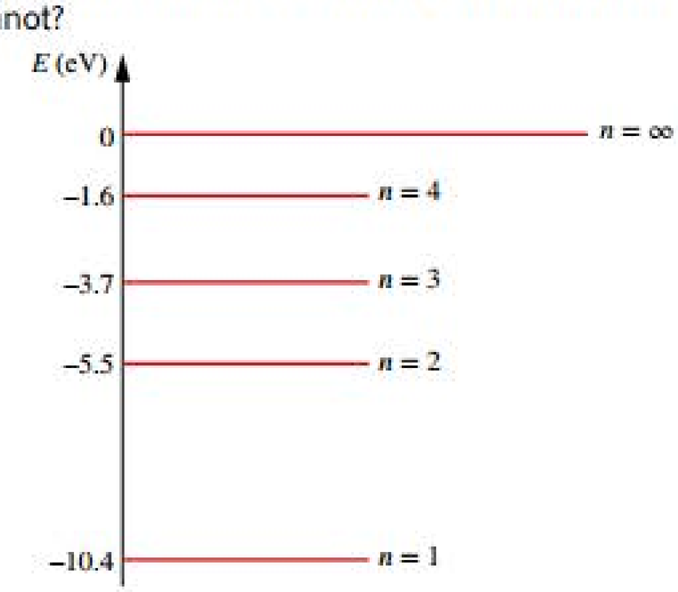
6 Calculate the energy of the photon required to move an electron in a hydrogen atom from its ground state ($n = 1$) to the $n = 4$ energy level. Refer to Figure 7.5.6, page 274.

7 Calculate the wavelength of the photons released in the transition described in Question 6.

8 Bohr's quantised model of the atom was a significant development; however, it was limited in application. What was the Bohr model unable to explain?

9 When an electron drops from the $n = 5$ energy level of the hydrogen atom to the $n = 2$ energy level, a 434 nm photon is released. If the $n = 2$ orbit has an energy of −3.4 eV, what is the energy of the $n = 5$ orbit?

10 Some of the energy levels for atomic mercury are shown in the diagram below. Why can a beam of electrons of 7 eV fired at mercury promote an electron from the $n = 1$ to $n = 3$ level but a photon of 7 eV cannot?



7.6 The quantum nature of light and matter

In order to explain the photoelectric effect, Einstein used the photon concept that Planck had developed. However, like many great discoveries in science, the development of the quantum model of light raised almost as many questions as it answered. It had already been well established that a wave model was needed to explain phenomena such as diffraction and interference. How could these two contradictory models be reconciled to form a comprehensive theory of light?

Answering this question was one of the great scientific achievements of the 20th century and led to the extension of the quantum model to matter as well as energy. It led to a fundamental shift in the way the universe is viewed. Some of the great scientists of that time are shown in the historic photograph in Figure 7.6.1.

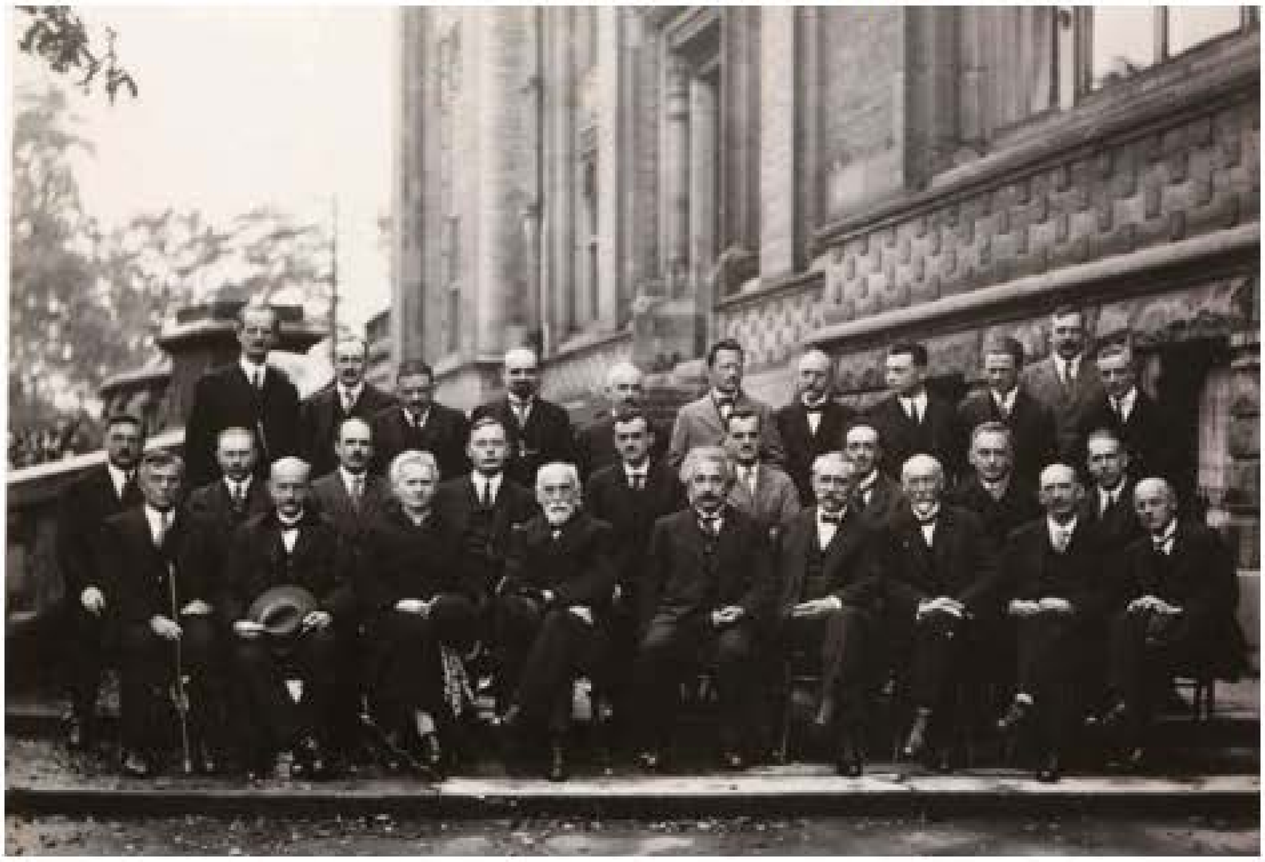


FIGURE 7.6.1 This photo shows the 5th Solvay conference in Brussels in 1927, which was attended by great scientists including Albert Einstein, Max Planck, Niels Bohr, Marie Curie, Paul Dirac, Erwin Schrödinger and Louis de Broglie. All of these scientists contributed to the current knowledge of the universe, the atom and quantum mechanics.

WAVE–PARTICLE DUALITY

In many ways, the wave and particle models for light seem fundamentally incompatible. Waves are continuous and are described in terms of wavelength and frequency. Particles are discrete and are described by physical dimensions such as their mass and radius.

In order to understand how these two sets of ideas can be used together, it is important to remember that scientists describe the universe using models. Models are analogies that are used to illustrate certain aspects of reality that might not be immediately apparent.

Physicists have come to accept that light is not easily compared to any other physical phenomenon. In some situations, light has similar properties to a wave; in other situations, light behaves more like a particle. This understanding is called **wave–particle duality** (Figure 7.6.2). In the century since Einstein did his work establishing quantum theory, many experiments have supported this duality and no scientist has come up with a better explanation (yet).

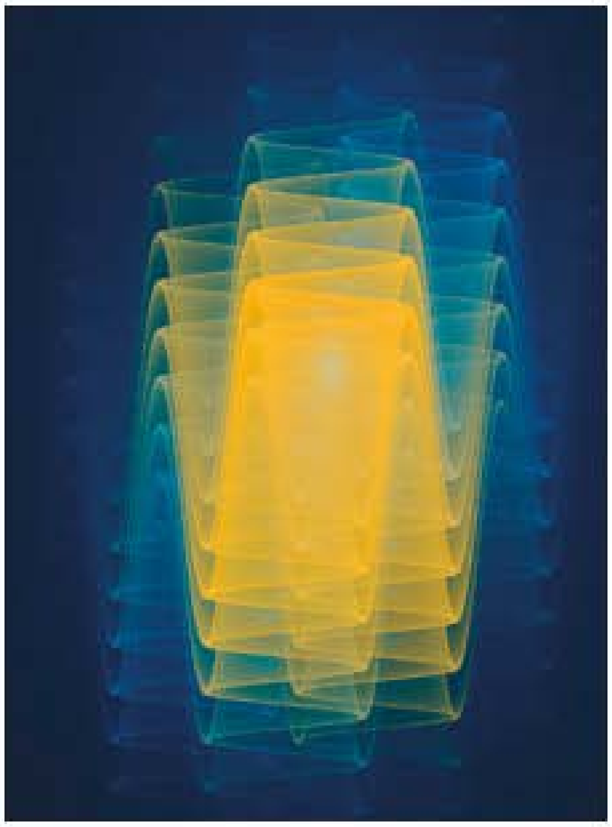


FIGURE 7.6.2 An artist's attempt to represent wave–particle duality.

Experimental evidence for the dual nature of light

In the early years of quantum theory, some scientists believed that the wave properties of light observed in Young's double-slit experiment might have been due to some sort of interaction between photons as they passed through the slits together.

To test this, experiments were done with light sources that were so dim that scientists were confident that only one photon was passing through the apparatus at a time. In this way, any interactions between photons could be eliminated. Over time, these experiments produced identical interference patterns to those done with bright sources (Figure 7.6.3), thus demonstrating the dual nature of light.

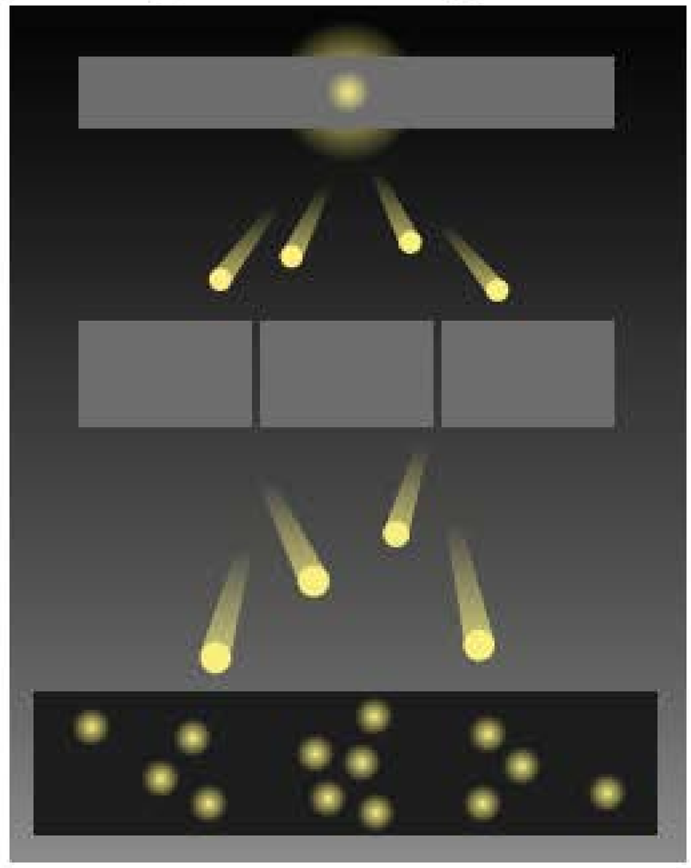


FIGURE 7.6.3 An interference pattern can be built up over time by a series of single photons passing through an apparatus like that used in Young's experiment, demonstrating the wave–particle duality of light.

Interestingly, when a detector is used to determine which slit the photon passes through, the wave pattern disappears and the photon acts like a particle. The act of trying to directly observe the photon changes the observation.

De Broglie's wave–particle theory

In 1924, the French physicist Louis de Broglie proposed a groundbreaking theory. He suggested that since light (which had long been considered to be a wave) sometimes demonstrated particle-like properties, then perhaps matter (which was considered to be made up of particles) might sometimes demonstrate wave-like properties.

He quantified this theory by predicting that the wavelength of a particle would be given by the equation:

$$\lambda = \frac{h}{p}$$

where λ is the wavelength of the particle (m)

p is the momentum of the particle ($kg\,m\,s^{-1}$)

h is Planck's constant

This is also commonly written as:

$$\lambda = \frac{h}{mv}$$

where m is the mass of the particle (kg)

v is the velocity of the particle ($m\,s^{-1}$)

The wavelength that de Broglie described, λ, is referred to as the **de Broglie wavelength** of matter.

Worked example 7.6.1

CALCULATING THE DE BROGLIE WAVELENGTH

Electrons in a famous experiment known as the Davisson–Germer experiment travelled at about $4.0 \times 10^6\,\mathrm{m\,s^{-1}}$. Calculate the de Broglie wavelength of these electrons given that the mass of an electron is 9.11×10^{-31} kg.	
Thinking	**Working**
Recall de Broglie's equation.	$\lambda = \frac{h}{mv}$
Substitute the appropriate values into the equation and solve it.	$\lambda = \frac{h}{mv}$ $= \frac{6.63 \times 10^{-34}}{9.11 \times 10^{-31} \times 4 \times 10^6}$ $= 1.8 \times 10^{-10}$ m or 0.18 nm

Worked example: Try yourself 7.6.1

CALCULATING THE DE BROGLIE WAVELENGTH

Calculate the de Broglie wavelength of a proton travelling at $7.0 \times 10^5\,\mathrm{m\,s^{-1}}$. The mass of a proton is 1.67×10^{-27} kg.

Worked example 7.6.2

CALCULATING THE DE BROGLIE WAVELENGTH OF A MACROSCOPIC OBJECT

Calculate the wavelength of a cricket ball of mass m = 160 g travelling at $150\,\mathrm{km\,h^{-1}}$.	
Thinking	**Working**
Convert mass and velocity to SI units.	m = 160 g = 0.16 kg $v = 150 \div 3.6 = 42\,\mathrm{m\,s^{-1}}$
Recall de Broglie's equation.	$\lambda = \frac{h}{mv}$
Substitute the appropriate values into the equation and solve it.	$\lambda = \frac{h}{mv}$ $= \frac{6.63 \times 10^{-34}}{0.16 \times 42}$ $= 9.9 \times 10^{-35}$ m

Worked example: Try yourself 7.6.2

CALCULATING THE DE BROGLIE WAVELENGTH OF A MACROSCOPIC OBJECT

Calculate the de Broglie wavelength of a person with a mass of 66 kg running at $36\,\mathrm{km\,h^{-1}}$.

PHYSICSFILE

Louis Victor Pierre Raymond de Broglie (1892–1987)

Louis de Broglie (Figure 7.6.4) was a French physicist. In 1924 he wrote a doctoral thesis entitled *Recherches sur la théorie des quanta* (Research on quantum theory), in which he presented his theory of the wave properties of particles—the de Broglie wave theory, based on the works of Einstein and Planck on wave–particle duality. Later, de Broglie developed his thesis and formulated the final de Broglie hypothesis. In 1929 he was awarded the Nobel Prize for his research. By applying de Broglie's theory it was possible, for example, to construct an electron microscope.

FIGURE 7.6.4 Louis de Broglie

It can be seen from Worked examples 7.6.1 and 7.6.2 that the de Broglie wavelength of an electron is smaller than that of visible light, but is still large enough to be measurable. However, the wavelength of an everyday object such as a cricket ball is extremely small (9.9×10^{-35} m). Hence, you will never notice the wave properties of everyday objects. To illustrate this, consider the observable wave behaviour of diffraction. Recall that for diffraction to be noticeable, the size of the wavelength needs to be comparable to the size of the gap or obstacle. Therefore for an everyday object, with its tiny wavelength, to produce a noticeable diffraction, it would need to pass through a gap much smaller than a fraction of a proton diameter!

ELECTRON DIFFRACTION PATTERNS

De Broglie's prediction that matter could exhibit wave-like behaviour was controversial. However, it was experimentally confirmed by the American scientists Clinton Davisson and Lester Germer in 1927, when they observed the diffraction patterns produced when they bombarded the surface of a piece of nickel with electrons (Figure 7.6.5).

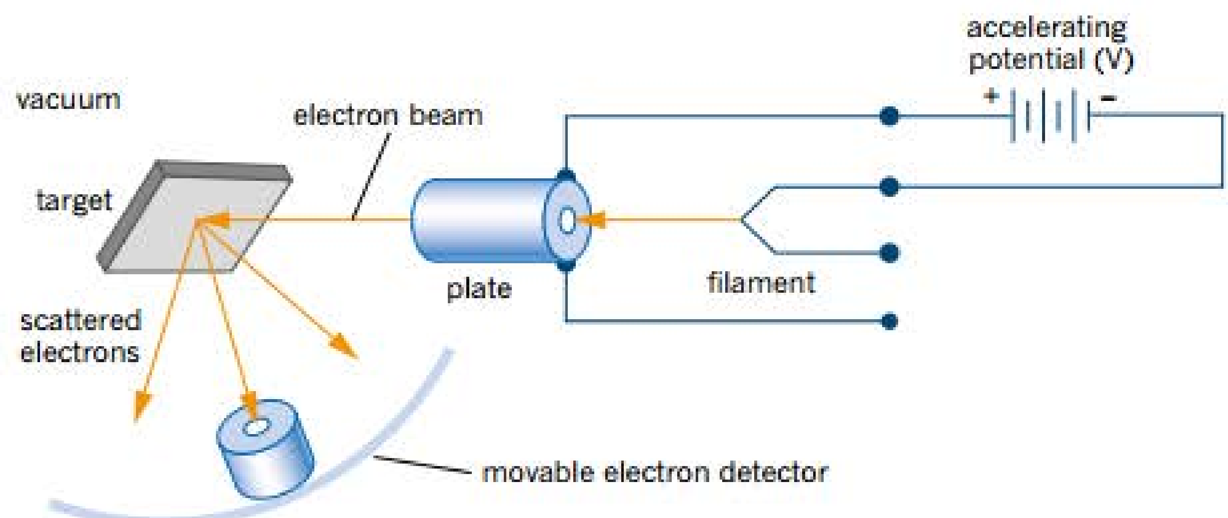


FIGURE 7.6.5 The Davisson and Germer apparatus to show electron scattering.

They used an electron 'gun', which provided a beam of electrons. The speed of the electrons was known because they had been accelerated through a known voltage. The detector could be swung around on an axis so that it could intercept electrons scattered from the nickel target in any direction in the plane shown.

Davisson and Germer found that as they moved their detector through the different scattering angles, they encountered a sequence of maximum and minimum intensities (Figure 7.6.6).

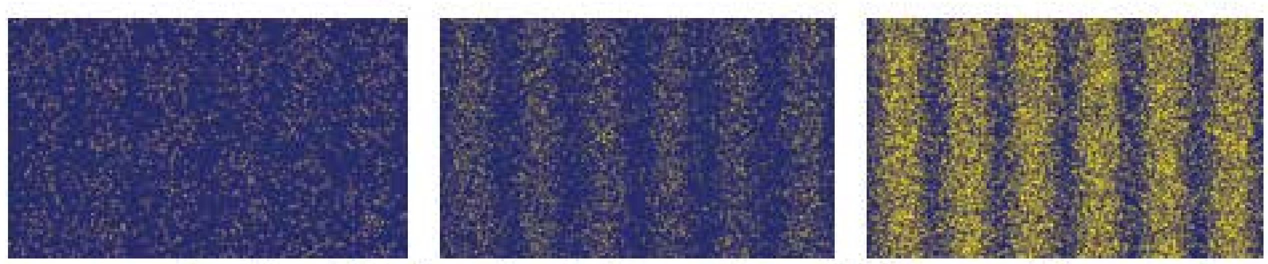


FIGURE 7.6.6 An electron diffraction pattern like the one observed by Davisson and Germer can be built up over time from repeated observations.

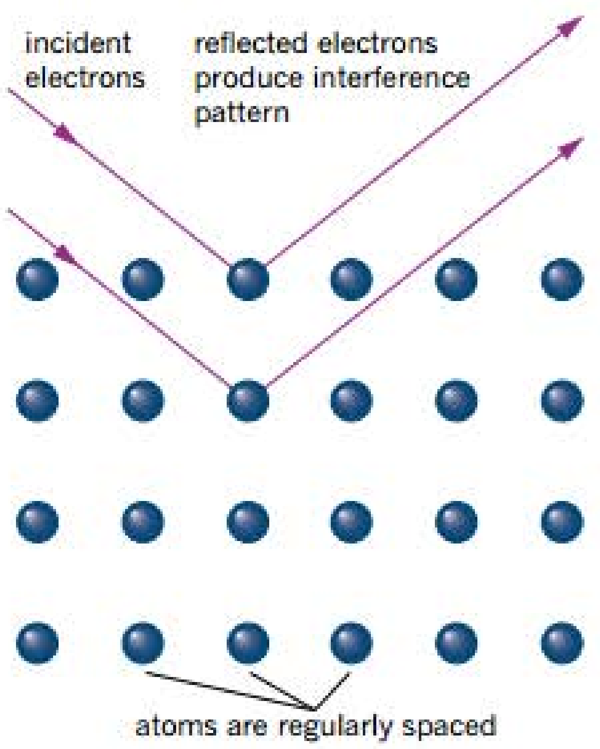


FIGURE 7.6.7 Electrons reflecting from different layers within the crystal structure create an interference pattern like those produced by a diffraction grating. If the path difference between the two rays is equal to a multiple number of wavelengths, constructive interference occurs; destructive interference occurs if the path difference is a multiple number plus half a wavelength.

Clearly, the electrons were being scattered by the different layers within the crystal lattice (Figure 7.6.7) and were undergoing interference. When Davisson and Germer analysed the diffraction pattern to determine the wavelength of the 'electron waves', they calculated a value of 0.14 nm, which was consistent with de Broglie's hypothesis.

Worked example 7.6.3

WAVELENGTH OF ELECTRONS FROM AN ELECTRON GUN

Find the de Broglie wavelength of an electron that has been accelerated from rest through a potential difference of 75 V. The mass of an electron is 9.11×10^{-31} kg and the magnitude of the charge on an electron is 1.6×10^{-19} C.

Thinking	Working
Calculate the kinetic energy of the electron from the work done on it by the electric potential. Recall from earlier chapters that $W = qV$.	$W = qV$ $= 1.6 \times 10^{-19} \times 75$ $= 1.2 \times 10^{-17}$ J
Calculate the velocity of the electron.	$E_k = \frac{1}{2}mv^2$ $v = \sqrt{\frac{2E_k}{m}}$ $= \sqrt{\frac{2 \times 1.2 \times 10^{-17}}{9.11 \times 10^{-31}}}$ $= 5.1 \times 10^6$ m s^{-1}
Use de Broglie's equation to calculate the wavelength of the electron.	$\lambda = \frac{h}{mv}$ $= \frac{6.63 \times 10^{-34}}{9.11 \times 10^{-31} \times 5.1 \times 10^6}$ $= 1.4 \times 10^{-10}$ m $= 0.14$ nm

Worked example: Try yourself 7.6.3

WAVELENGTH OF ELECTRONS FROM AN ELECTRON GUN

Find the de Broglie wavelength of an electron that has been accelerated from rest through a potential difference of 50 V. The mass of an electron is 9.11×10^{-31} kg and the magnitude of the charge on an electron is 1.6×10^{-19} C.

Comparing the wavelengths of photons and electrons

In the same year that Davisson and Germer conducted their experiment, other supporting evidence came from G.P. Thomson (son of J.J. Thomson, discoverer of the electron). Rather than scatter an electron beam from a crystal, Thomson produced a diffraction pattern by passing a beam of electrons through a tiny crystal. Thomson then repeated his experiment, using X-rays of the same wavelength in place of the electrons. The X-ray diffraction pattern was almost identical to the one made with electrons, as shown in Figure 7.6.8.

As the diffraction patterns obtained for the X-ray photons and electrons were the same, and as both were passed through the same 'gaps' to obtain these diffraction patterns, then an important conclusion could be made. The electrons must have a wavelength similar to that of the X-rays. Since their wavelengths are similar, the momentum of the electrons and the X-ray photons must also be comparable (but not their speeds).

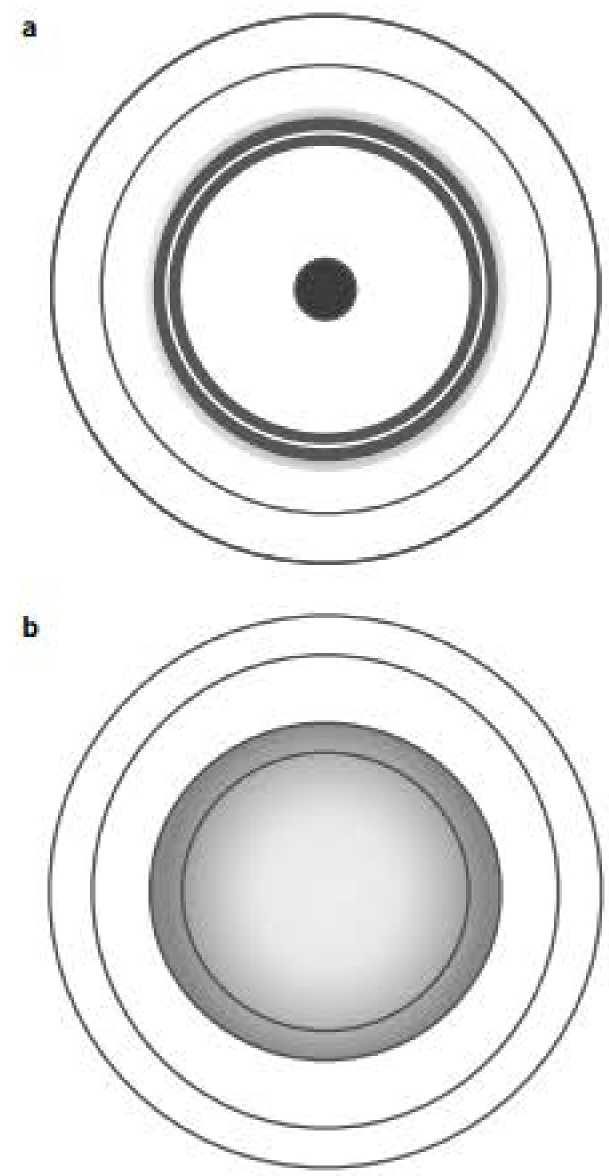


FIGURE 7.6.8 These diffraction patterns were taken by using (a) X-rays and (b) a beam of electrons with the same target crystal. Their similarity suggests a wave-like behaviour for the electrons and an electron de Broglie wavelength similar to that of X-rays.

EXTENSION

Electron microscopes

The discovery of the wave properties of electrons had an important practical application in the invention of the electron microscope. Just as an optical microscope makes use of the wave properties of photons to magnify tiny objects, so too can the wave properties of electrons be used to create magnified images (Figure 7.6.9).

One of the limitations of an optical microscope is that it can only create a clear image of structures that are similar in size to the wavelength of the light being used. This is because the light diffracts around these structures. So a light microscope is only useful for seeing things down to about 390 nm, the lower wavelength end of the visible light spectrum.

However, the wavelength of a beam of electrons is significantly smaller than the wavelength of a beam of visible light. For example, an electron microscope with electrons at an energy of 200 keV has a de Broglie wavelength of 2.51 pm (1 pm = 10^{-12} m). Due to lensing restrictions, the practical resolution limit for an electron microscope is 0.1 nm (1 nm = 10^{-9} m). Therefore, electron microscopes can create images with much finer detail than optical microscopes.

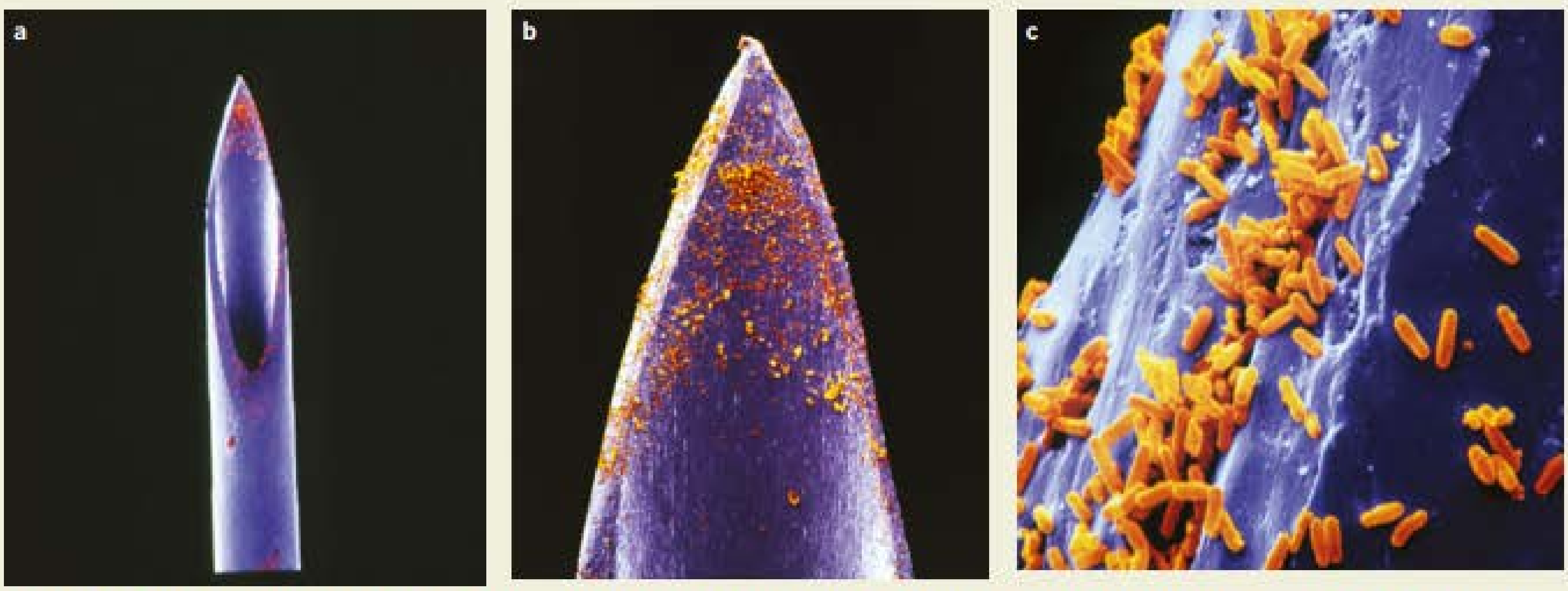


FIGURE 7.6.9 Images formed by an electron microscope: rod-shaped bacteria (orange) clustered on the point of a syringe used to administer injections. The magnifications are (a) ×9, (b) ×36 and (c) ×560 at 35 mm size.

YOUNG'S DOUBLE-SLIT EXPERIMENT USING ELECTRONS

In 1961, Claus Jönson at the University of Tübingen, Germany, performed the double-slit experiment with a beam of electrons. The result was an interference pattern that was the same as that obtained for photons, illustrating that electrons can behave as a wave. In 1965 physicist Richard Feynmann introduced the idea of single electrons producing interference patterns. This was demonstrated experimentally in 1974 by Guillo Pozzi and colleagues at the University of Bologna, Italy. Thus, a single electron interacts with itself, as does a single photon.

PHOTON MOMENTUM

An interesting corollary of de Broglie's hypothesis $\lambda = \frac{h}{p}$ is that if a particle such as an electron has a wavelength, λ, then a photon must have a momentum, p. This is quite counterintuitive, since photons do not have mass and all photons travel at the speed of light, c. Nevertheless, de Broglie's equation allows the momentum of a photon to be calculated.

Worked example 7.6.4

CALCULATING PHOTON MOMENTUM

Calculate the momentum of a photon of red light with a wavelength of 650 nm.	
Thinking	**Working**
Convert 650 nm to m.	$650\,\text{nm} = 650 \times 10^{-9}\,\text{m}$
Transpose de Broglie's equation to make momentum the subject.	$\lambda = \frac{h}{p}$ $p = \frac{h}{\lambda}$
Substitute in the appropriate values and solve for p.	$p = \frac{h}{\lambda}$ $= \frac{6.63 \times 10^{-34}}{650 \times 10^{-9}}$ $= 1.02 \times 10^{-27}\,\text{kg}\,\text{m}\,\text{s}^{-1}$

Worked example: Try yourself 7.6.4

CALCULATING PHOTON MOMENTUM

Calculate the momentum of a photon of blue light with a wavelength of 450 nm.

Clearly, the momentum of a single photon is tiny, which is why you will not feel any physical 'pressure' when light falls on you. However, it is possible to measure 'light pressure' using very sensitive equipment.

PHYSICS IN ACTION

Solar sailing

In interplanetary space, where other forces such as friction are negligible, light pressure can actually be used as a form of propulsion. Spacecraft such as the Mariner 10 and MESSENGER spacecraft, which both flew past Mercury and Venus, used deceleration caused by solar pressure to conserve fuel.

More recently, the Japanese Aerospace Exploration Agency launched IKAROS (Interplanetary Kite-craft Accelerated by Radiation of the Sun). IKAROS (Figure 7.6.10) is the first spacecraft to draw its primary propulsion from a *solar sail*. A traditional sail propels a ship using the change of momentum that occurs when air molecules bounce off it. Similarly, a solar sail gains propulsion from changes in photon momentum as light is reflected from it. The IKAROS spacecraft has a $196\,\text{m}^2$ reflective sail that produces a thrust of 1.12 mN.

FIGURE 7.6.10 The IKAROS spacecraft is the first interplanetary spacecraft to use solar-sail technology.

STANDING WAVES AND THE DUAL NATURE OF MATTER

Earlier in this section it was shown that small particles moving at very high speeds can be thought of as matter waves. Wave behaviour can be used to indicate the probability of the path of a particle. If particles can be thought of as matter waves, then these matter waves must be able to maintain steady energy values if the particles are to be considered stable.

De Broglie, the scientist who proposed the idea of matter having wavelengths, applied his approach to the discussion of Bohr's model for the hydrogen atom. He viewed the electrons orbiting the hydrogen nucleus as matter waves. He suggested that the electron could only maintain a steady energy level if it established a **standing wave**.

De Broglie reasoned that if an electron of mass m were moving with speed v in an orbit with radius r, this orbit would be stable if it matched the condition

$$mvr = n\frac{h}{2\pi}$$

where n is an integer.

This can be rearranged to

$$2\pi r = n\frac{h}{mv}$$

Since $2\pi r$ is the circumference, C, of a circle, and the de Broglie equation is $\lambda = \frac{h}{mv}$, this equation can be rewritten as $C = n\lambda$.

In other words:

> **The stable orbits of the hydrogen atom are those for which the circumference is exactly equal to a whole number of electron wavelengths.**

This can be visualised by imagining a conventional standing-wave pattern, like that of a vibrating string discussed in *Pearson Physics 11 Western Australia*, being looped around on itself in three dimensions as shown in Figure 7.6.11.

a

b

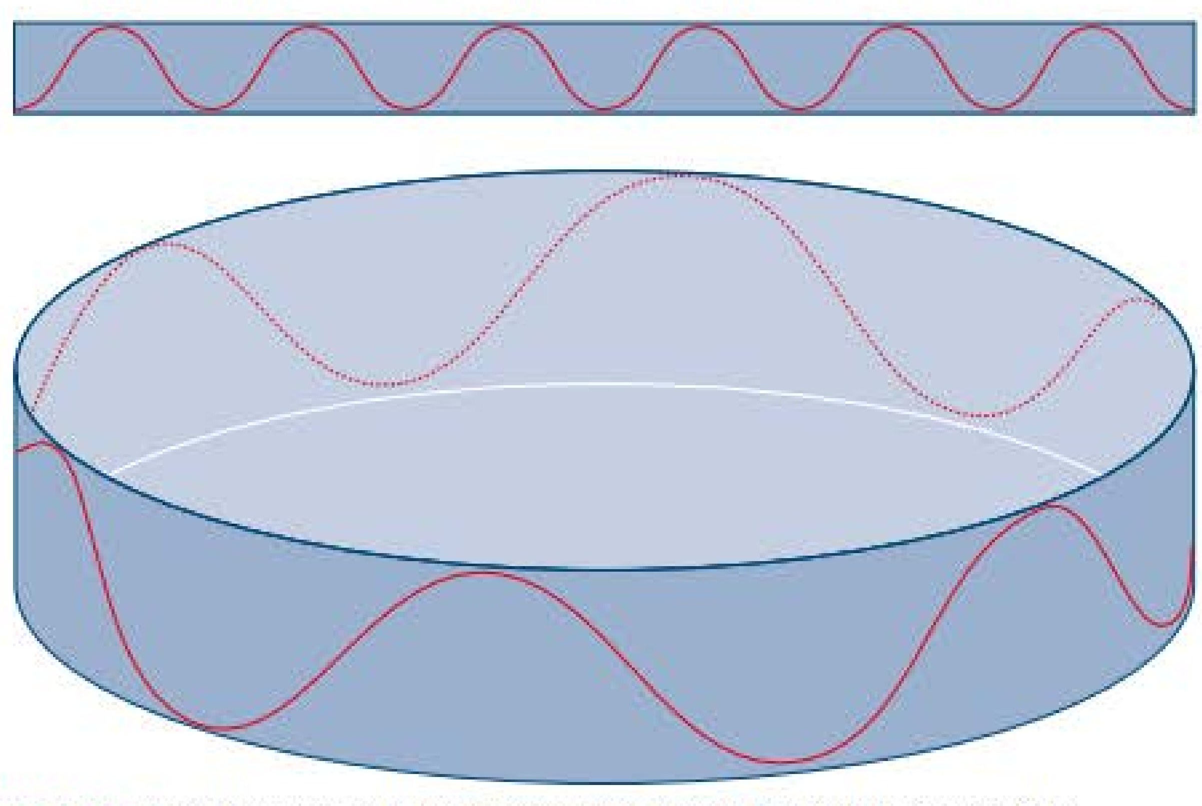


FIGURE 7.6.11 A standing wave pattern (a) can be looped around on itself to form (b), if the circumference of the circle is equal to a whole number of wavelengths.

If the circumference of the circle is not equal to a whole number of wavelengths, then destructive interference occurs, a standing wave pattern cannot be established and the orbit cannot represent an energy level (Figure 7.6.12).

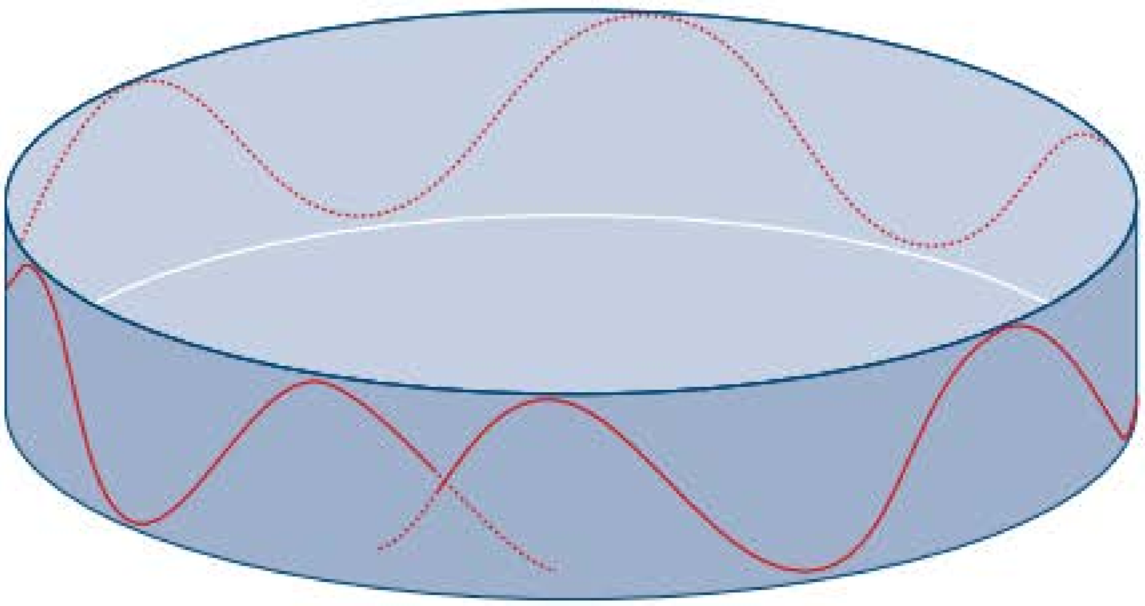


FIGURE 7.6.12 A circular standing wave pattern cannot be established if the circumference of the circle is not equal to a whole number of wavelengths.

A QUANTUM INTERPRETATION OF THE ELECTRON

In the early 20th century, scientists struggled to interpret the evidence of the dual wave–particle nature of energy and matter. Waves and particles are fundamentally different—waves are extended and continuous whereas particles are discrete. How two such different models could be combined to describe the fundamental building blocks of nature was a serious puzzle. As scientists delved into this mystery, they discovered fundamental limitations to their ability to explore the 'quantum' universe.

In 1925, the Austrian physicist Erwin Schrödinger built on the work of Niels Bohr by developing a mathematical equation that could describe the wave behaviour of electrons in situations other than the simple hydrogen atom (Figure 7.6.13).

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi$$

FIGURE 7.6.13 Schrödinger's wave equation. While outside the scope of this course, it is interesting to see what Schrödinger's equation looks like. The Greek symbol Ψ (psi) represents the wave function of the electron.

Quantum mechanics is the name now given to the area of physics in which the wave properties of electrons and other subatomic particles are studied and modelled. In Schrödinger's model, the wave properties of electrons are interpreted as representing the probability of finding an electron in a certain location. Schrödinger's equation has been used to calculate the regions of space in which an electron can be found in a hydrogen atom. These are now known as 'orbitals' or 'electron clouds' rather than orbits because they are complex three-dimensional shapes, as shown in Figure 7.6.14. Schrodinger's equation is successfully used to model significantly more complex atoms, molecules and solid structures.

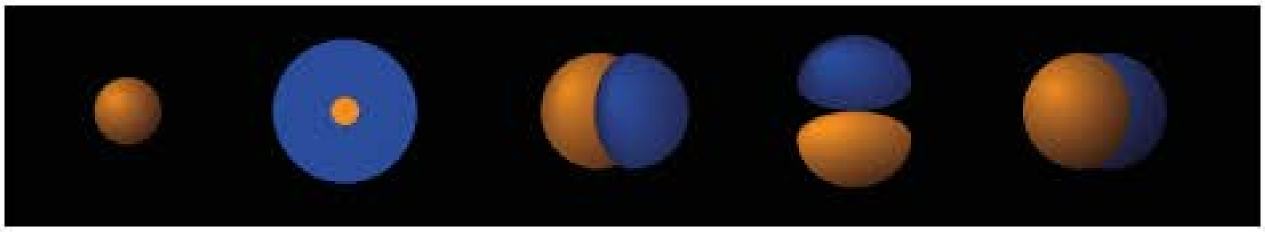


FIGURE 7.6.14 The shapes of the first five electron orbitals of a hydrogen atom.

LIMITS TO MODELS AT VERY SMALL SCALES

It was becoming clear to scientists that the nature of the universe at the very smallest scale is fundamentally different from the way the universe is perceived at the macroscopic scale.

In everyday life, each object has a clearly definable position and motion. The classical laws of physics, developed by scientists from Newton through to Maxwell, are all based on this assumption, which is so fundamental to human experience that it is hard to imagine a universe where this is not the case. However, there is no particular reason why tiny particles such as electrons and photons should be similar to larger objects such as balls or planets; scientists initially just extrapolated from their experience until the evidence showed that their assumptions were wrong.

PHYSICSFILE

Schrödinger's cat

Schrödinger described the strange, counterintuitive nature of quantum mechanical systems using a now-famous analogy known as Schrödinger's cat.

This is a thought experiment (i.e. Schrödinger did not actually perform the experiment) in which a cat is placed in a closed box with a flask of poison. A quantum mechanical system is set up such that there is a 50% chance of the flask being broken and the cat killed.

Schrödinger argued that until the box is opened to reveal the outcome of the experiment, the cat is considered as simultaneously alive *and* dead.

In a manner similar to the dual nature of light, the outcome (for the cat being alive or dead; for light being a wave or a particle) does not exist until an observation or measurement is made.

EXTENSION

Heisenberg's uncertainty principle

There are some quantities that are impossible to measure precisely. The first scientist to clearly identify this was the German physicist Werner Heisenberg (Figure 7.6.15). Heisenberg's uncertainty principle describes a limit to which some quantities can be measured.

For example, imagine you are trying to measure the exact location of an electron (Figure 7.6.16). The only way to do this would be to hit the electron with another particle such as a photon of light. However, as soon as the photon strikes the electron, it would cause the electron to move. In this way, the act of measurement causes a change in the value being measured. This is a general problem when trying to measure the location or motion of subatomic particles.

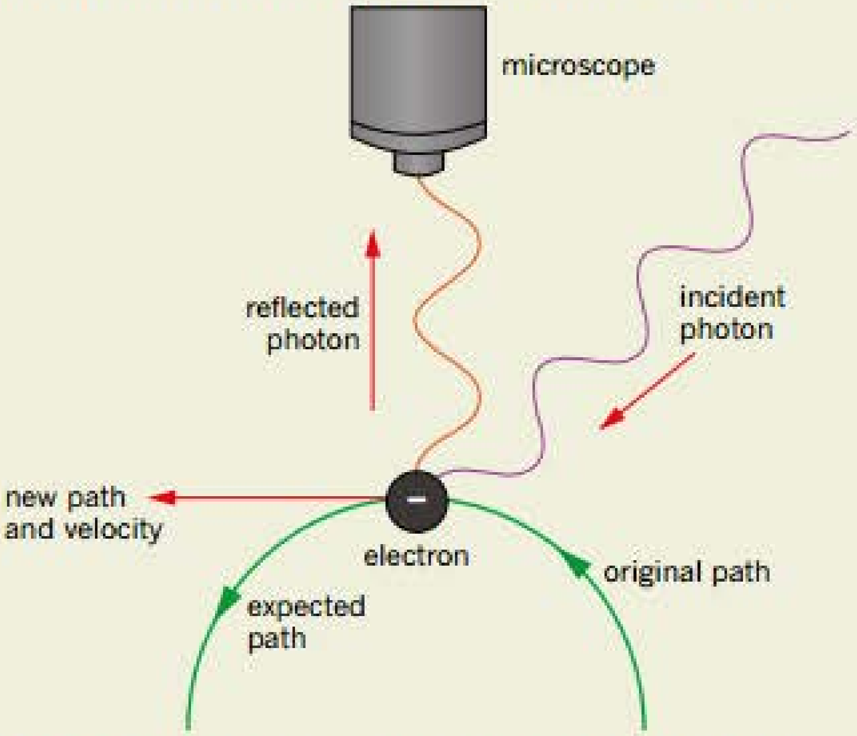


FIGURE 7.6.16 A thought experiment considering how an electron could be observed. The reflection of the photon needed to observe the electron introduces uncertainty in the position of the electron, making it unobservable.

Heisenberg described this problem with a formula that states that the product of the uncertainties in the position (Δx) and momentum (Δp) of a particle must always be greater than a certain minimum value related to Planck's constant (h):

$$\Delta x \Delta p \geq \frac{h}{4\pi}$$

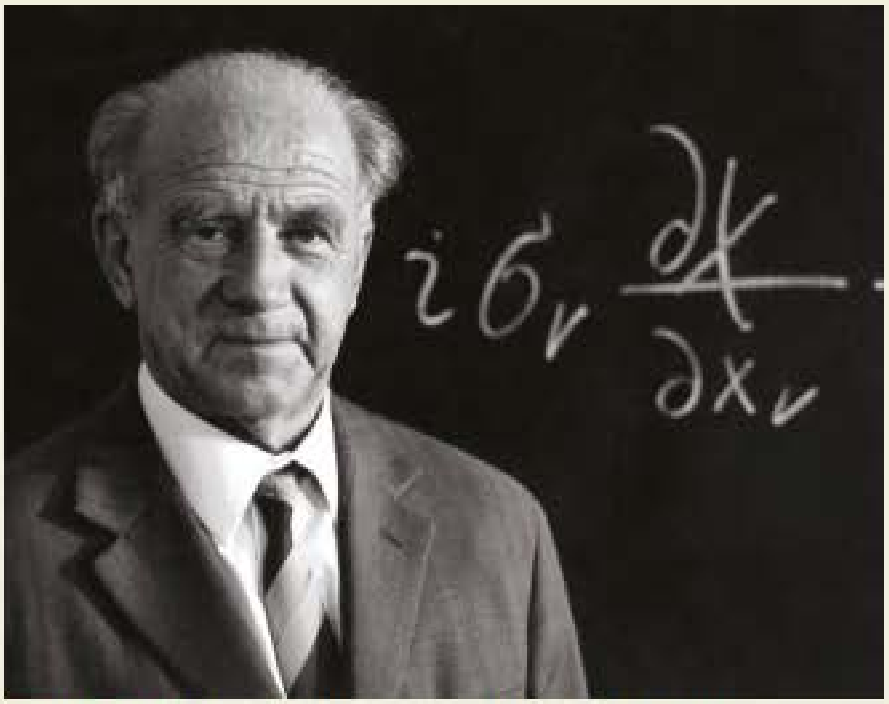


FIGURE 7.6.15 Werner Heisenberg (1901–76) won the Nobel Prize in Physics in 1932 for his work on the uncertainty principle. Heisenberg is regarded as a founder of quantum physics.

According to Heisenberg's uncertainty principle, the more exactly the position of a subatomic particle is known, the less is known about its momentum. Similarly, the more precisely the momentum of a particle is measured, the less certain is its exact position.

It can also be expressed as $\Delta E \Delta t \geq \frac{h}{4\pi}$, where ΔE is the uncertainty in energy and Δt the uncertainty in time

For the normal-sized world around us, the inclusion of Planck's constant, h, in the measure of uncertainty means that the level of uncertainty in determining the position of everyday objects is extremely small—in fact, virtually insignificant.

However, at an atomic scale, this level of uncertainty is substantial. And since everyday objects are made up of atoms containing subatomic particles such as electrons, the basic understanding of matter comes down to this fundamental property of all quantum mechanical systems.

Single-slit diffraction and the uncertainty principle

An experiment that can be used to illustrate Heisenberg's uncertainty principle is the single-slit diffraction of light or electrons.

If light from a laser is shone through a narrow adjustable slit, a diffraction pattern will form on the screen behind the slit (Figure 7.6.17), and if single photons pass through the slit, a diffraction pattern is also formed.

The observations can be explained by treating the laser light as a wave and considering the interference of different sections of the wavefront that pass through the slit. However, explaining this pattern in terms of the motion of individual photons is much more challenging.

Consider a single photon on its journey from the laser to the screen. As the photon passes through the slit, its position is known to some degree of certainty (Δx). According to Heisenberg, this means that there must therefore be an uncertainty about the photon's momentum (Δp), which is why the photon can end up at a variety of different places on the screen. Schrödinger's wave equation could be used to explain why some paths for the photon are much more likely than others.

Since the slit is adjustable, it could be made narrower. This would decrease the uncertainty about the photon's position, Δx, but increase the uncertainty about the photon's momentum, Δp. This is why narrowing the slit causes the diffraction pattern to spread out.

If the slit is removed, as shown in Figure 7.6.18, there are no constraints on the path of the light, the uncertainty about the position (Δx) of each individual photon is large. Correspondingly, the uncertainty about the momentum of the photon becomes very small and no diffraction occurs—all the photons end up very close together in a small spot in the middle of the screen.

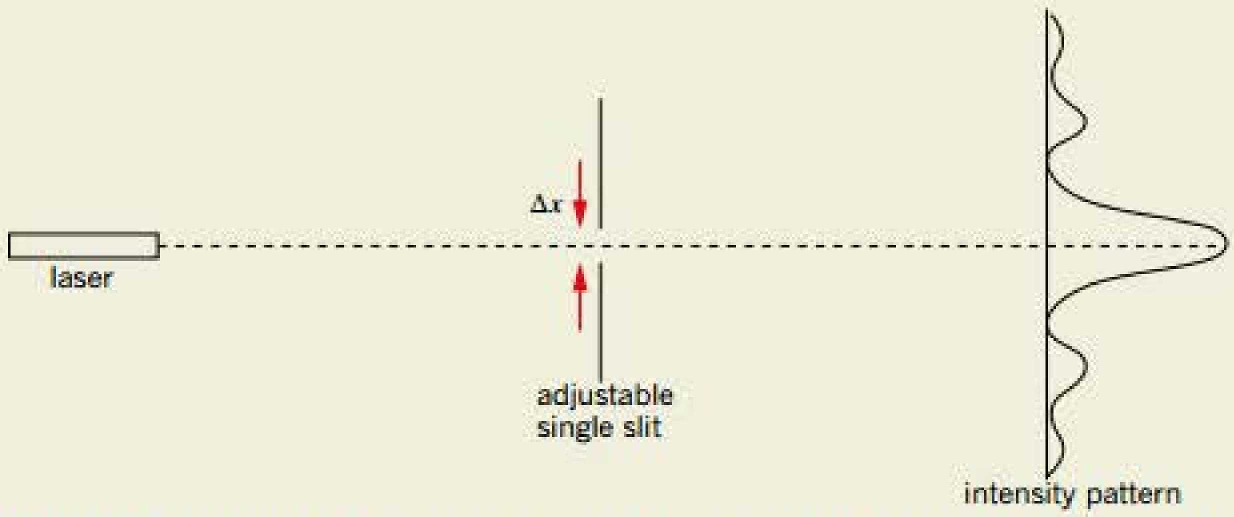


FIGURE 7.6.17 Demonstrating Heisenberg's uncertainty principle with diffraction through a single slit.

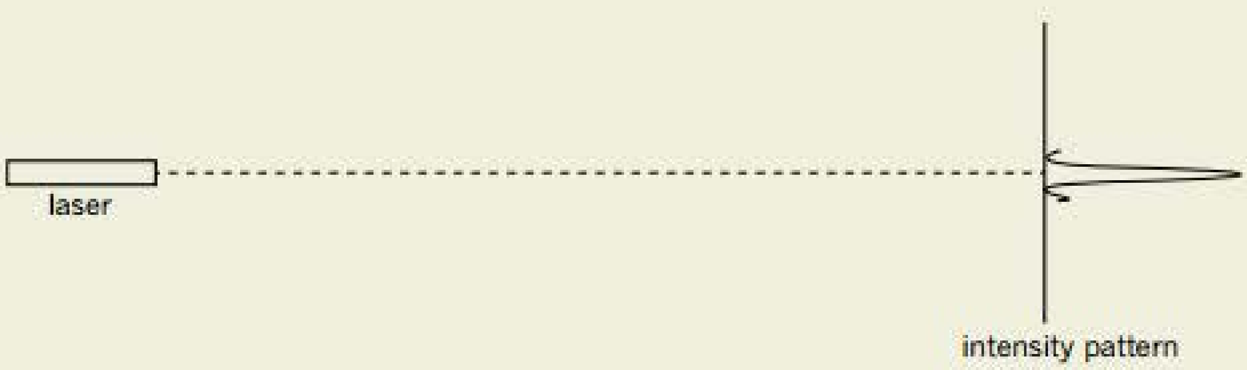


FIGURE 7.6.18 Without a slit, there is no uncertainty about the momentum of the photon and therefore no diffraction occurs.

PHYSICSFILE

Interpretation

Scientists have long argued about how we should interpret **Heisenberg's uncertainty principle.** In the early 20th century, many were unhappy with the implication of Schrödinger's work that the motion of an electron could only be described in terms of probabilities. Even Einstein famously argued against this interpretation, saying 'God does not play dice with the universe' (Figure 7.6.19).

However, experiments over the past century have confirmed that uncertainty is a fundamental property of the quantum mechanical universe.

FIGURE 7.6.19 Einstein's claim that 'God does not play dice' is challenged by modern understandings of quantum physics.

PHYSICS IN ACTION

Quantum computing and other quantum applications

A normal computer stores information in a series of zeros and ones known as 'bits'. However, in a quantum computer 'qubits' exist in various quantum mechanical states that are a superposition of zeros and ones; in the simplest example, the 00, 01, 10 and 11 states can exist simultaneously. These can be coded on quantum phenomena such as electron spin and this can vastly increase the computing power of the quantum computer over that of a conventional computer.

Quantum entanglement occurs when two entangled particles are so strongly correlated that measuring one particle instantly affects the other, even when the particles are separated by a large distance. This plays a role in applications such as quantum teleportation and quantum cryptography.

A QUANTUM VIEW OF THE WORLD

Bohr's model of the atom, explained in this section and Section 7.5, was a mixture of classical and quantum theories, partially recognising the wave–particle duality of light and matter. However, it allowed for the development of new, more-radical theories by physicists such as Schrödinger and Heisenberg. And intriguingly, quantum mechanics has confirmed certain aspects of the Bohr model, such as atoms existing only in discrete states of definite energy and the emission or absorption of photons of light when electrons make transitions from one energy state to another within an atom.

However, quantum mechanics goes much further—very much further than the brief introduction in this section. Quantum mechanics proposes that the dual nature of particles, particularly fundamental particles such as the electron, prevents knowing the position and speed of an object at the same time. Physicists can only calculate the probability that an electron will be observed at a particular place around an atom. In this view of the world there is some inherent unpredictability. In fact, it becomes meaningless to ask *how* an electron gets from one state to another when an atom emits or absorbs a photon of light—it just does.

7.6 Review

SUMMARY

- On the atomic level, energy and matter exhibit the characteristics of both waves and particles.
- The wavelength of a particle is given by the de Broglie equation:

  $$\lambda = \frac{h}{p}$$

  i.e. $\lambda = \frac{h}{mv}$
- Young's double-slit experiment is explained with a wave model but produces the same interference and diffraction patterns when one photon at a time or one electron at a time is passed through the slits.
- In particle-scattering experiments, beams of particles (electrons usually) are made to travel with a speed so that their matter wavelength approximates the interatomic spacing in a crystal. Consequently, a diffraction pattern is produced that can only be explained if matter has a wave-like nature.
- If photons and matter particles being scattered by the same crystal sample produce the same fringe spacing, then they must have the same wavelength and momentum.
- All matter, like light, has a dual nature. Through everyday experience we see matter as particle-like. However, on an atomic or molecular scale the wave-like nature of matter becomes important. This symmetry in nature—the dual nature of light and matter—is referred to as wave–particle duality.
- de Broglie viewed electrons as matter waves. His standing-wave model for electron orbits provided a physical explanation for electrons only being able to occupy particular energy levels in atoms. He suggested that the only way that the electron could maintain a steady energy level was if it established a standing wave.
- The quantised states of the atom are analogous to the quantised standing waves that are known to occur in physical objects such as strings.
- The nature of the universe at the very smallest of scales is fundamentally different from the way the universe is perceived at the macroscopic scale.
- Quantum mechanics is the study of the wave properties of electrons and other subatomic particles. The wave properties of an electron are interpreted as describing the probability of finding an electron at a particular point in space, hence the term 'electron cloud' or 'orbital' is used.
- Heisenberg's uncertainty principle results from wave–particle duality and states that it is not possible to know the exact position and momentum of a particle simultaneously.
- The single-slit diffraction pattern is not produced by photon interactions but by the probability of where a single photon may end up.

KEY QUESTIONS

1 What is the de Broglie wavelength of an electron travelling at $1.0 \times 10^6\,\text{m s}^{-1}$?

2 Calculate the speed of an electron that has a de Broglie wavelength of $4.0 \times 10^{-9}\,\text{m}$.

3 Which of the following conclusions can be drawn from de Broglie's investigation into the existence of matter waves?

A All particles exhibit wave behaviour.
B Only moving particles exhibit wave behaviour.
C Only charged particles exhibit wave behaviour.
D Only moving, charged particles exhibit wave behaviour.

4 In an experiment to determine the structure of a crystal, identical diffraction patterns were formed by a beam of electrons and a beam of X-rays with a frequency of $8.6 \times 10^{18}\,\text{Hz}$.

a Calculate the wavelength of the electrons.
b Calculate the speed of the electrons.

5 Explain why a cricket player does not have to consider the wave properties of a cricket ball while batting.

6 At what speed would a proton be travelling if it were to have the same wavelength as a gamma ray of energy $6.63 \times 10^{-14}\,\text{J}$? (Mass of a proton = $1.67 \times 10^{-27}\,\text{kg}$)

7 A charge q of mass m is accelerated from rest through a potential difference of V. Derive an expression that defines the de Broglie wavelength, λ, of the mass in terms of q, m and V.

8 A corollary of de Broglie's work on matter waves is that photons can be considered to have momentum. The momentum of photons, although small, has been measured under laboratory conditions. Use de Broglie's equation to find an equation for the momentum of a photon of wavelength λ.

9 Why can an electron microscope resolve images in finer detail than an optical microscope?

10 What do de Broglie's matter-wave concept and a bowed violin string have in common?

Chapter review

07

KEY TERMS

absorption spectrum
blackbody
coherent
constructive interference
critical angle
de Broglie wavelength
destructive interference
diffraction
diffraction pattern
dispersion
electromagnetic radiation
electromagnetic spectrum
electron-volt
emission spectrum
excited state
forward voltage
ground state
Heisenberg's uncertainty principle
incandescent
interference
laser
light-emitting diode
metal vapour lamp
monochromatic
path difference
photocurrent
photoelectric effect
photoelectron
photon
polarisation
quantum
quantum mechanics
reflection
refraction
refractive index
reverse voltage
Snell's law
standing wave
stopping voltage
synchrotron
threshold frequency
total internal reflection
wave–particle duality
work function

1 The diagram below demonstrates the phenomenon of:

A diffraction
B interference
C reflection
D refraction

2 Explain how the width of a double-slit interference pattern would change if all the variables were constant but a blue laser was replaced with a green laser.

3 Polarisation is an important phenomenon. What does it show about light?

A It can travel instantaneously at an infinite speed.
B It travels faster in materials such as water and air than in a vacuum.
C It is a longitudinal wave.
D It is a transverse wave.

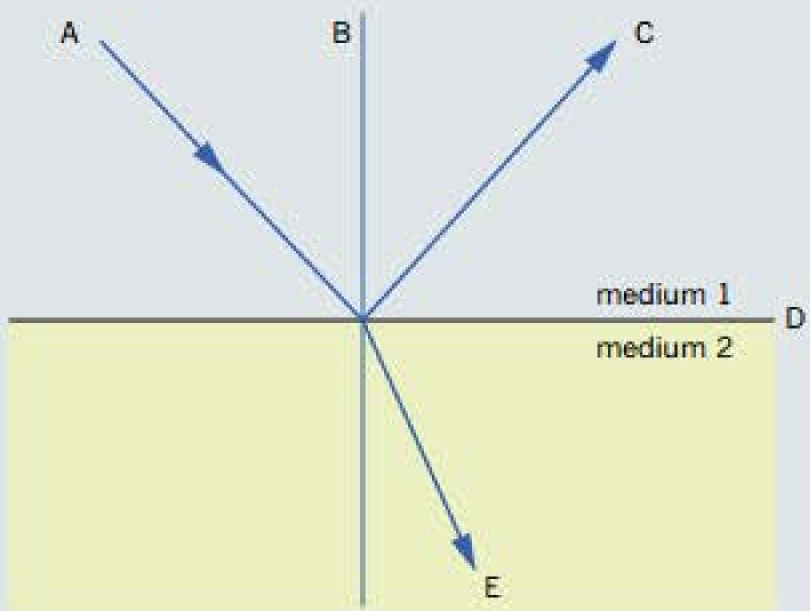
4 Explain briefly why snowboarders and sailors will benefit from wearing polarising sunglasses.

5 Red light (4.5×10^{14} Hz) has a wavelength of 500 nm in water. Calculate the speed of red light in water.

6 Choose the correct answers from those given in bold to complete the following sentence about refraction.

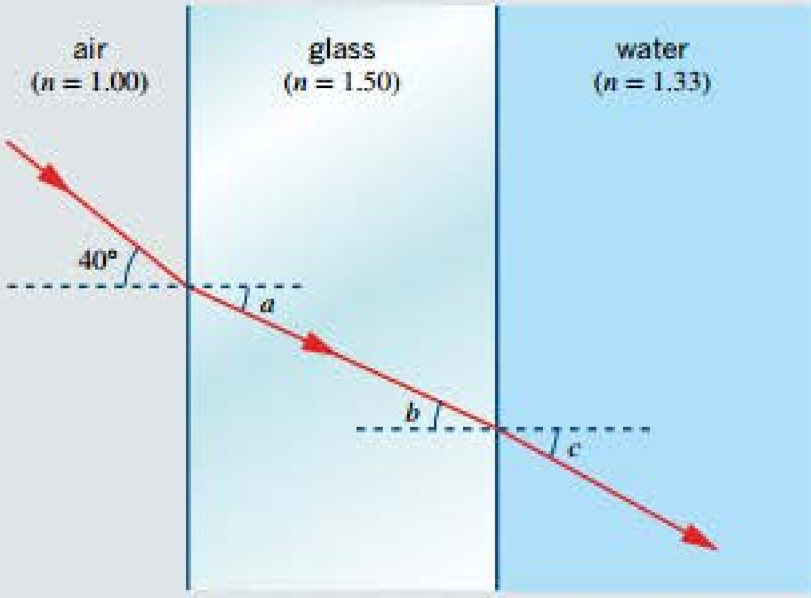
As light travels from quartz ($n = 1.46$) to water ($n = 1.33$), its speed **increases/decreases**, which causes it to refract **away from/towards** the normal.

7 The figure represents a situation involving the refraction of light. Identify the correct label for each of the lines from the choices provided: boundary between media, reflected ray, incident ray, normal, refracted ray.



8 The speed of light in air is 3.00×10^8 m s^{-1}. As light strikes an air–perspex boundary, the angle of incidence is 43.0° and the angle of refraction is 28.5°. Calculate the speed of light in perspex.

9 A ray of light travels from air, through a layer of glass and then into water as shown. Calculate angles *a*, *b* and *c*.



10 A ray of light exits a glass block. On striking the inside wall of the glass block, the ray makes an angle of 58.0° with the glass–air boundary. The index of refraction of the glass is 1.52. Calculate the:

a angle of incidence

b angle of refraction of the transmitted ray (assuming $n_{air} = 1.00$)

c angle of deviation

d speed of light in the glass.

11 A narrow beam of white light enters a crown glass prism with an angle of incidence of 30.0°. In the prism, the different colours of light are slowed to varying degrees. The refractive index for red light in crown glass is 1.50 and for violet light the refractive index is 1.53. Calculate the:

a angle of refraction for the red light

b angle of refraction for the violet light

c angle through which the spectrum is dispersed

d speed of the violet light in the crown glass.

12 Calculate the critical angle for light travelling between the following media.

	Incident medium	Refracting medium
a	ice ($n = 1.31$)	air ($n = 1.00$)
b	salt ($n = 1.54$)	air ($n = 1.00$)
c	cubic zirconia ($n = 2.16$)	air ($n = 1.00$)

13 When a light ray refracts, the difference between the angle of incidence and angle of refraction is known as the *angle of deviation*. Sort the following boundaries between media in order of increasing angle of deviation.

A water ($v = 2.25 \times 10^8\,\text{m s}^{-1}$) to diamond ($v = 1.24 \times 10^8\,\text{m s}^{-1}$)

B water ($v = 2.25 \times 10^8\,\text{m s}^{-1}$) to air ($v = 3 \times 10^8\,\text{m s}^{-1}$)

C air ($v = 3 \times 10^8\,\text{m s}^{-1}$) to diamond ($v = 1.24 \times 10^8\,\text{m s}^{-1}$)

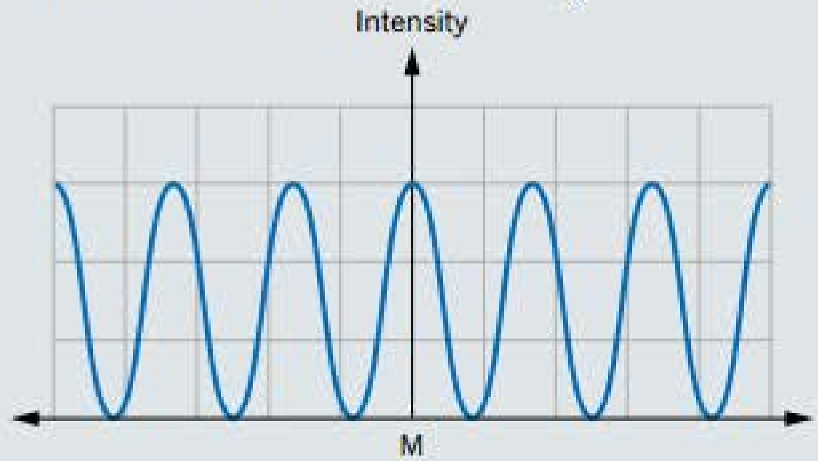
D glass ($v = 1.97 \times 10^8\,\text{m s}^{-1}$) to air ($v = 3 \times 10^8\,\text{m s}^{-1}$)

14 Light of an unknown wavelength emitted by a laser is directed through a pair of thin slits separated by 75 μm. The slits are 4.0 m from a screen on which bright fringes are 3.1 cm apart.

a Calculate the wavelength of the laser light in nm.

b Identify the unknown colour emitted by the laser.

15 The following diagram shows the resulting intensity pattern (simplified) after light from two slits reaches the screen in a double-slit experiment. Copy the diagram into your workbook and circle the points at which the path difference is equal to $1\frac{1}{2}\lambda$.



16 Arrange the types of electromagnetic radiation below in order of decreasing wavelength.

gamma rays, visible, microwaves, radio waves, X-rays, infrared, ultraviolet

17 What form of electromagnetic radiation is used in the following applications?

a mobile phone communication

b night-vision goggles

c medical imaging

18 An AM radio station has a frequency of 612 kHz . If the speed of light is $3 \times 10^8\,\text{m s}^{-1}$, calculate the wavelength of these waves to the nearest metre.

19 Describe Young's experiment and explain why it is considered evidence for the wave theory of light.

20 Explain briefly why a microwave oven is tuned to produce electromagnetic waves of a particular frequency.

21 What is the energy, in electron-volts, of light with a frequency of 6.0×10^{14} Hz?

22 What is the approximate value of the energy in J of a quantum of light with energy of 5.0 eV?

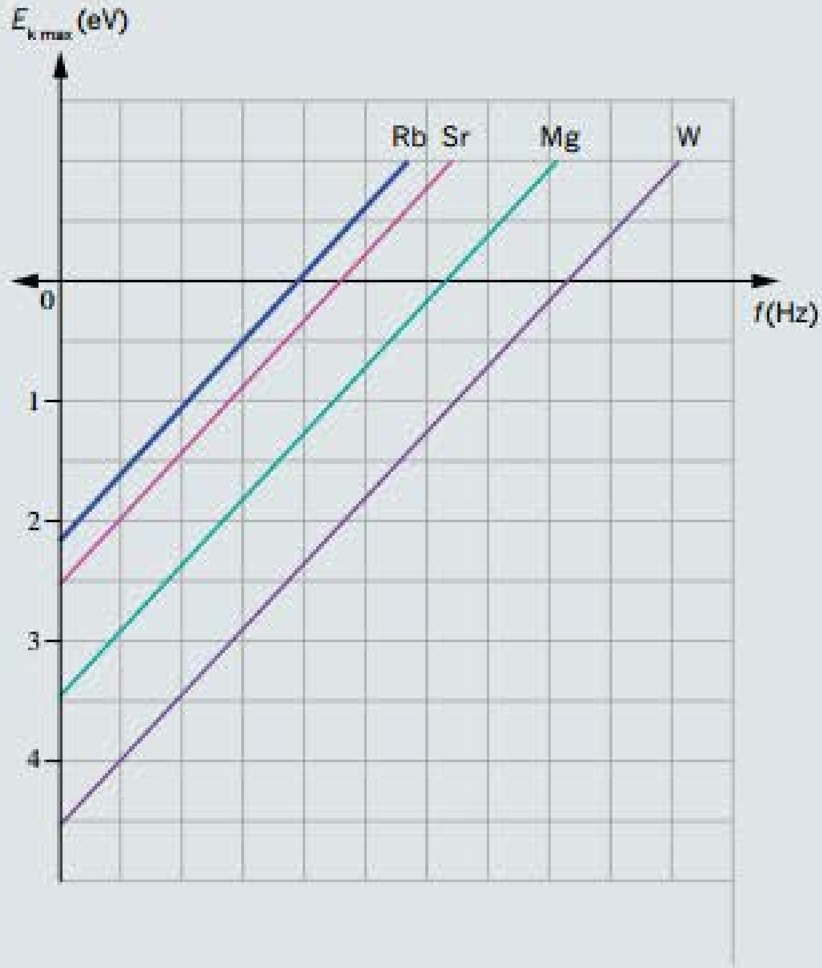
23 What name is given to the electrons released from a metal surface due to the photoelectric effect?

24 If the work function for nickel is 5.0 eV, what is the threshold frequency for nickel?

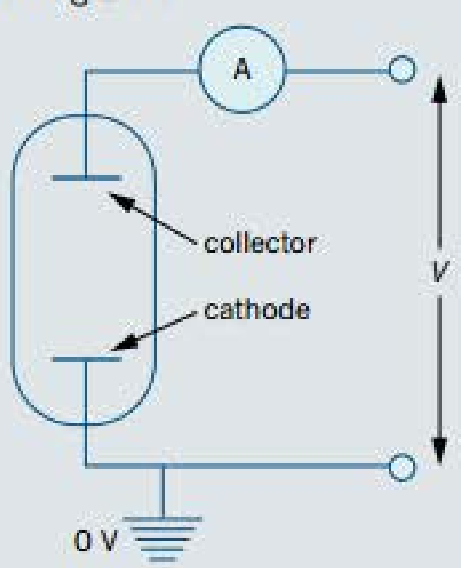
25 Platinum has a threshold frequency of 1.5×10^{15} Hz. Calculate the maximum kinetic energy, in electron-volts, of the emitted photoelectrons when ultraviolet light with a frequency of 2.2×10^{15} Hz shines on it.

26 The stopping voltage obtained using a particular photocell is 1.95 V. Determine the maximum kinetic energy of the photoelectrons in electron-volts.

27 From the graph, determine the value of the work function for each metal.



28 The cathode of a particular photocell, shown below, is coated with rubidium. Incident light of varying frequencies is directed onto the cathode of the cell and the maximum kinetic energy of the photoelectrons is logged. The results are summarised in the following table.



Frequency (Hz) × 10^{14}	$E_{k\,max}$ (eV)
5.20	0.080
5.40	0.163
5.60	0.246
5.80	0.328
6.00	0.411
6.20	0.494

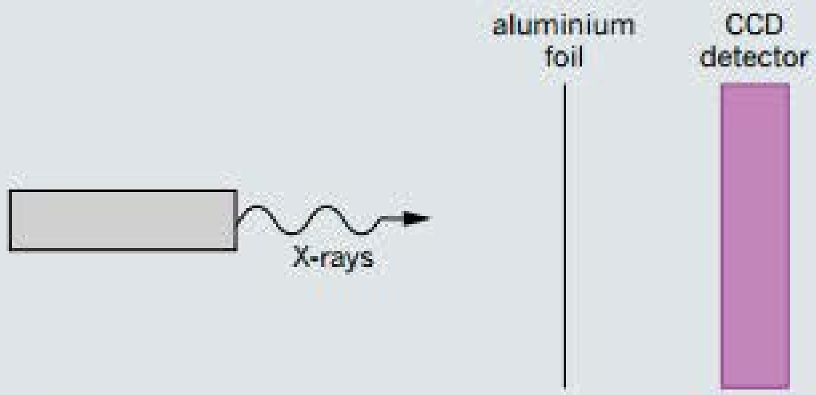
a Plot the points from the table on a graph.

b Calculate the gradient of the graph.

c Based on the graph of the experimental results, what is the threshold frequency for rubidium?

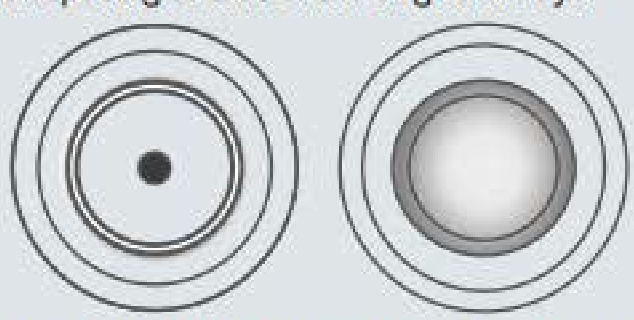
d Will red light of wavelength 680 nm cause photoelectrons to be emitted from the rubidium surface? Justify your answer.

29 In an X-ray diffraction experiment, a beam of X-rays from a synchrotron is directed onto a thin sheet of aluminium foil. The X-rays are scattered by the foil and detected via a CCD (charge-coupled device) behind the foil, which forms a digital image of the resulting pattern.



a If the wavelength of the X-rays is 260 pm (260×10^{-12} m), what is the energy of the X-rays?

b The CCD displays an image on a computer screen of the diffraction pattern formed once the X-rays have passed through the foil. This is shown in the diagram on the left below. A beam of accelerated electrons is then substituted for the X-rays. This is shown in the diagram on the right below. A very similar diffraction pattern is observed. Why do the electrons produce a diffraction pattern with a similar spacing to that of the original X-rays?



c Based on the assumption that the two diffraction patterns have the same radius, what is the momentum of the electrons?

30 Davisson and Germer conducted an experiment in which electrons were scattered after being fired at a target.

a What was observed by a detector moving through the scattering angles?

b What was the implication of this?

31 A particular atom has four energy levels. In this context, what does it mean to say that the levels are *quantised?*

32 Calculate the frequency of the photon produced when an electron in a hydrogen atom drops from the $n = 3$ energy level to its ground state, $n = 1$. Refer to Figure 7.5.6 on page 274.

33 Explain why the development of the Bohr model of the hydrogen atom was significant in the development of a comprehensive understanding of the nature of light.

34 Describe the relationship between the colours seen in the emission and absorption spectra of hydrogen.

35 By referring to the behaviour of electrons, explain how light is produced in an incandescent (filament) light globe.

36 A student has measured the wavelength and threshold voltage of several semiconductor LEDs. Use the student's results to plot a graph and find an experimental value for Planck's constant.

Semiconductor	Threshold voltage (V)	λ (nm)
GaSb	0.68	1828
GaAs	1.43	869
GaP	2.25	553
ZnS	3.6	345

37 Which of the following would have the longest wavelength?

A electron, $m = 9.1 \times 10^{-31}$ kg, $v = 7.5 \times 10^{6}$ m s^{-1}

B blue light, $\lambda = 470$ nm

C X-ray, $f = 5 \times 10^{17}$ Hz

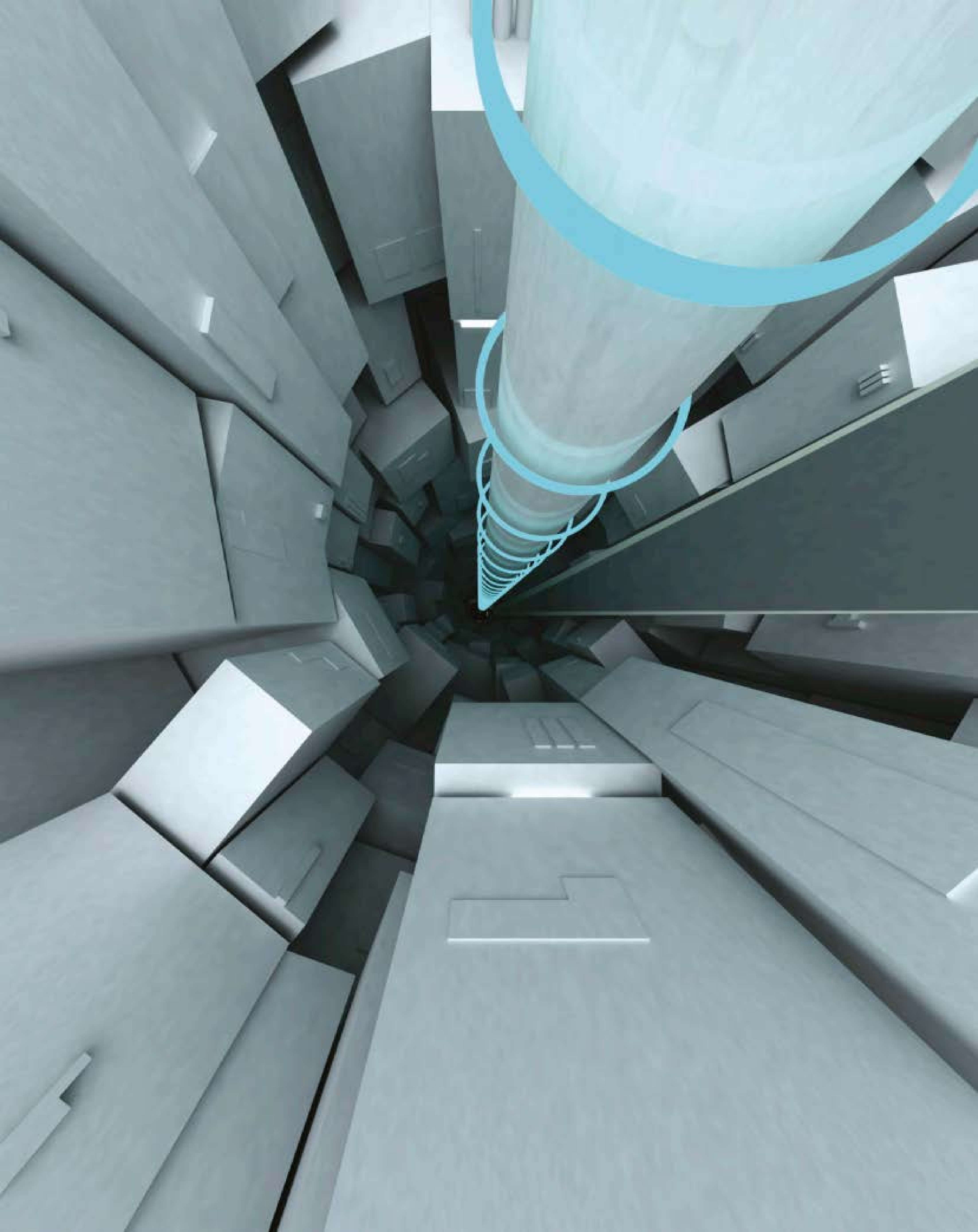
D proton, momentum $= 1.7 \times 10^{-21}$ kg m s^{-1}

38 Calculate the de Broglie wavelength of a 40 g bullet travelling at 1.0×10^{3} m s^{-1}.

39 Would wave behaviours such as diffraction be noticeable for the bullet described in Question 38?

40 According to Heisenberg's uncertainty principle, if the uncertainty in position is decreased, what will happen to the uncertainty in momentum?

41 If a photon of a very short wavelength were to collide with an electron, what would be the effect on the position of the electron?



CHAPTER 08

Special relativity

In this chapter, you will explore the concepts of classical physics, as described by Galileo and Newton, and the evidence that pointed towards the need for some different thinking. Einstein's special relativity is presented as a solution to the problem of classical physics at speeds approaching the speed of light.

Science as a Human Endeavour

Research studies of cosmic rays show that interactions between cosmic rays and the upper atmosphere produce muons. These particles have a lifetime of about two microseconds and should have ceased to exist before reaching the surface of the Earth. However, because they are travelling near the speed of light, the time dilation effect allows them to complete their journey. Continuing research in the field of high-energy physics is important for improving our understanding of our world and its origins.

Science Understanding

- observations of objects travelling at very high speeds cannot be explained by Newtonian physics. These include the dilated half-life of high-speed muons created in the upper atmosphere, and the momentum of high-speed particles in particle accelerators
- Einstein's special theory of relativity predicts significantly different results to those of Newtonian physics for velocities approaching the speed of light
- the special theory of relativity is based on two postulates: that the speed of light in a vacuum is an absolute constant, and that all inertial reference frames are equivalent
- motion can only be measured relative to an observer; length and time are relative quantities that depend on the observer's frame of reference
  This includes applying the relationships

$$\ell = \ell_0\sqrt{\left(1-\frac{v^2}{c^2}\right)} \qquad t = \frac{t_0}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}} \qquad u = \frac{v+u'}{1+\frac{vu'}{c^2}} \qquad u' = \frac{u-v}{1-\frac{uv}{c^2}}$$

- relativistic momentum increases at high relative speed and prevents an object from reaching the speed of light
  This includes applying the relationship

$$p_v = \frac{mv}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}}$$

- the concept of mass–energy equivalence emerged from the special theory of relativity and explains the source of the energy produced in nuclear reactions
  This includes applying the relationship

$$E = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

8.1 Einstein's theory of special relativity

Galileo and Newton developed theories of motion. These theories allowed the relative motion of low-speed objects to be modelled mathematically. This section presents the observations that challenged Galilean relativity and Newtonian physics, and explains the key principles that led to the new physics described by Albert Einstein (Figure 8.1.1), known as the theory of special relativity.

FIGURE 8.1.1 Albert Einstein statue in Washington, D.C.

EINSTEIN'S BRILLIANT THEORY

FIGURE 8.1.2 Einstein as a teenager.

When Albert Einstein was just 5 years old, his father gave him a compass. He was fascinated by the fact that it was responding to some invisible field that enveloped the Earth. His curiosity was aroused and, fortunately for physics, he never lost that curiosity. In his teens he turned his attention to the question of light (Figure 8.1.2).

Perhaps it was lucky that in his early twenties Einstein was not part of the physics 'establishment'. He was working as a patent clerk in the Swiss Patent Office. It was an interesting enough job, but it left him time to think about electromagnetic waves (light) and their relationship to the Galilean principle of relativity.

Galileo opposed the conventional thinking of his time and suggested that the natural state of objects is not to be at rest, but in a state of uniform motion. This means that an object will continue in its state of motion without any force acting on it at all. For Galileo, a state of rest was not special; it was simply an example of uniform motion where the magnitude of the velocity was zero. He was able to support this theory mathematically in what is now known as the principle of inertia, often referred to as Newton's first law.

The idea that there is nothing special about a velocity of zero forms the basis of the Galilean principle of relativity. In Aristotle's world, zero velocity was something very special; to Galileo (and later Newton), zero velocity was no different, in principle, to any other velocity. Galileo argued that a force acting for a period of time could change the velocity of an object from 0 to $50\,\text{m}\,\text{s}^{-1}$, and that the same force acting for the same period of time could also change the object's velocity from $-25\,\text{m}\,\text{s}^{-1}$ to $+25\,\text{m}\,\text{s}^{-1}$, or from $1000\,\text{m}\,\text{s}^{-1}$ to $1050\,\text{m}\,\text{s}^{-1}$.

Galileo was particularly interested in relative motion. One of his famous experiments involved the dropping of a cannon ball from the top of the mast of a moving ship. Galileo found that the motion of the cannon ball was not affected by the motion of the ship; the cannon ball landed next to the base of the mast. His principle of relativity was that you cannot tell if you are moving or not without looking outside of your own **frame of reference**.

Based on the work of Galileo, Isaac Newton established detailed models for the motion of objects such as planets, moons and comets, even falling oranges. According to his equations, the velocity of objects can be calculated relative to any frame of reference, as long as the velocity of the frame of reference is known. The Newtonian principle that the velocities of objects and frames of reference can be added together to determine the velocity of the object in another frame of reference is common throughout his equations and laws.

Consider an object moving in one frame of reference, A. This frame of reference is moving in another frame of reference, B. The velocity of the object in B is given by:

$$v_{\text{object in B}} = v_{\text{object in A}} + v_{\text{A in B}}$$

A practical example of this could be when a person runs forwards in the carriage of a train. Here, the train is frame of reference A, and the track along which the train moves is frame B. Imagine that the person runs at $5\,\text{m}\,\text{s}^{-1}$ forwards, while the train travels at a velocity of $20\,\text{m}\,\text{s}^{-1}$ forwards. The velocity of the person relative to B, the track, is:

$$\begin{aligned} v_{\text{person along track}} &= v_{\text{person in train}} + v_{\text{train along track}} \\ &= 5 + 20 \\ &= 25\,\text{m}\,\text{s}^{-1} \end{aligned}$$

That is, the person is moving with a velocity of $25\,\text{m}\,\text{s}^{-1}$ forwards when measured against the track.

Einstein was a typical theoretician; the only significant experiments he ever did were thought experiments, or ***Gedanken*** experiments, as they are called in German. Many of his *Gedanken* experiments involved thinking of situations that involved two frames of reference moving with a steady relative velocity, in which the principles of Galilean relativity applied. Newton had referred to these as **inertial frames of reference**, as the law of inertia applied within them.

Einstein and Galilean relativity

Einstein decided that the elegance of the principle of Galilean relativity was such that it simply had to be true. Nature did not appear to have a special frame of reference, and Einstein could see no reason to believe that there was one waiting to be discovered. In other words, there is no such thing as an absolute velocity. It is not possible to have a velocity relative to space itself, only to other objects within space. So the velocity of any object can always be stated as relative to some other object. In the case of the person running on the train, their velocity can be stated as either $5\,\text{m}\,\text{s}^{-1}$ relative to the train or $25\,\text{m}\,\text{s}^{-1}$ relative to the track.

Einstein expanded the Galilean principle to state that all inertial frames of reference must be equally valid, and that the laws of physics must apply equally in any frame of reference that is moving at a constant velocity. So there is no physics experiment you can do that is entirely within a frame of reference that will tell you that you are moving. In other words, as you speed along in your *Gedanken* train with the blinds down, you cannot measure your speed. You can tell if you are accelerating easily enough: just hang a pendulum from the ceiling. However, the pendulum will hang straight down whether you are travelling steadily at $100\,\text{km}\,\text{h}^{-1}$ or are stopped at the station. Consider Figure 8.1.3(a) and (b) on page 306. There is no way of telling which of the trains is stationary relative to the ground, or which is moving at a constant velocity.

Einstein decided that the relativity principle could not be abandoned. Recall that Einstein was, at the time, thinking about the relationship between light and relativity. Whatever the explanation for the strange behaviour of light, it could not be based on a flaw in the principle of Galilean relativity.

PHYSICSFILE

Frames of reference

The usual frame of reference is the Earth's surface, but this does not have to be the case. When a car is said to be 'doing' $60\,\text{km}\,\text{h}^{-1}$, the assumption is that the car's velocity is measured by a radar gun that is at rest relative to the road. However, if the reading is made by a radar gun mounted on a police car travelling at a constant velocity, the observer's frame of reference is different. From the frame of reference of a police car following behind at $50\,\text{km}\,\text{h}^{-1}$ relative to the road, the same car is only travelling at $10\,\text{km}\,\text{h}^{-1}$.

If you consider the Earth's surface to be the inertial frame of reference, then the Earth is stationary in that frame of reference. However, you know that the Earth rotates about its axis, completing one revolution per day. You also know that the Earth revolves around the Sun and that the solar system is moving through space about the Milky Way's centre, and this galaxy also moves through space. So the Earth is definitely not stationary.

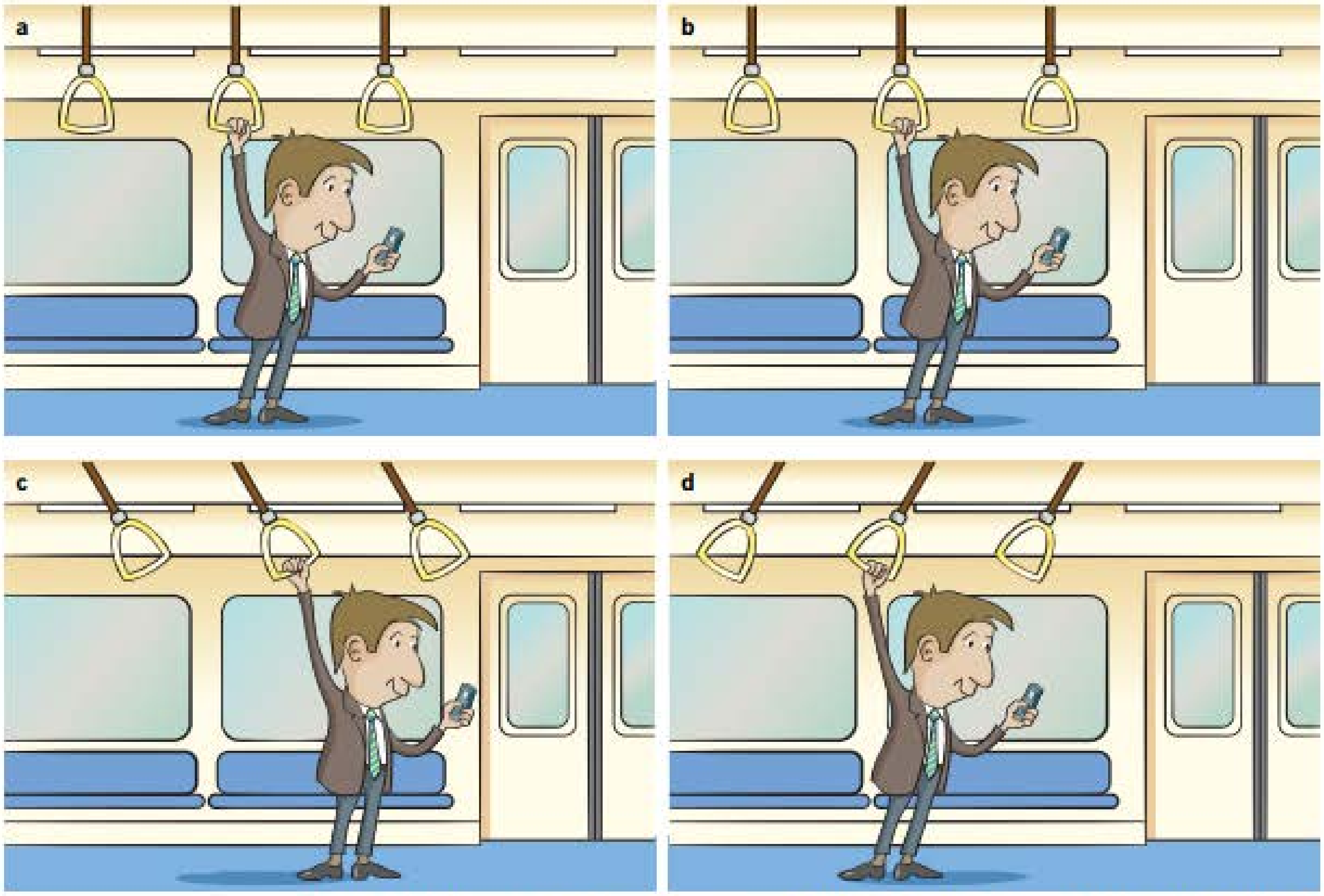


FIGURE 8.1.3 There is no observation or experiment that shows the difference between two inertial frames of reference (a) and (b). In one of the situations illustrated, the train is stationary, and in the other it is moving smoothly at 100 km h^{-1}. There is no observation that will tell which one is which. In (c) and (d), the motion of the handles hanging from the ceiling of the train indicate that these trains are not moving at a constant speed.

Einstein's fascination with the nature of light had led him to a deep understanding of Maxwell's work on the electromagnetic nature of light waves. He was convinced of the elegance of Maxwell's equations and their prediction of a constant speed of light. Most physicists believed that the constant speed predicted by Maxwell's equations referred to the speed of light relative to a **medium** (a substance it travelled through). It was thought that the speed predicted would be the speed in the medium in which light travelled, and the measured speed would have to be adjusted for one's own speed through that medium.

As light travelled through the vacuum of space between the Sun and Earth, clearly the medium was no ordinary material. Physicists gave it the name **aether**, as it was an 'ethereal' substance. It was thought, following Maxwell's work, that the aether must be some sort of massless, rigid medium that 'carried' electric and magnetic fields.

This was a real problem for Einstein. A speed of light that is fixed in the aether and which depended on the velocity of an inertial frame in the aether would be in direct conflict with the principle of Galilean relativity, which Einstein was reluctant to abandon.

Resolving the problem of the aether

As in any conflict, the resolution is usually found by people who are prepared to look at it in new ways. This was the essence of Einstein's genius. Instead of looking for faults in what appeared to be two perfectly good principles of physics, he decided to see what happened if they were both accepted, despite the apparent contradiction.

EXTENSION

The Michelson–Morley experiment

The existence of an aether appeared to be a serious blow for the principle of relativity. It seemed that there might be, after all, a frame of reference attached to space itself. If this was the case, there was the possibility of an absolute zero velocity.

Scientists needed to test the idea of electromagnetic waves moving through the aether. Since the Earth is in orbit around the Sun, an aether wind should be blowing past the Earth. This suggested to American physicist Albert Michelson that it should be possible to measure the speed at which the Earth was moving through the aether by measuring the small changes in the speed of light as the Earth changed its direction of travel. For example, if the light was travelling in the same direction as the Earth, through the aether, the apparent speed should be slower than usual, i.e. at $c - v$ (Figure 8.1.4). It would be as if the light was travelling against an aether 'wind' created by the motion of the Earth through it. If the light was travelling against the Earth's motion, the apparent speed should be faster as it would be travelling with the 'wind' at $c + v$ (Figure 8.1.4). The differences would be tiny, less than 0.01%, but Michelson was confident that he could measure them.

In the 1880s Michelson, and his collaborator Edward Morley, set up a device known as an interferometer. The device cannot measure the speed of light but it can detect changes in the speed of light that might have been due to the aether wind. In fact, the interferometer was used to attempt to measure the very small differences in the time taken for light to travel in two mutually perpendicular directions. They were able to rotate the whole apparatus and hoped to detect the small difference that should result from the fact that one of the directions was to be the same as that in which the Earth was travelling and the other at right angles. However, they found no difference. Perhaps, then, the Earth at that time was stationary with respect to the aether? Six months later, however, when the Earth would have to be travelling in the opposite direction relative to the aether, there was still no difference in the measured speeds. Other people performed similar experiments, always with the same null result. In whatever direction the Earth was moving it seemed to be at rest in the aether. Or perhaps there was no aether at all.

While Michelson and Morley's results were consistent with Maxwell's prediction that the speed of light would always appear to be the same for any observer, the apparent absurdity of such a situation led most physicists to believe that some flaw in the theory behind the experiment, or in its implementation, would soon be discovered. Einstein, however, wondered about the consequences of actually accepting their prediction about the speed of light but at the same time holding on to the relativity principle.

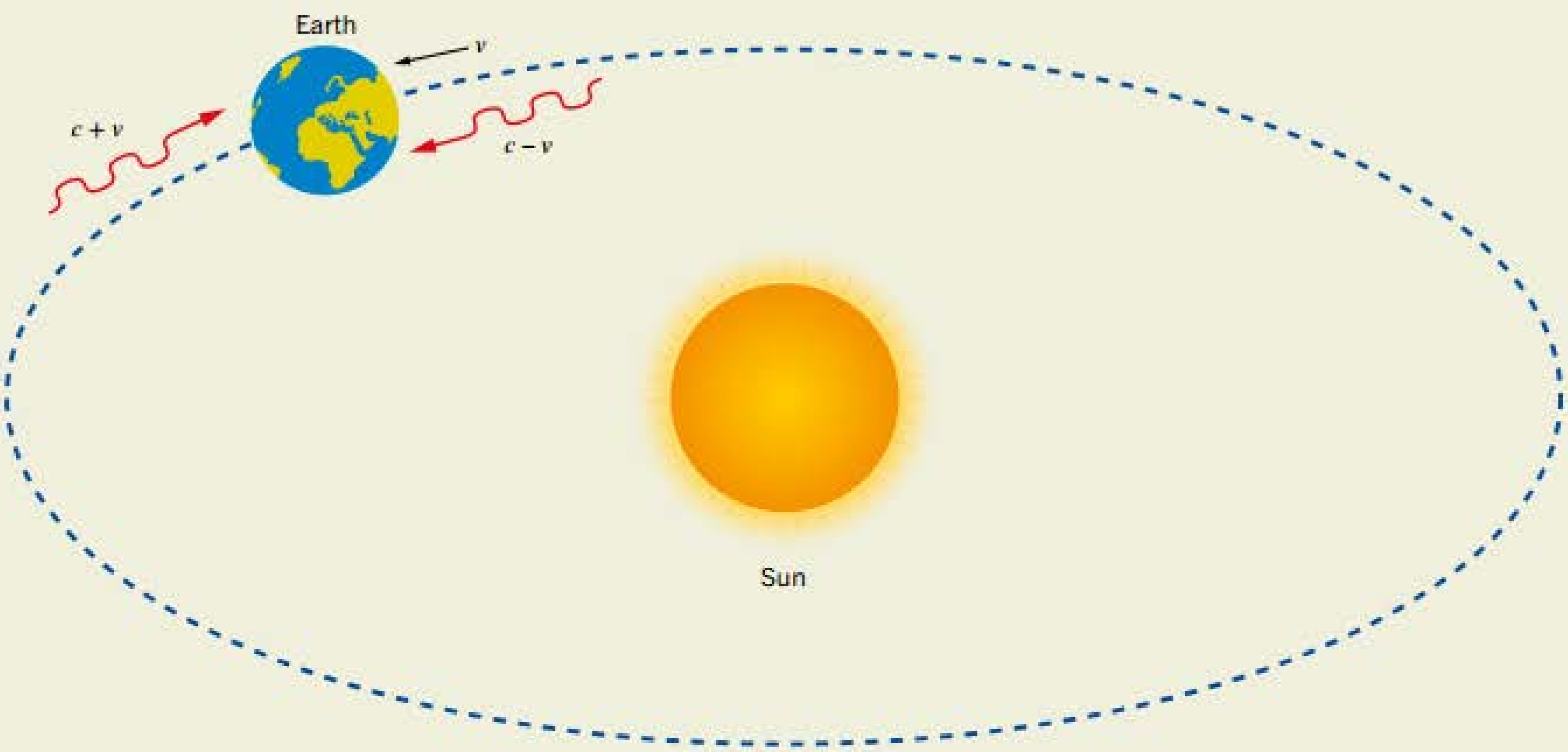


FIGURE 8.1.4 The basic principle of the Michelson–Morley experiment. If the aether is fixed relative to the Sun, and the light is travelling at c relative to the aether and in the same direction as the Earth, the apparent speed should be less than c, i.e. $c - v$. If the light was travelling in the opposite direction to Earth, light should appear faster than c, i.e. $c + v$.

So Einstein swept away the problem of the aether, saying that it was simply unnecessary. It had been invented only to be a medium for light waves, and no one had found any evidence for its existence (refer to the Michelson–Morley experiment in the Extension box on page 307). Electromagnetic waves, he said, could apparently travel through empty space without a medium. Doing away with the aether, however, did not solve the basic conflict between the absolute speed of light and the principle of relativity.

EINSTEIN'S THEORY OF SPECIAL RELATIVITY

Einstein's two postulates:

- The laws of physics are the same in all inertial (non-accelerated) frames of reference.
- The speed of light has a constant value for all observers regardless of their motion or the motion of the source.

(The first postulate means that there is no preferred frame of reference and so is sometimes stated as: No law of physics can identify a state of absolute rest.)

Although Einstein accepted both Galileo's and Maxwell's theories despite the apparent contradiction, this still left the question of how two observers travelling at different speeds could see the same light beam travelling at the same speed. The answer, Einstein said, was in the very nature of space and time.

In 1905 he sent a paper to the respected physics journal *Annalen der Physik* entitled 'On the electrodynamics of moving bodies'. In this paper he put forward two simple **postulates** (statements assumed to be true) and followed them to their logical conclusion. It was this conclusion that was so astounding.

Einstein's postulates

The first postulate is basically that of Newton, but Einstein extended it to include the laws of electromagnetism so elegantly expressed by Maxwell. The second postulate simply takes Maxwell's prediction about the speed of electromagnetic waves in a vacuum at face value.

These two postulates sound simple enough; the only problem was that, according to early Newtonian physics, they were contradictory.

Consider the example illustrated in Figure 8.1.5. Binh is in his spaceship travelling away from Clare at a speed v, and Clare turns on a laser beam to signal Binh. The first postulate seems to imply that the speed of the laser light, as measured by Binh, should be $c - v$, where c is the speed of light in Binh's frame of reference. This is what you would expect if, for example, you were to measure the speed of sound as you travel away from its source; as your velocity gets closer to the speed of sound, the slower the sound waves appear to be travelling.

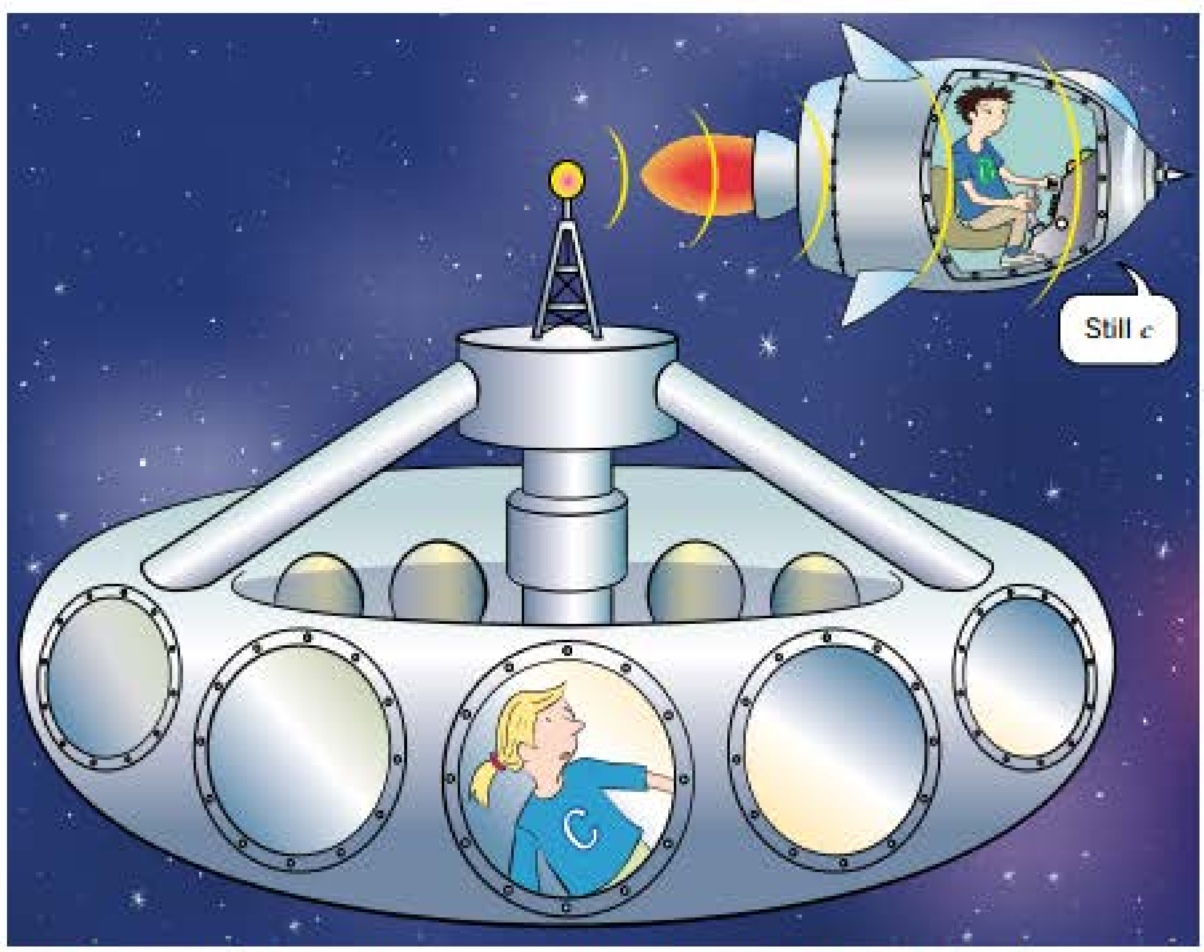


FIGURE 8.1.5 Einstein's two postulates are seemingly contradictory. His first postulate indicates that the speed of the laser light, as measured by Binh, should be $c - v$, whereas his second postulate indicates it should be c. Einstein revisited Newton's assumptions to resolve this problem.

The second postulate, however, tells you that when Binh measures the speed of Clare's laser light, he will find it to be c; that is, $3.00 \times 10^8 \, \text{m s}^{-1}$. So at first glance, these two postulates appear to be mutually exclusive. To resolve this problem, Einstein went back to the assumptions on which Newton based his theories.

Newton's assumptions

In 1687, Isaac Newton published his famous *Principia*. At the start of this incredible work, which formed the basis for all physics in the following two centuries and beyond, he notes the following assumptions.

The following two statements are assumed to be evident and true:

- *Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external.*
- *Absolute space, in its own nature, without relation to anything external, remains always similar and immovable.*

Newton based all his laws on these assumptions of space and time being constant, uniform and straight. In Newton's space you would expect a 1-metre rule to be the same length whether it is held vertically or horizontally, north–south or east–west, in your classroom or in the International Space Station. It would be the same length to anyone viewing it regardless of its position or velocity. This also meant that one second in one reference frame would be the same as one second in any other reference frame.

Einstein realised that these assumptions made by Newton might not be valid, at least not on scales involving huge distances and speeds approaching that of light. He also realised that the only way his two postulates could both be true was if space and time were not fixed and unchangeable. Based on his postulates, Einstein devised a form of mechanics called relativistic mechanics to distinguish it from that based on Newton's laws, now called classical mechanics. It turned out that classical mechanics was a good approximation of relativistic mechanics for the vast majority of everyday phenomena because the velocities involved were well below that of light. The whole theory created on the basis of these postulates that makes reference to inertial frames was called the special theory of relativity.

Einstein's *Gedanken* train

To illustrate the consequences of accepting the two postulates he put forward, Einstein discussed a simple thought experiment. It involves a train, moving at a constant velocity.

Amaya and Binh have boarded Einstein's *Gedanken* train and Clare is outside it, beside the track (Figure 8.1.6). This train has a flashing light bulb set right in the centre of the carriage. Amaya and Binh are watching the flashes of light as they reach the front and back walls of the carriage. They find that the flashes reach the front and back walls at the same time, which is not surprising. Outside, Clare is watching the same flashes of light. Einstein was interested in when Clare saw the flashes reach the end walls.

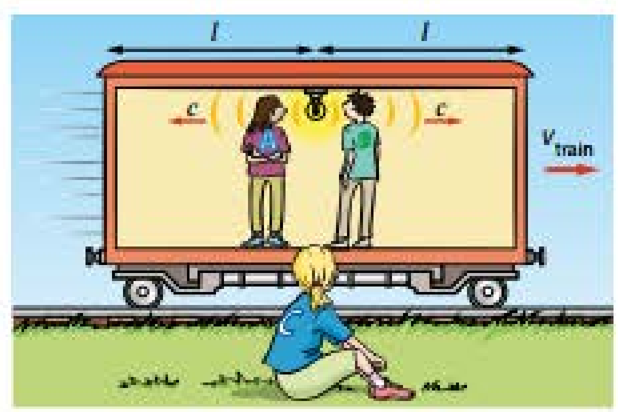


FIGURE 8.1.6 Amaya and Binh see the light take the same time, $\frac{l}{c}$ seconds, to reach the front and back walls.

To appreciate Einstein's ideas, you need to contrast them with what you would normally expect. Consider a situation in which Amaya and Binh are rolling balls towards opposite ends of a train carriage. It is important to appreciate that, while Clare, the outside observer, sees the ball's velocity differently from Amaya and Binh, the times at which various events (balls hitting the ends of the carriage) occur must be the same.

If you had discussed a pulse of sound waves travelling from the centre of the train, you would find exactly the same result: Clare always agrees with Amaya and Binh that the time taken for the balls, or sound waves, to reach the end walls is the same. But what about light?

Einstein's second postulate tells you that all observers see light travel at the same speed. Amaya, Binh and Clare will all see the light travelling at $3.00 \times 10^8 \, \text{m s}^{-1}$; they do not add or subtract the speed of the train.

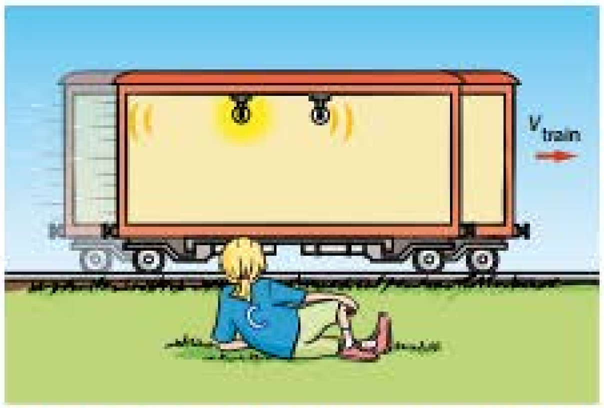


FIGURE 8.1.7 Clare sees the light reach the back wall first, and then the front wall.

PHYSICSFILE

Measurement in a thought experiment

The people in the *Gedanken* train would need extremely good measuring devices, such as an atomic clock, and amazingly quick reflexes to take their measurements.

Under normal circumstances, there is no chance of detecting the lack of simultaneity of light beams hitting the front and back walls of a train. This is because the differences in time are about a millionth of a microsecond, well beyond the capacity of even the best stopwatches. The reflexes required to see the light reach the back wall and then see the light encounter the front wall would also be beyond human ability.

If Clare sees the light travelling at the same speed in the forward and backward directions, she will see the light reach the back wall first (refer to Figure 8.1.7). This is because that wall is moving towards the light, whereas the front wall is moving away from the light, and so the light will take longer to catch up to it. This is against the principles of Newtonian physics. Amaya and Binh saw the light flashes reach the ends of the carriage at the same time; Clare saw them reach the walls at different times.

The idea that two events that are **simultaneous** (occur at the same time) for one set of observers but are not simultaneous for another is outrageous.

Simultaneity and spacetime

The big difference between the situation for light and that for balls, or sound, is the strange notion that both sets of observers see the speed of light as exactly the same. The velocity of a thrown ball or the velocity of sound in Amaya and Binh's frame of reference will always be different from that in Clare's frame of reference by exactly the velocity of the train. For light, however, there is no difference. As a result, events that are simultaneous for one set of observers are *not simultaneous* for the others. This is a very strange situation that is referred to as a lack of simultaneity.

Although Einstein's *Gedanken* experiments are purely hypothetical, other experiments based on these ideas are well within the capacity of modern experimental physics. In all cases they confirm Einstein's ideas to a high degree of accuracy.

Einstein said that the only reasonable explanation for how two events that were simultaneous to one set of observers were not simultaneous to another, is that time itself is behaving strangely. The amount of time that has elapsed in one frame of reference is not the same as that which has elapsed in another (Figure 8.1.8).

In the example shown in Figure 8.1.7, Amaya and Binh saw the light flashes that went forwards and backwards take the same time to reach the walls. In Clare's frame of reference the times were different. Time, which has one dimension, seems to depend on the frame of reference in which it is measured, and a frame of reference is just a way of defining three-dimensional space. Clearly time and space are somehow interrelated. This four-dimensional relationship, which includes the three dimensions of space and the one dimension of time, is called **spacetime**. Special relativity is all about spacetime.

This was a profound shock to the physicists of Einstein's time. Many of them refused to believe that time was not the constant and unchanging quantity that it was assumed always to have been. And to think that it might 'flow' at a different rate in a moving frame of reference was too mind-boggling for words. That could mean that if you went for a train trip, your clock would go slower, and you would come back having aged slightly less than those who stayed behind.

Einstein's idea was that time and distance are relative. They can have different values when measured by different observers. Simultaneous events in one frame of reference are not necessarily simultaneous when observed from another frame of reference. This is difficult to comprehend at first and will take some time to fully appreciate. Our basic understanding of time and distance (and perhaps mass too) needs adjustment when objects travel at close to the speed of light. A certain observer might see light travelling through a distance d in a time t at a speed c. A different observer might see light travelling through a different distance, d', in a different time, t', but still at the same speed, c.

Probably because of the tiny differences involved and the highly abstract nature of the work, many physicists simply disregarded the concepts and got on with their work. They thought it could never have any practical results.

FIGURE 8.1.8 The famous clock tower in Bern, Switzerland, near Einstein's apartment. Its hands move at one minute per minute, but only in the same frame of reference as the clock.

OBSERVATIONS THAT NEWTON'S LAWS CAN'T EXPLAIN

With the invention of more accurate measuring devices for time and distance, it became evident that some of the measurements of events didn't agree with the predicted values. These predicted values were based on Newton's laws acting in a framework of Galilean relativity.

Atomic clocks

Measuring time is an exercise in precision, replicating an interval of one second over and over again, until 86 400 (24 × 60 × 60) of them equals the time for one rotation of the Earth, or one day. There have been many mechanical solutions to this problem in the past using cogs and levers, weights and dials. The accuracies of these devices varied, with some of them gaining or losing seconds or minutes per day.

To correct your clock you would need to frequently adjust it against a standard clock. To help in this recalibration, radio stations would broadcast a time signal, so you could set your clock each day. Typically they would broadcast a series of five beeps counting down to each hour. You could also phone a number that would tell you 'at the tone it will be six o'clock … beep'.

For scientists, clocks with this level of inaccuracy could only be reliable for measuring events to one or two decimal places, which is fine for verifying relatively slow motion. Such clocks could not differentiate between two events occurring over a much smaller time interval.

Before 1967, the standard of one second was based on a fraction of the time it took for the Earth to orbit the Sun, a far-from-ideal standard. From 1967 onwards, the basis for the unit of time was changed to be a certain number of transitions of the outermost electron of a caesium-133 isotope. In fact, one second is now defined as: 9 192 631 770 oscillations of the 6s electron (the valence electron) of the Cs-133 isotope. The remarkable precision of this oscillation resulted in atomic clocks (the first of which is shown in Figure 8.1.9) with an accuracy of 1 second in 1.4 million years, and the ability to measure time to an incredible number of decimal places. It is at these levels of measurement that the predictions of Newton's laws of motion vary from the measured values.

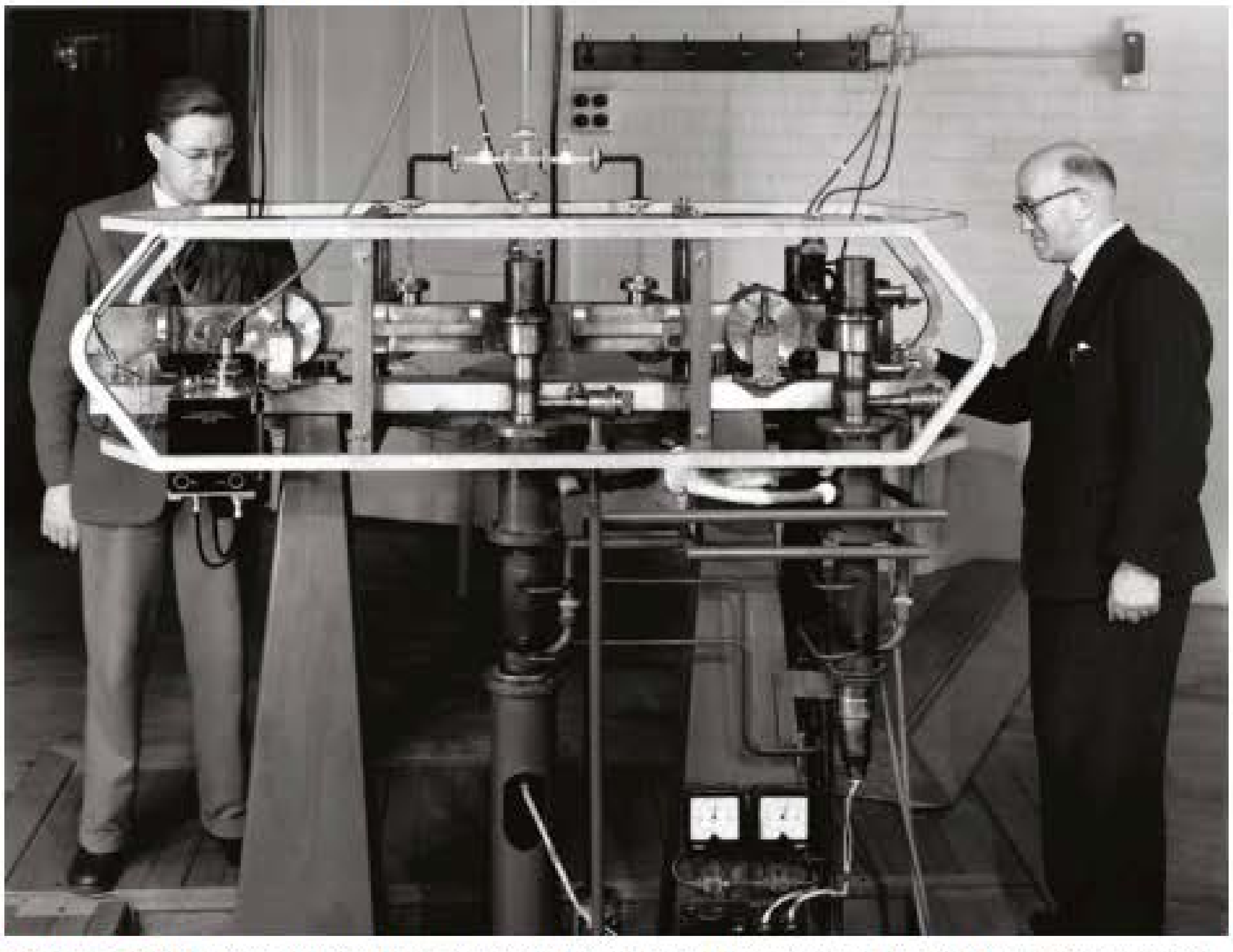


FIGURE 8.1.9 The first atomic clock, developed in 1955, and used to set the standard of one second from 1967.

Long-lived muons

When certain unstable subatomic particles (such as pions, which have a precisely known decay rate) are accelerated to almost the speed of light, their life spans are measured to be longer than when the particles are stationary. For example, the mean lifetime of the positive pion, π^+, is 0.000 000 026 033 s (26.033 ns), when it is stationary relative to the atomic clock that is measuring it. However, when it is moving at 99% of the speed of light, its mean lifetime as measured by the stationary atomic clock is 184.54 ns. This means that the moving pion exists seven times longer than a stationary pion.

High-energy cosmic rays interact with the nuclei of oxygen atoms 15 km above the surface of the Earth to create a cascade of high-velocity subatomic particles. One of these particles is a muon, which is unstable. The mean lifetime of a stationary muon, as measured by an atomic clock, is 0.000 002 196 s (2.196 μs). The muons created by cosmic radiation typically travel at 99.97% of the speed of light, so at this speed Newtonian physics would predict that a muon would travel about 659 m:

$$s = v\Delta t$$
$$= 0.9997 \times 3 \times 10^8 \times 2.196 \times 10^{-6}$$
$$= 658.6\text{ m (or roughly 659 m)}$$

After 10 lifetimes, you can expect there to be essentially no muons remaining. So after beginning at a height of 15 km and travelling through a distance of 6.58 km, to a height of about 8.42 km above the surface of the Earth, you would expect that no muons would be detected on Earth.

However, muons created by cosmic radiation are actually detected at the surface of the Earth. This means that the fast-moving muons have existed for a much longer period of time than they should have. A muon that strikes the surface would have existed at least 22.8 times its predicted life span as a stationary muon, based on Newtonian physics. Once again, Newtonian physics and Galilean relativity cannot explain this observation. Section 8.2 'Time dilation' explains why this happens.

PHYSICS IN ACTION

Particles gaining mass

When an object moves in a circular path, it does so as a result of a centripetal force that acts towards the centre of the circular path. Centripetal force therefore acts continuously at a right angle to the velocity of the object. There are a number of actions that could cause the centripetal force on an object, such as the tension in a string tied to a rubber stopper or the gravitational force of the Earth on the Moon.

Another action that causes circular motion is the force on a charged particle that is moving at right angles to a magnetic field. The equation that represents the relationship between the magnetic force (F_B) and the centripetal force (F_C) is:

$$F_B = F_C$$
$$qvB = \frac{mv^2}{r}$$
$$r = \frac{mv}{qB}$$
$$r = \frac{m}{qB} \times v$$

The final relationship shows that, if the mass m, charge q and magnetic field B are all constant, then the radius of the circular path is directly proportional to the velocity of the charged particle. So, theoretically, if the velocity increases by a factor of 2, then the radius will also increase by a factor of 2. However, this is not the case.

In circular accelerating devices, such as cyclotrons and the Australian Synchrotron, it is evident that, as the velocity of a charged particle increases, the radius of its path also increases, but to a much greater degree than that expected. According to the relationship shown above, if the charge q and the magnetic field B don't change, then the only explanation for the extra increase in radius is that the mass of the particle m must increase.

In fact, the mass of an electron travelling at 99.999 99% of the speed of light seems to increase to 6000 times the mass of an electron at rest. There is no explanation for this phenomenon in Galileo's relativity or Newtonian physics. This will all be explained later in the chapter, using Einstein's theories.

EXTENSION

Real-world relativity

The global positioning system

The global positioning system (GPS) is a satellite system that is now part of everyday life. For it to function accurately, adjustments need to be made to its component satellites to compensate for the effects of both special and general relativity. There are many uses for GPS including aircraft and motor vehicle navigation, driverless vehicles, crop planting, spraying and fertilising, both military and civilian drones, and the control of autonomous trucks in the mining industry. The GPS device provides accurate real-time information on position, speed and direction of travel on the Earth's surface.

The GPS system is configured using a network of satellites that are in various orbits around the Earth at an altitude of just over 20000 km, a speed of around 3900 $m\,s^{-1}$ and a period of around 12 hours (Figure 8.1.10). The satellites are placed in various orbital planes that are set so that four or more of them are visible to any point on the Earth's surface at any one time. Their orbits are not geosynchronous; that is, they do not stay above the same point on the Earth's surface.

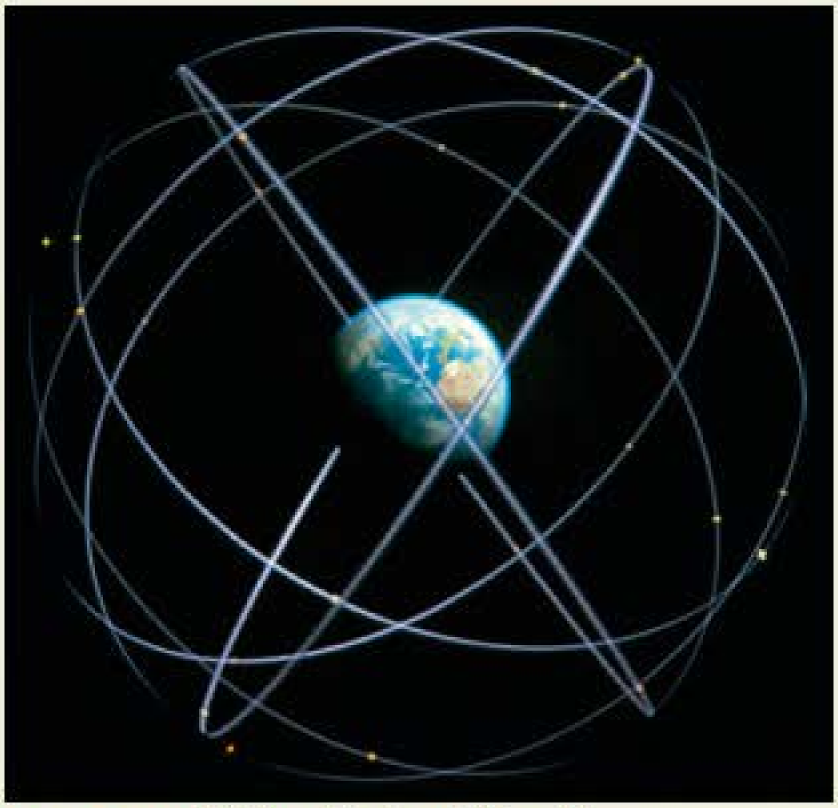


FIGURE 8.1.10 Global positioning satellite orbits.

Each satellite is fitted with an atomic clock that has accuracy in the order of one nanosecond per day. It is the comparison of the time signal received from each satellite visible to the GPS device that determines the position, speed and course of the device. The resultant data received is compared based on the accurately known position of each satellite by a process called trilateration and an absolute position on the Earth's surface determined within 5–10 metres in just a few seconds.

For positional accuracy to be achieved, the time on each GPS satellite must be known to within 20–30 nanoseconds. To achieve such accuracy at the speed and height at which these satellites are travelling, the effects of both special and general relativity must be taken into consideration. The satellite is travelling at a high velocity relative to any point on the Earth's surface, so special relativity would have the GPS device 'see' the satellite's atomic clock running slower than the same clock on Earth (time dilation). Calculations show this to be about 7 μs per 24 hours.

General relativity predicts that a clock closer to a massive object will 'tick' slower than one that is further away from that object. The implication of this is that the clock in orbit on the satellite will suffer a time contraction (tick faster) relative to the same clock on Earth. The atomic clock on the satellite will be ahead of the atomic clock on Earth. This results in a time gain on the satellite of about 45 μs in 24 hours. This phenomenon exists because the curvature of spacetime results in time running slower in a stronger gravitational field. The further you are from a massive object, such as the Earth, the faster time will pass. A discussion of this concept is beyond this course at this time.

Every GPS satellite is therefore affected by two separate relativistic effects with opposite consequences. One gives a gain in 24 hours of around 45 μs, the other a loss of around 7 μs. The resultant gain is therefore around 38 μs per 24-hour period. Although this does not seem to be much of a difference, in terms of time measurements made on Earth, it introduces a very significant and serious error when applied to the GPS satellites. As the signal travels at the speed of light, the resultant positional error to an Earth-based GPS device would be about 10 km in this 38 μs period.

Before each satellite is launched the engineers and technicians who design and work on global positioning systems include parameters that compensate for the general relativistic effects. They do this by slowing down the rate of the satellite's atomic clock. The special relativity requirements are, however, not fully known until the satellite is in its fixed orbit. Once the satellite is in orbit these parameters are adjusted by an inbuilt computer to achieve the necessary clock speed.

Relativity is therefore no longer an abstract theory but a reality applied to many situations.

EXTENSION

Gravity waves

Einstein's general theory of relativity predicted the existence of gravity waves that carry energy and angular momentum as ripples through spacetime at the velocity of light. These waves were considered capable of deflecting a beam of light (photons) from a distant star when it was in the gravitational field of a large mass such as the Sun. The deflection calculated using Einstein's theory gave an answer that was greater than that from Newton's gravity; a result not unexpected because Newton's physics was developed around non-relativistic speeds.

Einstein's theory was purely theoretical and at that time had no empirical evidence to support it. Results obtained in 1919 from the observation of a solar eclipse gave empirical evidence that light from a distant star underwent deflection as predicted by Einstein's calculations. These results were seriously questioned at the time, but were substantiated by Professor Alexander Ross, professor of physics at the University of Western Australia, when he led a group to do similar observation of a solar eclipse in 1922 at Wallal near Broome, Western Australia.

In 1974, observations of a pair of neutron stars in close orbit about each other were made by Russell Hulse and Joe Taylor from the University of Massachusetts. They found that the period of orbit of this star pair (measured extremely precisely) about each other was decreasing. These observations agreed with the predictions of Einstein's theory that orbiting stars would spiral in towards each other as they radiated gravity waves. Taylor and Hulse received the Noble Prize in Physics in 1993 for the work they did on the energy loss from binary pulsars and their resultant proof from this of the existence of gravity waves.

In 1995 Australia established a national consortium involving the Australian National University, the University of Western Australia, Adelaide University and the CSIRO to develop laser interferometry technology and join a world consortium, called LIGO (Light Interferometer Gravitational Observatory), searching for gravity waves. The Australian facility was set up at Gingin, 100 km north of Perth, in 1999. In September 2015 success was achieved by LIGO when gravity waves were detected almost simultaneously at the USA sites of Hanford and Livingston. Fifty-six Australian scientists were involved in this discovery.

8.1 Review

SUMMARY

- Galileo struggled to change long-held ideas about the solar system and the motion of objects under the influence of forces.
- Galileo suggested that the natural state of objects is not to be at rest, but to be in a state of uniform motion.
- Galilean relativity holds that there is nothing special about zero velocity, and that a force acting for a period of time will cause the same change in the velocity of an object, no matter what the initial velocity might be.
- Galilean relativity also states that a change in velocity will be identical, whether measured from a stationary frame of reference or a non-zero inertial frame of reference.
- Newtonian physics incorporated the concept of inertial frames of reference from Galilean relativity.
- Einstein decided that Galileo's principle of relativity was so elegant it simply had to be true, and he was also convinced that Maxwell's electromagnetic equations, and their predictions, were sound.
- Einstein's two postulates of special relativity can be abbreviated to:
  1. The laws of physics are the same in all inertial frames of reference.
  2. The speed of light is the same for all observers.
- Einstein realised that accepting both of these postulates implied that space and time were not absolute and independent, but were related in some way.
- Two events that are simultaneous in one frame of reference are not necessarily simultaneous in another.
- This implies that time measured in different frames of reference might not be the same. Time and space are related in a four-dimensional universe of spacetime.
- Observations of the lifetimes of subatomic particles that are accelerated to high speeds indicate that they exist for longer than when they are stationary.
- High-speed muons created in Earth's upper atmosphere should not last long enough to reach Earth's surface, but they do. The moving muons have longer lifetimes than stationary muons.

KEY QUESTIONS

1 Based on Galileo's theories, which of the following is correct?

A A force is required to accelerate an object.
B No force is required to change the velocity of an object.
C A force is required to maintain an object at zero velocity.
D A force is required to maintain the velocity of an object.

2 Which of the following is the best definition of an inertial frame of reference?

A one in which the object is stationary
B one in which the object is accelerating
C one in which the object is moving at a constant velocity
D one in which the object is undergoing constant force

3 Why did the physicists of the late 19th century feel the need to invent the idea of the aether?

A It was required to satisfy the principle of relativity.
B It was required to satisfy Maxwell's equations.
C They thought that it would be impossible that totally empty space could occur in nature.
D They thought that there should be a medium that carries light waves just as air carries sound waves.

4 Which of the following is/are reasonably good inertial frames of reference? More than one answer is possible.

A an aircraft in steady flight
B an aircraft taking off
C a car turning a corner
D a car driving up a hill of constant slope at a steady velocity

8.1 Review *continued*

5 Calculate the velocity, relative to the ground, of a cattle dog running north at 1.80 m s^{-1} on the back of a moving cattle truck moving at 22.25 m s^{-1} south.

6 Two spaceships are travelling for a while with a constant relative velocity. Then one begins to accelerate. A passenger with a laser-based velocity measurer sees the relative velocity increase. How could this passenger tell whether it was his own or the other ship that began to accelerate?

7 Tom, who is in the centre of a train carriage moving at constant velocity, rolls a ball towards the front of the train, while at the same time he blows a whistle and shines a laser towards the front of the train. Compared with Tom, what will Jana, who is on the ground outside the train, observe regarding the speed of the ball, the sound and the light?

8 If the speed of sound in air is 340 m s^{-1}, at what speed would the sound from a fire truck siren appear to be travelling in the following situations?

a You are driving towards the stationary fire truck at 30 m s^{-1}.

b You are driving away from the stationary truck at 40 m s^{-1}.

c You are stationary and the fire truck is heading towards you at 20 m s^{-1}.

d You are driving at 30 m s^{-1} and draw level with the siren as you overtake the fire truck, which is travelling at 20 m s^{-1} in the same direction.

9 In order to resolve the apparent conflict resulting from his two postulates, Einstein rejected some of Newton's assumptions. Which of the following statements is a consequence of this?

A Time is not constant in all frames of reference.

B Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external.

C One second in any inertial frame of reference is the same as one second in any other inertial frame of reference.

D Space and time are independent of each other.

10 Anna is at the front end of a train carriage moving at 10 m s^{-1}. She throws a ball back to Ben, who is 5 m away at the other end of the carriage. Ben catches it 0.2 s after it was thrown. Chloe is watching all this from the side of the track.

a At what velocity does Chloe see the thrown ball travelling?

b How far, in Chloe's frame of reference, did the ball move while in flight?

c How long was it in flight in Chloe's frame of reference?

11 Imagine that the speed of light has suddenly slowed down to only 50 m s^{-1} and this time Anna (still at the front of the 5 m-long train carriage moving at 10 m s^{-1} in Question 10) sends a flash of light towards Ben.

a From Anna's point of view, how long does it take the light flash to reach Ben?

b How fast was the light travelling in Ben's frame of reference?

c In Chloe's frame of reference, how far did the train travel in 0.1 s?

d How fast was the light travelling in Chloe's frame of reference?

e Approximately, when did Chloe see the light reach Ben?

12 Why was the development of atomic clocks important to the advancement of Einstein's special theory of relativity?

13 Complete the following sentences by selecting the correct term in bold.

Muons have **very short/prolonged** lives. On average, muons live for approximately 2.2 **s/μs**. Their speeds are measured as they travel through the atmosphere. A muon's speed is **about a tenth of/very similar to** the speed of light. According to Newtonian laws, muons **should/should not** reach the Earth's surface. However, many **do/do not**.

8.2 Time dilation

Extremely precise atomic clocks like that shown in Figure 8.2.1 enable very short-lived events to be measured to a large number of decimal places. At this level of precision, some unusual observations have been made regarding the life spans of some high-speed subatomic particles when compared to the life spans of those same particles at rest. This section explores the concept of time dilation as an explanation for these observations.

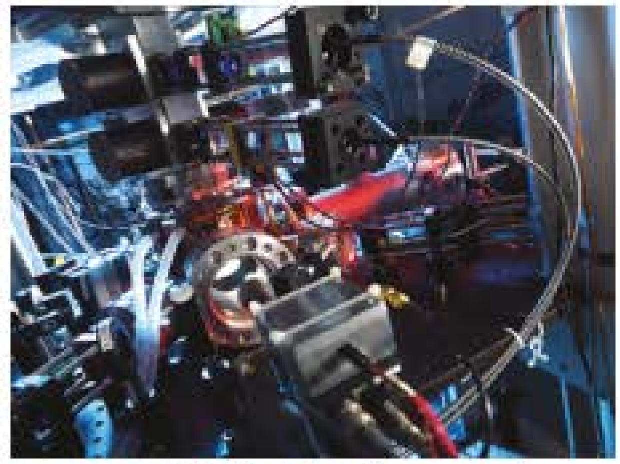


FIGURE 8.2.1 The duration of one second can be measured very precisely using a caesium atomic clock like this one.

TIME IN DIFFERENT FRAMES OF REFERENCE

The consequences of Einstein's two postulates have been discussed, in general terms, when they are applied to a simple *Gedanken* situation, such as a moving train. Observers inside the train see two simultaneous events while those outside see the same two events occurring at different times. Certainly the differences are extremely small and would not be noticeable by an observer in any actual train, unless they had an atomic clock. For aircraft flying at supersonic speeds, the differences, although very small, become measurable by the most precise clocks. For subatomic particles, such as pions in accelerators like the Australian Synchrotron, the differences in time become more significant, and so in situations like this, in which speeds approach the speed of light, it is important to use calculations that take Einstein's theory into account.

The light clock

If you want to observe **time dilation** in the moving train or among moving subatomic particles, you need to watch a clock in a moving reference frame to see if it is actually going slower. The term 'dilation' in this context means slower.

Consider Amaya and Binh riding in a *Gedanken* spaceship that can travel at speeds close to the speed of light. Clare is going to watch from a space station that, according to Clare, is a stationary frame of reference. Amaya and Binh have taken along a clock, which (it is assumed) Clare can read, even from a long distance away.

Like any clock, this clock is governed by a regular oscillation that defines a period of time.

Amaya's *Gedanken* clock has a light pulse that bounces back and forth between two mirrors. One mirror is on the floor and the other on the ceiling, as shown in Figure 8.2.2. When a light pulse oscillates from one mirror to the other and back, you can consider that period of time to be 'one unit'. Clare has an identical clock in her own space station, which she can compare to Amaya's clock.

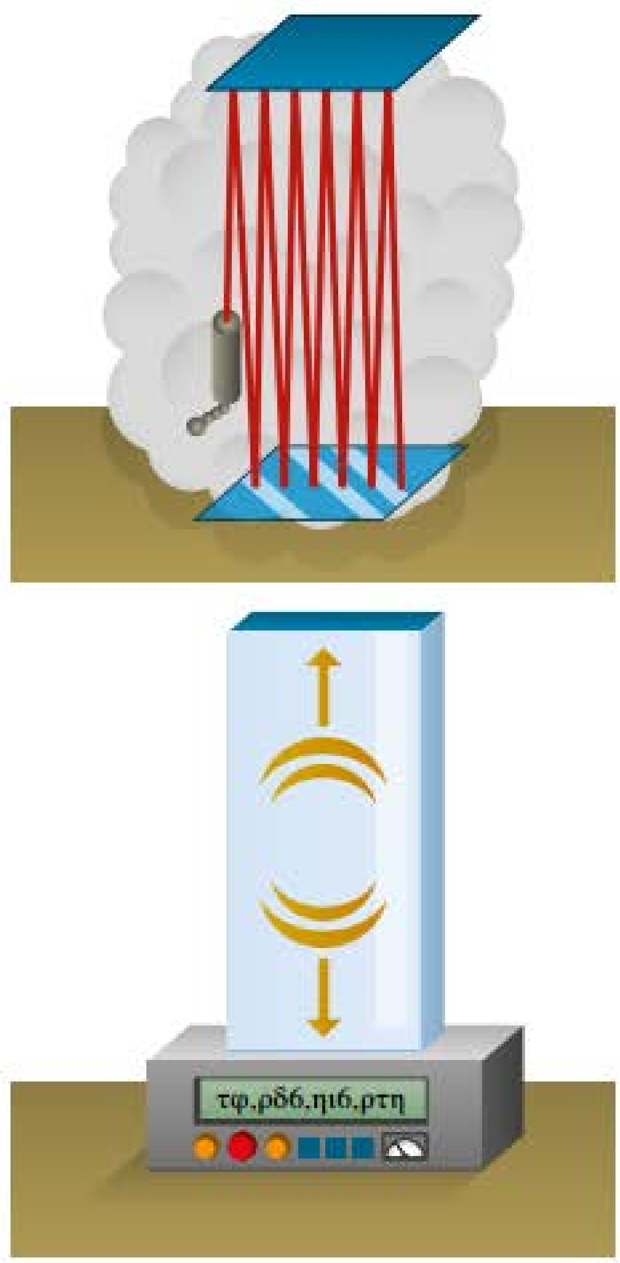


FIGURE 8.2.2 The *Gedanken* light clock 'ticks' each time the light pulse reflects off the bottom mirror.

The advantage of this clock is that it can be used to predict how motion will affect it by using Pythagoras' theorem and some algebra. The clock has been set up in the spaceship so that the light pulses oscillate up and down a distance d that is at right angles to the direction of travel. The distance d is shown by a black arrow in the centre position of the moving spacecraft in Figure 8.2.3 (page 318). As the spaceship speeds along, the light will trace out a zigzag path, as shown by the red dotted line in Figure 8.2.3.

Only one of the oscillations of the light pulse needs to be considered, as all the other oscillations will have the same geometry. One 'unit of time' will be the time taken for the light pulse to oscillate once. In the frame of reference of the spaceship, Amaya and Binh see a unit of time equal to t_a. Clare, from her frame of reference, will see a different time, t_c. The relationship between these two times will now be determined.

Amaya and Binh see the light pulse travel at the speed of light, c, along the distance $2d$, from the bottom mirror to the top and back again in time t_a. So the distance that the light pulse travels is given by:

$$2d = c \times t_a$$

However, Clare sees the light travel a longer path that is shown as the red dotted line in Figure 8.2.3.

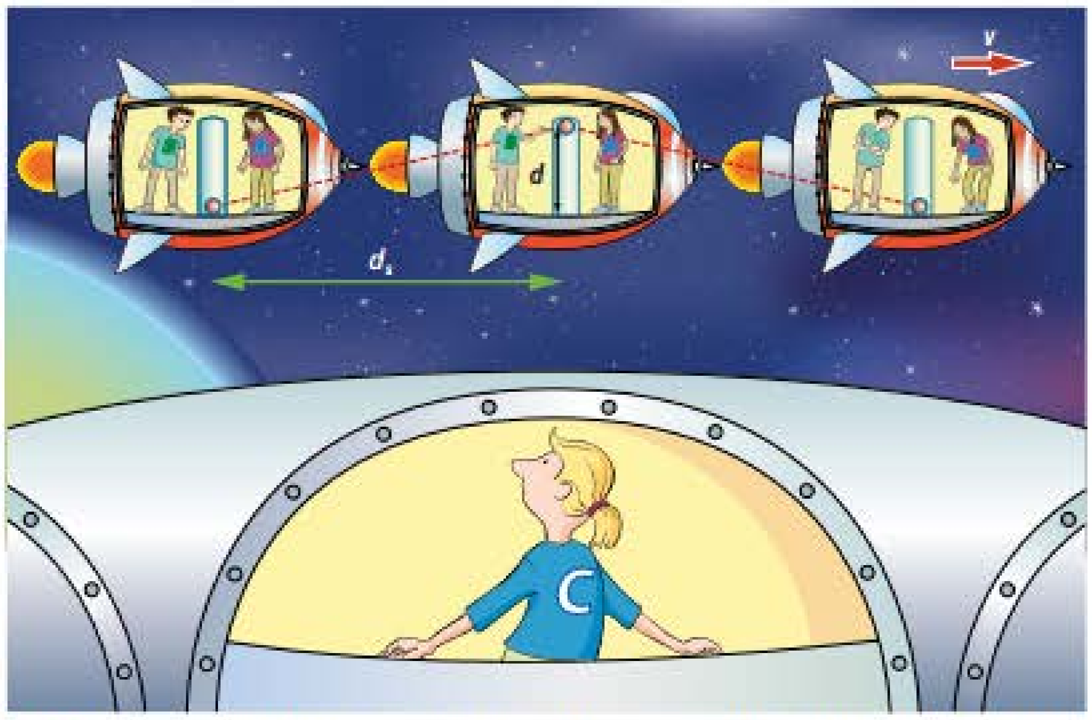


FIGURE 8.2.3 Clare can see that in one unit of time the *Gedanken* light clock 'ticks' each time the light pulse reflects off the bottom mirror. She also sees that the light pulses travel a zigzag path between the mirrors.

The ship moves with a speed v, and so in one unit of time as seen by Clare, t_c, the spaceship will travel a distance $2 \times d_s$, equal to the velocity multiplied by the time taken for Clare to see one oscillation:

$$2d_s = v \times t_c$$

Consider only half of the light oscillation for now. The light pulse not only travels the vertical distance d in the clock, but also travels forwards as the spaceship moves through the distance d_s, making the combined distance d_c. Therefore, according to Pythagoras's theorem:

$$d_c^2 = d^2 + d_s^2$$

$$d_c^2 = d^2 + \left(\frac{vt_c}{2}\right)^2$$

$$d_c = \sqrt{\left(d^2 + \left(\frac{vt_c}{2}\right)^2\right)}$$

Clare sees this light pulse travelling twice this combined distance at the speed of light c, in a period of time t_c measured on her clock. So:

$$2d_c = c \times t_c$$

Equating and rearranging the two expressions for d_c gives:

$$\frac{c \times t_c}{2} = \sqrt{\left(d^2 + \left(\frac{vt_c}{2}\right)^2\right)}$$

$$c \times t_c = 2 \times \sqrt{\left(d^2 + \left(\frac{vt_c}{2}\right)^2\right)}$$

$$c \times t_c = \sqrt{\left(4d^2 + 4 \times \left(\frac{vt_c}{2}\right)^2\right)}$$

$$t_c = \frac{\sqrt{4d^2 + (vt_c)^2}}{c}$$

From Amaya and Binh's frame of reference, in which they see the light pulse travelling a distance $2d$ at speed c in a time t_a, the previously given equation can be rewritten in terms of d as:

$$d = \frac{c \times t_a}{2}$$

Note that you have used the same value for c in both of these equations, something you would never do in **classical physics**, but something Einstein insists you must do.

Substituting this expression for d into the previous equation gives:

$$t_c = \frac{\sqrt{4\times\left(\frac{ct_a}{2}\right)^2 + (vt_c)^2}}{c}$$

Now square both sides and simplify to make t_c^2 the subject:

$$t_c^2 = \frac{\frac{4(ct_a)^2}{4} + (vt_c)^2}{c^2}$$

$$t_c^2 = \frac{c^2t_a^2 + v^2t_c^2}{c^2}$$

$$t_c^2 = \frac{c^2t_a^2}{c^2} + \frac{v^2t_c^2}{c^2}$$

$$t_c^2 = t_a^2 + \frac{v^2t_c^2}{c^2}$$

Group the terms with t_c^2 together and factorise:

$$t_c^2 - \frac{v^2t_c^2}{c^2} = t_a^2$$

$$t_c^2\left(1-\frac{v^2}{c^2}\right) = t_a^2$$

Take the square root of both sides and make t_c the subject:

$$t_c\sqrt{\left(1-\frac{v^2}{c^2}\right)} = t_a$$

$$t_c = \frac{t_a}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}}$$

As v can never be larger than c, the denominator in the equation above must be less than one. Any number divided by a number less than one must result in a larger number, so $t_c > t_a$

This final equation shows that the time that Clare measures, t_c, is greater than the time that Amaya and Binh measure, t_a, for the same event.

PHYSICSFILE

The zigzag path of light

Mathematically, you can see that time dilation results from the strange behaviour of light. As light travels on the diagonal zigzag path, it does so at speed c, not at a faster speed resulting from the additional component of the spaceship's motion as, for example, would be true for a boat zigzagging across a river as it is carried along by the current.

DILATION OF TIME

In Einstein's equation for time dilation, the symbol t is used to represent the time that a stationary observer (Clare) measures using a stationary clock for an event that the observer sees occurring in a moving frame of reference. The symbol t_0 is then the time that passes on the moving clock, which is also known as the **proper time**.

The factor by which the proper time is multiplied is given the symbol gamma, γ, so that:

$$t = t_0\gamma$$

where $\gamma = \frac{1}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}}$

v is the speed of the moving frame of reference

c is the speed of light in a vacuum (3×10^8 m s^{-1})

t is the time observed in the stationary frame

t_0 is the time observed in the moving frame (proper time)

The physicist H.A. Lorentz first introduced the factor γ in an attempt to explain the results of the Michelson–Morley experiment, so it is often known as the **Lorentz factor**.

Table 8.2.1 and Figure 8.2.4 show the effect of varying the value of v on the value for γ.

From the data in Table 8.2.1, a velocity of $300\,\text{m}\,\text{s}^{-1}$ results in a Lorentz factor of essentially 1. So for relatively low-speed spaceships, a stationary observer measures the oscillation of light in the light clock on the spaceship to be the same as in her own stationary light clock. This implies that time is passing at essentially the same rate in both frames of reference.

TABLE 8.2.1 The value of the Lorentz factor at various speeds

v ($\text{m}\,\text{s}^{-1}$)	$\frac{v}{c}$	γ
3.00×10^2	0.000001	1.000000000
3.00×10^5	0.00100	1.0000005
3.00×10^7	0.100	1.005
1.50×10^8	0.500	1.155
2.60×10^8	0.866	2.00
2.70×10^8	0.900	2.29
2.97×10^8	0.990	7.09
2.997×10^8	0.999	22.4

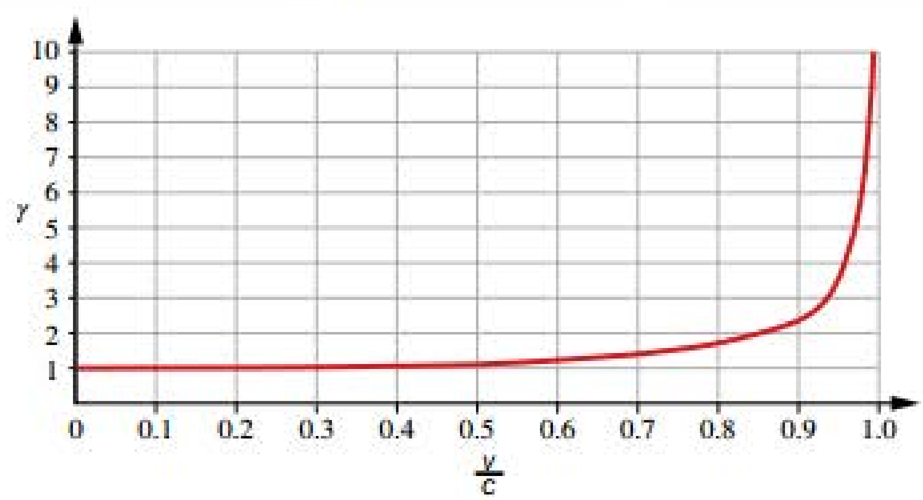


FIGURE 8.2.4 The graph of the Lorentz factor versus $\frac{v}{c}$.

EXTENSION

The Lorentz factor

The Lorentz factor becomes so close to 1 for values of v less than about $0.0001c$ that the expression can't be used with a normal calculator. Fortunately, there is a simple way to find the value of γ for speeds less than about 1% of c. The binomial expansion of the term $(1 - x)^n$ tells you that:

$(1 - x)^n \approx (1 - nx)$ provided that $x \ll 1$.

For the Lorentz factor this means that:

$$x = \left(\frac{v}{c}\right)^2 \text{ and } n = -\frac{1}{2}$$

and so:

$$\gamma = 1 + \frac{1}{2}\left(\frac{v}{c}\right)^2$$

Thus the part of the factor that is greater than 1 can simply be found from the term:

$$\frac{1}{2}\left(\frac{v}{c}\right)^2$$

Sometimes it is useful to make v the subject in the equation for the Lorentz factor. This produces:

$$v = c\sqrt{\left(1 - \frac{1}{\gamma^2}\right)}$$

When the spaceship is travelling at 0.990c, a stationary observer like Clare will measure that a single oscillation of light in the spaceship's light clock will take seven oscillations of her own stationary light clock. According to Clare, time for the objects and people in the moving frame of reference has slowed down to one-seventh of 'normal' time.

As the speed approaches the speed of light, time in the moving frame, as viewed from the stationary frame, slows down more and more. So, if you were able to see the clock travelling on a light wave, the clock would not be 'ticking' at all. In other words, time would be seen to stand still.

It is important to realise that Amaya and Binh do not perceive their time slowing down at all. To them, their clock keeps ticking away at the usual rate and events in their frame of reference take the same time as they normally would. It is the series of events that Clare sees and measures in Amaya and Binh's frame that go slowly. Binh and Amaya are moving in slow motion because, according to Clare's observations, time for them has slowed down (Figure 8.2.5).

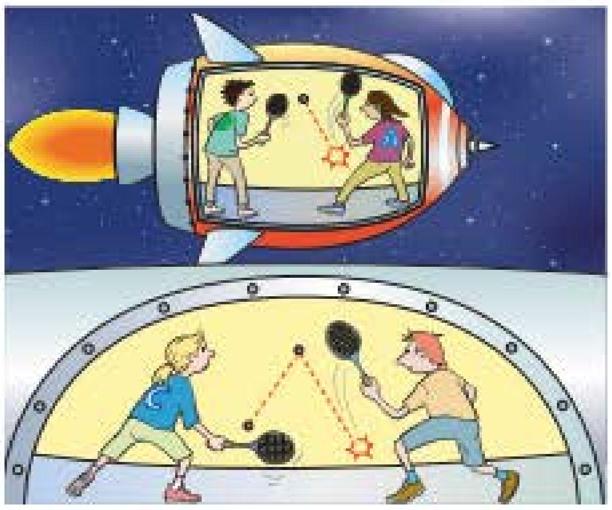


FIGURE 8.2.5 As Clare watches Amaya and Binh play space squash, the ball seems to be moving much more slowly than in her own game.

Worked example 8.2.1

TIME DILATION

Assume *Gedanken* conditions exist in this example. A stationary observer on Earth sees a very fast car passing by, travelling at $2.50 \times 10^8\,\text{m s}^{-1}$. In the car is a clock on which the stationary observer sees 3.00 s pass. Calculate how many seconds pass by on the stationary observer's clock during this observation. Use $c = 3.00 \times 10^8\,\text{m s}^{-1}$.

Thinking	Working
Identify the variables: the time for the stationary observer is t, the proper time for the moving clock is t_0 and the velocities are v and the constant c.	$t = ?$ $t_0 = 3.00\,\text{s}$ $v = 2.50 \times 10^8\,\text{m s}^{-1}$ $c = 3.00 \times 10^8\,\text{m s}^{-1}$
Use Einstein's time dilation formula and the Lorentz factor.	$t = t_0\gamma$ $= \dfrac{t_0}{\sqrt{1-\frac{v^2}{c^2}}}$
Substitute the values for t_0, v and c into the equation and calculate the answer, t.	$t = \dfrac{3.00}{\sqrt{1-\frac{(2.50\times10^8)^2}{(3.00\times10^8)^2}}}$ $= \dfrac{3.00}{0.55277} = 5.43\,\text{s}$

Worked example: Try yourself 8.2.1

TIME DILATION

Assume *Gedanken* conditions exist in this example. A stationary observer on Earth sees a very fast scooter passing by, travelling at $2.98 \times 10^8\,\text{m s}^{-1}$. On the wrist of the rider is a watch on which the stationary observer sees 60.0 s pass. Calculate how many seconds pass by on the stationary observer's clock during this observation. Use $c = 3.00 \times 10^8\,\text{m s}^{-1}$.

LOOKING BACK TO THE STATIONARY OBSERVER

So far you have been looking at the situation from Clare's point of view, not Amaya's and Binh's. Galileo said that all inertial frames of reference are equivalent. It follows then that, according to Amaya and Binh, as they look out their window at Clare in her space station receding from them, they can consider that it is they who are at rest and it is Clare in her space station moving away at a velocity near the speed of light. This is what Galileo's principle of relativity and Einstein's first postulate are all about.

If Amaya and Binh watch the light clock in Clare's space station, they see that time has slowed down for Clare, as they would observe Clare's moving light clock oscillation taking longer than their stationary light clock oscillation. This raises the question: Whose time actually runs slowly?

The answer is that they are both right. The whole point of relativity is that you can only measure quantities relative to some particular frame of reference, not in any absolute sense. Certainly Amaya and Binh see Clare as though in slow motion and Clare sees them in slow motion. Remember that there is no absolute frame of reference and so there is no absolute clock ticking away the absolute 'right' time. All that you can be sure of is that time in your own inertial frame of reference is ticking away at a rate of one second per second.

PHYSICSFILE

Is light slowing down?

Recently there has been publicity given to research that has suggested that the speed of light is slowing down. Some have even suggested that Einstein's theory of relativity itself is under threat. The research, based on analysis of light from very distant quasars, actually suggests that there have been very small changes in what is called the fine structure constant, which is made up of three more basic constants: the speed of light, the charge on an electron and Planck's constant.

Prominent theoretical physicist Professor Paul Davies and others have suggested that if the evidence is correct, then it is probably the speed of light that is changing. If proved correct, no doubt this new data will modify some aspects of relativity, but to suggest that it will overturn relativity is a wild exaggeration.

EXPLAINING HIGH-ALTITUDE MUONS

In Section 8.1 'Einstein's theory of special relativity', the surprising observation of high-speed muons originating 15 km up in the atmosphere and yet reaching the surface of the Earth was discussed. It can only be explained if the mean lifetime of the short-lived particles is extended far beyond their normal mean lifetime.

Time dilation provides the explanation to this unusual observation.

The 'normal' mean lifetime of a muon is about 2.2 μs. However, this is the mean lifetime when measured in a stationary frame of reference. Muons travel very fast; in fact a speed as great as $0.999c$ is possible. At this speed, an observer on Earth will see the lifetime of a muon as far greater than 2.2 μs:

$$t = t_0\gamma$$

$$= \frac{t_0}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$= \frac{2.2\times10^{-6}}{\sqrt{1-\frac{(0.999c)^2}{c^2}}}$$

$$= \frac{2.2\times10^{-6}}{\sqrt{1-0.999^2}}$$

$= 49.21\ \mu s$ (which is more than 22 times as long as in the stationary frame)

An observer on Earth would see the muon's time run much slower. The slower time means that many muons live long enough to reach the Earth's surface.

EXTENSION

The Twin Paradox

If Clare sees time for Amaya and Binh running slowly, then Amaya and Binh will age slowly. But if Amaya and Binh see that time for Clare has slowed down, then Clare will age more slowly. So what happens when Amaya and Binh decide to turn their spaceship around and come home? Who will have aged more?

To solve this paradox, or contradiction, Einstein described a thought experiment in which one of a set of twins heads off on a long space journey, while the other twin stays on Earth.

The travelling twin finds that when she returns, her remaining twin has become quite elderly (refer to Figure 8.2.6). Although each twin is in constant motion relative to the other, they both see the other twin ageing more slowly. So why did the twin on the spaceship age more slowly than the twin on Earth?

The key to this paradox is that only one twin has spent the entire time in a single inertial (non-accelerating) frame of reference. The other twin spent some time in two different inertial frames of reference, one on the way out and the other on the way back. The fact that the travelling twin experienced at least three different periods of acceleration or deceleration has very little effect on the difference in time for the two twins. This means that Einstein's theory of general relativity is not the major factor in the explanation of the twin paradox.

The explanation can be found by considering the Doppler effect on the frequency at which the clock signals sent by each twin are received by the other twin. Consider the situation of twins, Linda, who likes to stay at home, and Yvette, who likes to travel. Yvette leaves on a trip in a spaceship travelling at 80% of the speed of light. Each twin has a clock; but, instead of having standard clocks, each twin has a clock that sends out a light signal once a year. Linda's clock sends out a pulse of light at midnight on New Year's Eve that would eventually catch up to Yvette as she travels away from Earth. Yvette's clock also sends out a light signal at the start of her year.

There is a distinction that must be made here between the time that each twin experiences—Yvette's time will be dilated time and Linda's time will be proper time. These distinctions are not normally made, because you can't tell which one of two inertial frames of reference is moving relative to the other. In this case, however, we know that Yvette is the one travelling and therefore is observed to live in dilated time, while Linda is not moving and lives in proper time. We know this because in Yvette's spaceship there is a chandelier that was observed to hang down at an angle from the vertical during certain stages of the trip. This indicates that Yvette's spaceship was accelerating or decelerating, so she was the one moving. In a spaceship travelling at 80% of the speed of light, it would take three of Yvette's dilated years for the first of stay-at-home Linda's time signals to reach her travelling sister. Similarly, it would take three of Linda's proper years for Yvette's first time signal to reach the twin on Earth. This is consistent with the idea of relative motion, and that neither inertial frame of reference is special.

In the time that the travelling twin Yvette travels for 15 of her dilated years, she would have received five yearly time signals from the stay-at-home twin Linda. Linda would have received five time signals from Yvette in 45 proper years. At this point, Yvette decides to come home. Yvette turns her spaceship around and she is now in a different inertial frame of reference, but it will still take 15 of her dilated years to get home. Now, as Yvette is travelling back towards the approaching time signals sent by Linda, she will receive three time signals from Linda during each one of her dilated years. So in the 15 dilated years it takes for Yvette to travel back, she will receive 45 yearly signals from Linda. Linda, on the other hand, will receive three time signals from Yvette each proper year, so she will receive a total of 15 yearly signals in five proper years. The total trip for Yvette took 30 dilated years, 15 out and 15 back, so Yvette has aged by 30 years. Linda remained on Earth and had 50 birthdays, so she is now 20 years older than Yvette.

Although it is often called a paradox, there is actually nothing impossible or illogical about this story. Einstein himself pointed out that, due to the Earth's rotation, and therefore centripetal acceleration, a clock on the Earth's equator would run a little more slowly than one at the poles. This has now actually been found to be the case. In fact, in 1971 accurate atomic clocks were flown around the world on commercial flights. When compared with those left behind, the difference of about a quarter of a microsecond was just what Einstein's theory predicted. Now there are many satellites in orbit around the Earth, so the theory has been well and truly tested many times. Indeed, global positioning systems must take the relativistic corrections into account to ensure their accuracy.

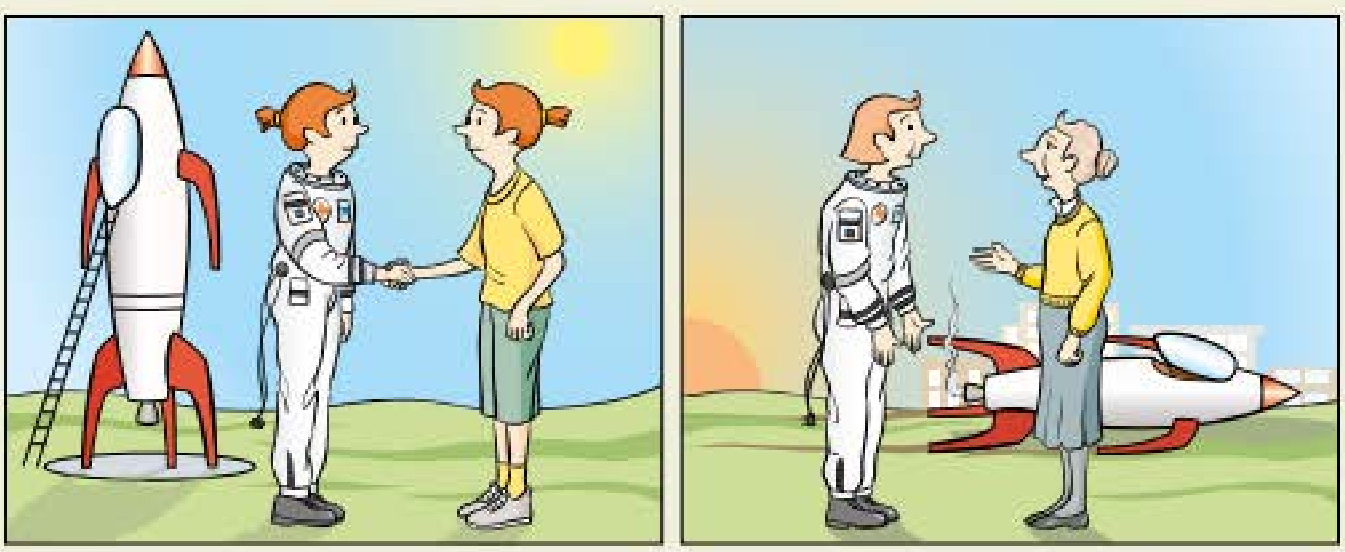


FIGURE 8.2.6 The Twin Paradox describes the phenomenon where one twin ages less quickly than the other after travelling in a non-inertial frame.

8.2 Review

SUMMARY

- The pulses in a light clock in a moving frame of reference have to travel further when observed from a stationary frame.
- Because of the constancy of the speed of light, this effectively means that time appears to have slowed in a moving frame.
- Time in a moving frame seems to flow more slowly according to the equation $t = t_0\gamma$ where t_0 is the time in the moving frame (proper time), t is the time observed from the stationary frame and γ is the Lorentz factor:
$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$
- Observers in relative motion both see time slowing in the other frame of reference; that is, each sees the other ageing more slowly.
- If one observer accelerates in order to return to meet the other, then that accelerated observer will have aged less than the other.
- Time dilation provides the explanation for the occurrence of muons reaching the Earth's surface after originating 15 km up in the upper atmosphere, when they should all decay within 7 km of their journey, according to classical physics.

KEY QUESTIONS

For the following questions, assume Gedanken *conditions exist and let* $c = 3 \times 10^8\,\text{m s}^{-1}$ *unless stated otherwise.*

1 Complete the following sentences by selecting the correct term from those in bold.

In a device called a **light/mechanical/digital** clock, the **speed/oscillation/wavelength** of light is used as a means of measuring **time/mass**, as the speed of light is **unknown/variable/constant** no matter from which inertial frame of reference it is viewed.

2 To what does the term 'proper time', t_0, refer?

3 An observer is standing on a train platform as a very fast train passes by at a speed of $1.75 \times 10^8\,\text{m s}^{-1}$. The observer notices the time on a passenger's phone as the passenger drops the phone to the floor. According to the clock on the phone, the phone takes 1.05 s to hit the floor. Calculate how much time has passed on the platform's clock during this time.

4 An observer standing on a comet is watching as a satellite approaches at a speed of $2.30 \times 10^8\,\text{m s}^{-1}$. The observer times on her watch that the solar panels on the satellite unfold in 75.0 s. Calculate how much time the observer sees as having passed on the satellite's clock.

5 A student standing by the side of a road sees a very fast sports car driving past. The driver times on his car's clock that it takes 5.50 s for the student to pick up her bag. If the car is moving at a speed of $2.75 \times 10^8\,\text{m s}^{-1}$, calculate how much time the driver sees has passed on the student's watch as she picks up the bag.

6 If Anna saw Ben fly by at 0.5*c*, how long, in her frame, would it take Ben's clock to tick 1 second?

7 Anna's *Gedanken* light clock has a height of 1 m between the mirrors, and relative to Chloe her spaceship is travelling at 90% of the speed of light ($c = 3.0 \times 10^8\,\text{m s}^{-1}$). One tick is the time for light to go from one mirror to the other.

- **a** How far does the light flash travel in Anna's frame of reference in one tick, t_A?
- **b** What is the tick time, t_A, for the clock in Anna's frame?

As the light takes a zigzag path in her frame, Chloe sees the clock ticking at a slower rate, t_C.

- **c** In terms of c and t_C, what is the length of the zigzag path that the flash travels in one tick in Chloe's frame?
- **d** What is the tick time of the clock in Chloe's frame?
- **e** What is the ratio of Chloe's tick to Anna's tick?

8 A muon created at an altitude of 15.0 km above the Earth is moving at a speed of 0.992 times the speed of light. The mean lifetime of a muon at rest is 2.20×10^{-6} s.

- **a** Calculate the lifetime of the moving muon as timed by a stationary observer.
- **b** Using classical physics equations and the results from part (a), calculate the non-relativistic distance and the relativistic distance travelled by the moving muon during one lifetime.

9 A high-speed, subatomic particle is accelerated by a linear accelerator to a speed of $2.83 \times 10^8\,\text{m s}^{-1}$. A researcher measures that it only leaves, on average, a track that is 2.50 cm long in the bubble chamber. Calculate the mean lifetime of the same particle if it was at rest relative to the researcher and her timer.

10 Briefly explain why Einstein said that a clock at Earth's equator should run slightly slower than one at the Earth's poles. Why do we not find this to be a problem?

8.3 Length contraction

The previous section described how time can only be measured relative to some particular frame of reference, but not in any absolute sense. Because of the constancy of the speed of light, this effectively means that time appears to have slowed in a moving frame relative to the frame of an observer. Einstein describes how space and time are interrelated, so it follows then that space, and therefore length, is not absolute (Figure 8.3.1). This section explores the effect on the length of an object based on the motion of the object in an inertial frame of reference.

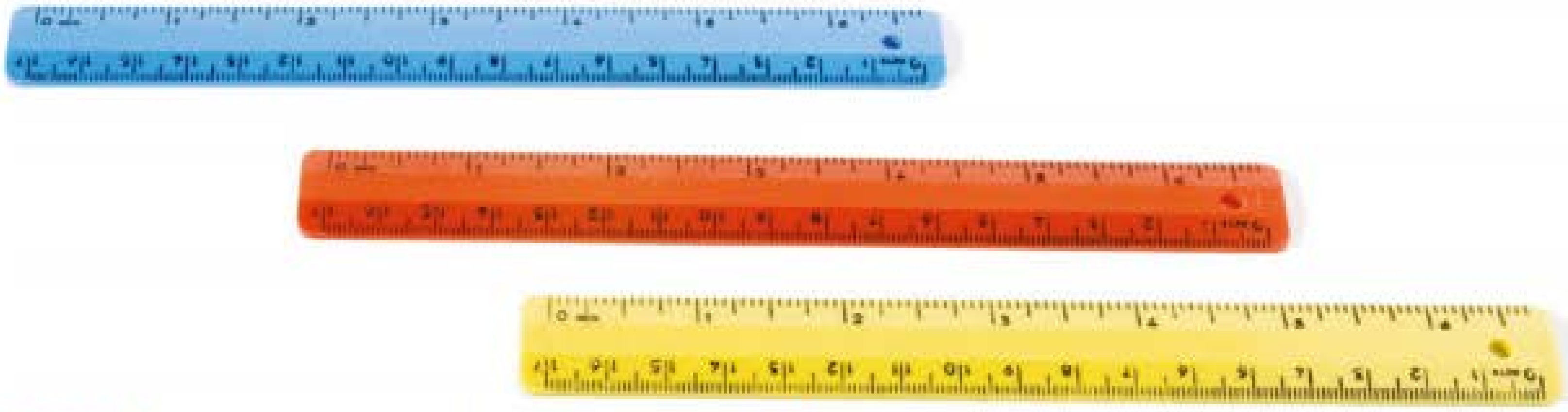


FIGURE 8.3.1 Length is relative to the frame of reference and the direction of motion.

LENGTH IN DIFFERENT INERTIAL FRAMES

You already have a clue to the fact that lengths depend on who is doing the measuring and the frame of reference in which they make their measurement.

The light clock analysis is appropriate to compare the proper time on the clock in the moving frame of reference (*observed by* Clare in the examples provided in Section 8.2) and the time measured on a clock in the stationary frame (*with* Clare). The light clock was used as it only depends on light, not some complicated mechanical arrangement that may well include other factors that are altered by relative motion. There was, however, one other condition in this clock analysis—that both Amaya and Clare would agree on the distance, d, between the mirrors. This enabled the two expressions for d to be equated in order to find the relationship between proper time, t_0, and time, t.

The clock was deliberately set up in the spaceship so that this light path, of distance d, was perpendicular (at right angles to) the velocity. Distances in this perpendicular direction are unaffected by motion. Indeed, Einstein showed that perpendicular distances are unaffected and relative motion affects length only in the direction of travel (refer to Figure 8.3.2).

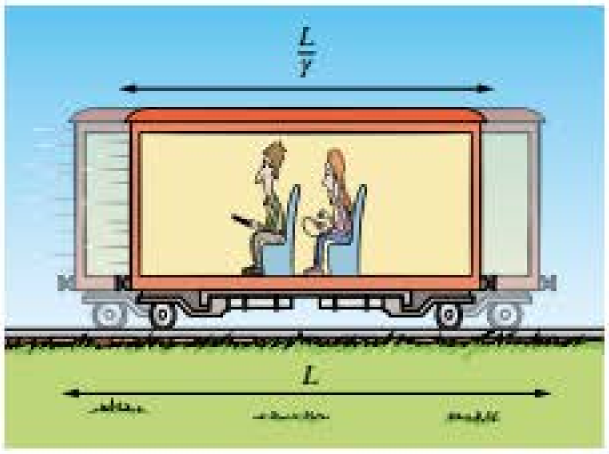


FIGURE 8.3.2 Einstein showed that the length of a moving object is foreshortened by the Lorentz factor, γ. The height and width of the carriage, howver, remain unchanged.

Length contraction

Consider the *Gedanken* situation in which Clare is standing on a train platform while Amaya and Binh pass by at a speed v. Both Clare and Binh want to measure the length of the train platform on which Clare is standing. Using a measuring tape, Clare measures the length of the platform (which is at rest according to her) as L_0, and says that Binh and Amaya cover this distance in a time equal to:

$$t = \frac{L_0}{v}$$

Binh observes the platform passing in a time t_0, as he and Amaya move past the station. The relationship between the time in Binh's frame of reference and the time that Clare measures is:

$$t_0 = \frac{t}{\gamma}$$

$$t_0 = t\sqrt{1 - \frac{v^2}{c^2}}$$

Substituting the equation for t into the previous equation for t_0 gives us:

$$t_0 = \frac{L_0}{v}\sqrt{1-\frac{v^2}{c^2}}$$

Binh sees the platform moving at a velocity of v relative to him, so he can say that the distance from the start to the end of the platform is:

$$L = vt_0$$

Substituting the equation for t_0 above into the equation for L gives us:

$$L = v \times \frac{L_0}{v}\sqrt{1-\frac{v^2}{c^2}}$$

This simplifies to:

$$L = L_0\sqrt{1-\frac{v^2}{c^2}}$$

This is Einstein's **length contraction** equation, which incorporates the Lorentz factor. This equation shows that an object with a **proper length** of L_0, when measured at rest, will have a shorter length L, parallel to the motion of its moving frame of reference, when measured by an observer in a stationary frame of reference. The proper length is contracted by a factor of $\frac{1}{\gamma}$. Length contraction can be represented as:

$$L = \frac{L_0}{\gamma}$$

where $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$

L_0 is the proper length, i.e. the length measured at rest, in the stationary frame of reference

L is the length in the moving frame, measured by an observer

Worked example 8.3.1

LENGTH CONTRACTION

Assume *Gedanken* conditions exist in this example. A stationary observer on Earth sees a very fast car travelling by at $2.50 \times 10^8\,\text{m}\,\text{s}^{-1}$. When stationary, the car is 3.00 m long. Calculate the length of the moving car as seen by the stationary observer. Use $c = 3.00 \times 10^8\,\text{m}\,\text{s}^{-1}$.

Thinking	Working
Identify the variables: the length measured by the stationary observer is L, the proper length of the car is L_0 and the velocities are v and the constant c.	$L = ?$ $L_0 = 3.00\,\text{m}$ $v = 2.50 \times 10^8\,\text{m}\,\text{s}^{-1}$ $c = 3.00 \times 10^8\,\text{m}\,\text{s}^{-1}$
Use Einstein's length contraction formula and the Lorentz factor.	$L = \frac{L_0}{\gamma}$ $= L_0\sqrt{1-\frac{v^2}{c^2}}$
Substitute the values for L_0, v and c into the equation and calculate the answer, L.	$L = 3.00 \times \sqrt{1-\frac{(2.50\times10^8)^2}{(3.00\times10^8)^2}}$ $= 3.00 \times 0.553$ $= 1.66\,\text{m}$

Worked example: Try yourself 8.3.1

LENGTH CONTRACTION

Assume *Gedanken* conditions exist in this example. A stationary observer on Earth sees a very fast scooter travelling by at $2.98 \times 10^8\,\text{m s}^{-1}$. The stationary observer measures the scooter's length as 45.0 cm. Calculate the proper length of the scooter, measured when the scooter is at rest.

LOOKING OUT OF THE WINDOW

So far you have been looking at situations in which objects that are in a moving frame of reference are seen as being shorter in the direction of the motion according to an observer who is in a stationary frame of reference. You can also apply length contraction to the distance that a moving object covers as it travels at very high speed.

Recall that no inertial frame of reference is special. Consider Amaya and Binh in their spacecraft. According to them, they are stationary and it is space itself that rushes by at high speed. As space zooms by Amaya and Binh, they are travelling a proper distance of 384 400 km from the Earth to the Moon. This is the proper length as it is measured by a device that is in the same frame of reference as the Earth and the Moon. As Binh looks out of the window, he sees a much shorter distance to travel.

Worked example 8.3.2

LENGTH CONTRACTION FOR DISTANCE TRAVELLED

Assume *Gedanken* conditions exist in this example. A spaceship is travelling at 0.997*c* from Earth to the Moon. The proper distance from the Earth to the Moon is 384 400 km. When the pilot looks out of the window, the distance between the Earth and the Moon looks much less than that. Calculate the distance that the pilot sees.	
Thinking	**Working**
Identify the variables: the length seen by the pilot is L, the proper length of the distance is L_0 and the velocity is v.	$L = ?$ $L_0 = 384\,400\text{ km}$ $v = 0.997c\text{ m s}^{-1}$
Use Einstein's length contraction formula and the Lorentz factor.	$L = \frac{L_0}{\gamma}$ $= L_0\sqrt{1-\frac{v^2}{c^2}}$
Substitute the values for L_0 and v into the equation. Cancel c and calculate the answer, L.	$L = 384\,400 \times \sqrt{1-\frac{(0.997c)^2}{c^2}}$ $= 384\,400 \times \sqrt{1-(0.997)^2}$ $= 384\,400 \times 0.0774$ $= 29\,753$ $= 29\,800\text{ km}$

Worked example: Try yourself 8.3.2

LENGTH CONTRACTION FOR DISTANCE TRAVELLED

Assume *Gedanken* conditions exist in this example. A stationary observer on Earth sees a very fast train approaching a tunnel at a speed of 0.986*c*. The stationary observer measures the tunnel's length as 123 m long. Calculate the length of the tunnel as seen by the train's driver.

The result from Worked example: Try yourself 8.3.2 leads to an interesting phenomenon. If the proper length of the train is 100 m, then the driver could park the train in the 123 m tunnel with 11.5 m of tunnel extending beyond each end of the train. But when the train is moving at 0.986*c*, then according to the train driver the train will not fit in the tunnel. There will be approximately 39.8 m of train extending past each end of the tunnel. This phenomenon is illustrated in Figure 8.3.3.

Similarly, a train that is longer than the tunnel will fit completely inside the tunnel if the length of the train was measured by a stationary observer as it was moving past very quickly. In this scenario, the length of the train would be contracted according to the stationary observer (refer to Figure 8.3.4).

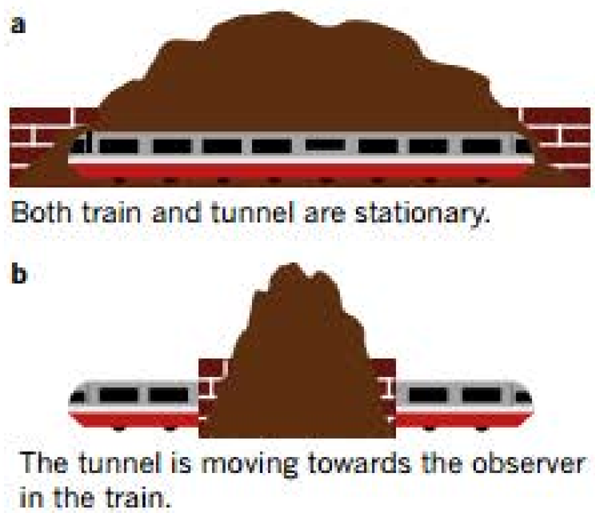


FIGURE 8.3.3 The train both fits in the tunnel and doesn't fit in the tunnel, depending on your frame of reference.

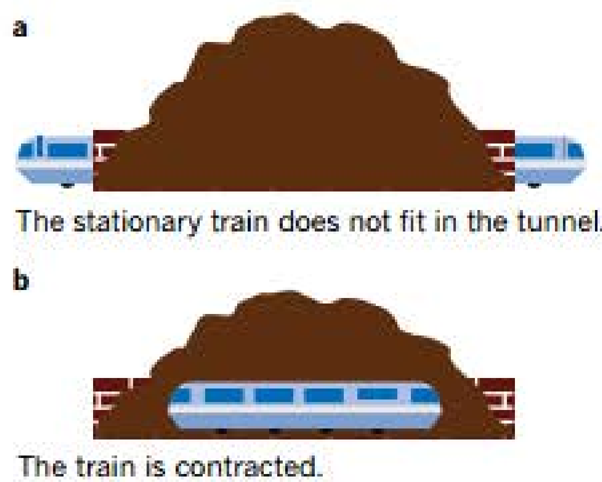


FIGURE 8.3.4 The train both doesn't fit in the tunnel and does fit in the tunnel, also depending on your frame of reference.

PROPER TIME AND PROPER LENGTH

The time t_0 and the length L_0 are referred to as the proper time and proper length. They are the quantities measured by the observer, who is in the same frame of reference as the event or the object being measured.

Proper time

The proper time is the time between two events that occur at the same point in space. For example, when a light bulb in the train flashes and Amaya measures the time for the flash to reflect off a mirror and return to her, then she has measured the proper time. This is because the stopwatch remained at the point in space inside the frame of reference in which the light originated and ended up. Proper time is illustrated in Figure 8.3.5.

It is important that a clock isn't moved from one place to another if you want to measure proper time. This is because, as soon as the clock is in motion, the time for that clock slows slightly.

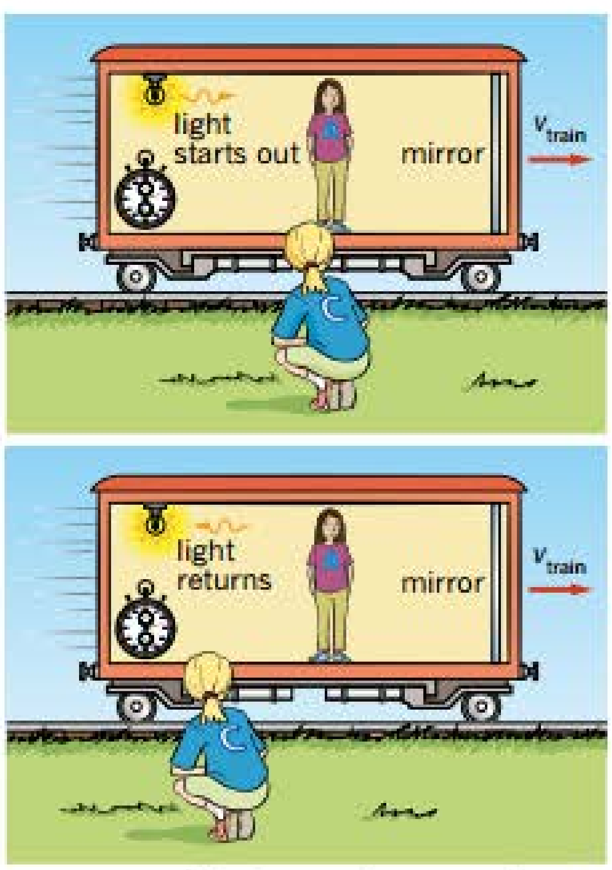


FIGURE 8.3.5 A clock measuring proper time. The clock is positioned at the place where the event started (the light starting out) and is at the same place when the event ends (the light returning).

Proper length

The proper length is the distance between two points whose positions are measured by an observer at rest with respect to the two points.

Recall the example of Amaya on a train and Clare beside the track observing the passing train. As Amaya reads her measuring tape at either end of the carriage and is at rest with respect to the train, her measurement of the carriage is the proper length. Clare's measurement of the carriage will be of the contracted length.

Clare, on the other hand, measures the length of the platform as the proper length, while Amaya and Binh see the platform as contracted in length. Remember that length contraction occurs only in the direction of travel, not in any perpendicular direction. To Clare, the carriage will appear shortened, but its width and height (the dimensions of the train perpendicular to the direction of travel) will remain unaltered.

An example of length contraction is shown with a tennis ball in Figure 8.3.6. The length in the direction of the motion is contracted, but the height is not.

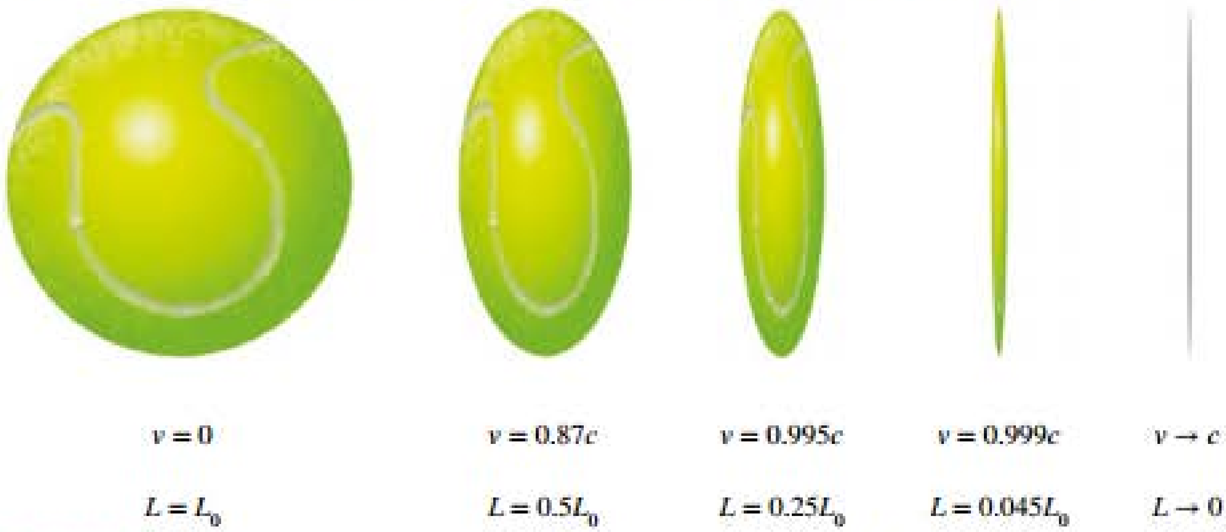


FIGURE 8.3.6 As the tennis ball moves faster to the right, its length in this dimension is contracted, but its height and depth remain the same.

FINAL THOUGHTS

Length contraction and time dilation are easy to confuse. One way to remember how it works is to think that stationary clocks tick faster and an object is longest when viewed from its own frame of reference. When viewed from a frame of reference in which objects are seen to be moving, they appear shorter and their clocks tick slower. All lengths and all clocks seem normal when viewed from within their own frame of reference.

8.3 Review

SUMMARY

- The theory of special relativity states that time and space are related. Motion affects space in the direction of travel.
- A moving object will appear shorter, or appear to travel less distance, by the inverse of the Lorentz factor, γ. Einstein's length contraction equation is given by:

$$L = \frac{L_0}{\gamma}$$

where L_0 is the proper length in the stationary frame, L is the contracted length as seen in the moving frame and γ is the Lorentz factor:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

- The proper time, t_0, is the time measured by an observer at the same point in an inertial frame of reference.
- The proper length, L_0, is the length measured by an observer at rest with respect to the object being measured.

8.3 Review *continued*

KEY QUESTIONS

For the following questions, assume Gedanken *conditions exist and let* $c = 3 \times 10^8\,m\,s^{-1}$ *unless stated otherwise.*

1 To what does the term 'proper length', L_0, refer?

2 If you are standing on Earth and observe a speeding rocket ship, what do you notice about its dimensions? Select from the following:

A Its length (in the direction of travel) is shorter than normal.

B Its length (in the direction of travel) is longer than normal.

C Its height (at right angles to the direction of travel) is shorter than normal.

D Its width (at right angles to the direction of travel) is shorter than normal.

3 An observer is standing on a train platform as a very fast train passes by at a speed of $1.75 \times 10^8\,m\,s^{-1}$. The observer notices that a passenger is holding a 1-metre rule in line with the direction that the train is moving. Calculate the length of the 1-metre rule that the stationary observer sees.

4 An observer standing on a comet is watching as a satellite approaches at a speed of $2.30 \times 10^8\,m\,s^{-1}$. The observer knows that the proper length of the satellite in the direction of its motion is 5.25 m. Calculate the length of the satellite that the observer sees as it passes.

5 A builder makes a mistake and builds a garage too short for the owner's car to fit in. The proper length of the garage is 1.50 m and the proper length of the car is 3.50 m. The builder suggests that if the owner drives fast enough, the builder could stand by the garage and the car would fit.

a Calculate the speed at which the car would need to travel to just fit in the garage when observed by the builder.

b Explain why the car owner would not be happy about the builder's suggestion, by calculating the length of the garage as seen by the driver.

6 An observer on a platform measures the time for a train carriage, moving at 0.99c, to pass her by. What time has she measured, t or t_0? Explain.

7 According to a speed (v) versus distance travelled (L) graph, which of the following is true?

A At the maximum speed, the distance travelled is the largest.

B Velocity and distance travelled are directly proportional variables.

C At values close to the speed of light, the distance travelled is near to zero.

D None of the above.

8 Emily is standing by the side of the track, watching Dan run in an 800 m race.

a At what speed must Dan run in order for the race to be only 400 m long in his frame of reference?

b Emily notices that Dan is thinner than he normally is, but just as wide and just as tall. Calculate Dan's thickness while he is running as a fraction of his normal thickness while standing still.

9 A jet plane zooms at a speed of $660\,m\,s^{-1}$ past an observer standing on the ground. If the length of the jet is 23.5 m when parked on the tarmac, calculate the length of the moving jet, as seen by the observer.

10 An astronaut in her spaceship is speeding at 0.900c to the Moon. She is holding, in the direction in which the spaceship is moving, a fishing rod that is 2.75 m long.

a Determine the length of the rod as observed by an astronaut in the International Space Station.

b What is the length of the fishing rod as observed by the astronaut in the spaceship?

8.4 Relativistic momentum and energy

RELATIVISTIC MOMENTUM

If a rocket ship like the one in Figure 8.4.1 is travelling at $0.99c$, why can't it simply turn on its rocket motor and accelerate up to c, or more? A full answer to this question was not given in Einstein's original 1905 paper on relativity. Some years later he showed that as the speed of a spaceship approaches c, its **momentum** increases, but this is not reflected in a corresponding increase in speed.

Although his analysis is beyond the scope of this course, you can get a feel for his approach if you take some short cuts.

The acceleration, a, of any object is inversely proportional to its mass, m, the mass that appears in Newton's second law:

$$F = ma$$

Newton originally stated this law as: a force, F, is equal to the rate of change in momentum p; that is:

$$F = \frac{\Delta p}{\Delta t}$$

A change in momentum is classically defined as the change in the product of the mass, m, and the velocity, v. If you rearrange the above equation and substitute the relationship $\Delta p = m\Delta v$, you get:

$$F\Delta t = m\Delta v$$

Now you see that time is involved; but you know that at relativistic speeds time is not the constant entity it was once believed to be.

FIGURE 8.4.1 This rocket ship is moving at 0.99c and accelerating, and yet it can never reach a speed of c.

Imagine that you have a rocket ship accelerating from rest to a high speed as viewed by an observer in a stationary frame of reference. You can say that the change in momentum of the ship will be given by:

$$F\Delta t_0 = m\Delta v$$

where t_0 is the time in the ship's frame of reference, and $m\Delta v$ is just the classical Newtonian change in momentum.

In the stationary observer's frame, the time is dilated:

$$\Delta t = \gamma\Delta t_0$$

$$\Delta t_0 = \frac{\Delta t}{\gamma}$$

PHYSICSFILE

Travel at the speed of light

Einstein said that at the speed of light distances shrink to zero and time stops. No ordinary matter can reach c, but light always travels at c. Strange though it may seem, for light there is no time. It appears in one place and disappears in another, having got there in no time (in its own frame of reference, not ours!). When you stay still, you travel through spacetime in the time dimension only. Light does the opposite: all its spacetime travel is through space and none through time.

Substituting Δt_0 into the change of momentum equation:

$$F\frac{\Delta t}{\gamma} = m\Delta v$$

$$F\Delta t = \gamma m\Delta v$$

That is, the impulse, as seen by the stationary observer, is equal to the product of the Lorentz factor, γ, and the change in Newtonian momentum. This means that as the spaceship approaches the speed of light, the impulse is multiplied by a factor that grows very rapidly. You can interpret this as meaning that the change in momentum in the stationary observer's frame of reference is equal to:

$$\Delta p = \gamma m\Delta v$$

$$\Delta p = \gamma\Delta p_0$$

If you assume an object starts at zero velocity, the final relativistic momentum becomes:

$$p = \gamma mv \quad \text{or} \quad p = \frac{1}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}}mv$$

$$p = \gamma p_0$$

where p_0 is the momentum mv, as you would define it in classical mechanics, and p is the relativistic momentum.

If velocity, v, is needed when the mass and relativistic increase in momentum is known, the formula $p = \gamma mv$ can be rearranged to give the following:

$$v = \frac{p}{m\sqrt{\left(1+\frac{p^2}{m^2c^2}\right)}}$$

The momentum increases very rapidly as a spaceship approaches the speed of light. You might argue that this is expected—after all, momentum is a function of velocity. If you graph the relativistic momentum, p, against the velocity, v, and on the same graph show the classical momentum, you can see that the relativistic momentum increases at a rate far greater than it would if it were due to the increase in velocity alone (Figure 8.4.2).

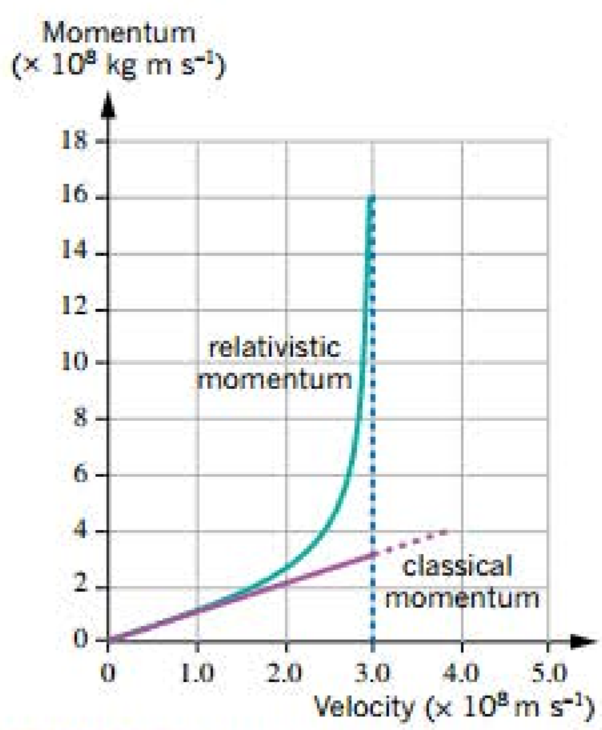


FIGURE 8.4.2 The relationship between classical momentum and velocity, and the relationship between relativistic momentum and velocity, for a 1 kg mass.

This result can be interpreted by thinking of the mass as a quantity that also increases at high speeds. Thus there is a relationship between the rest mass, m, which is the mass measured while the object is at rest in the frame of reference, and the relativistic mass, γm, which is the mass measured as the object is moving relative to the observer.

As the Lorentz factor increases with the increase in the velocity, then the relativistic mass also increases. Einstein was never happy with the term 'relativistic mass', and preferred that people only spoke of the relativistic momentum of an object.

Notice that the classical treatment would allow the object to have a speed greater than the speed of light, but the relativistic treatment causes the mass to become very large so that the speed of light is never actually reached.

Now return to the example of the rocket ship that is attempting to increase its velocity to the speed of light. With the increase in the relativistic mass of the rocket ship, it becomes harder for the force of the engines to cause a change in velocity. The closer the rocket ship approaches c, the greater the impulse that is required to accelerate the ship to the speed of light. In fact, as the velocity approaches c, the relativistic mass, γm, approaches infinity. You can now see why your rocket ship cannot reach the speed of light.

Worked example 8.4.1 illustrates this point. Notice that the result in part (b) shows that if you double the impulse required to get the rocket ship to $0.99c$, then you will only add $0.007c$ to your top speed. When you've completed Worked example: Try yourself 8.4.1, consider the change in velocity achieved by tripling the impulse.

Worked example 8.4.1

RELATIVISTIC MOMENTUM

a Calculate the momentum, as seen by a stationary observer, provided to a rocket ship with a rest mass of 1000 kg, as it goes from rest up to a speed of 0.990*c*. Assume *Gedanken* conditions exist in this example.

Thinking	Working
Identify the variables: the rest mass is *m*, and the velocity of the rocket ship is *v*.	$\Delta p = ?$ $m = 1000\,\text{kg}$ $v = 0.990 \times 3.00 \times 10^8\,\text{m s}^{-1}$
Use the relativistic momentum formula.	$p = \gamma m v$
Substitute the values for *m* and *v* into the equation and calculate *p*.	$p = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} m v$ $= \frac{1}{\sqrt{1-\frac{0.990^2 c^2}{c^2}}} \times 1000 \times 0.990 \times 3.00 \times 10^8$ $= 2.11 \times 10^{12}\,\text{kg m s}^{-1}$

b If twice the relativistic momentum from part (a) is applied to the stationary rocket ship, calculate the new final speed of the rocket ship in terms of *c*.

Thinking	Working
Identify the variables: the rest mass is *m*, and the relativistic momentum of the rocket ship is *p*.	$p = 2 \times 2.11 \times 10^{12}$ $= 4.21 \times 10^{12}\,\text{kg m s}^{-1}$ $m = 1000\,\text{kg}$ $v = ?$
Use the relativistic momentum formula, rearranged.	$p = \gamma m v$ $p = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} m v$ $v = \frac{p}{m\sqrt{\left(1+\frac{p^2}{m^2 c^2}\right)}}$
Substitute the values for *m* and *p* into the rearranged equation and calculate the answer *v*.	$v = \frac{p}{m\sqrt{\left(1+\frac{p^2}{m^2 c^2}\right)}}$ $= \frac{4.21 \times 10^{12}}{(1000)\sqrt{1+\frac{\left(4.21 \times 10^{12}\right)^2}{1000^2\left(3.00 \times 10^8\right)^2}}}$ $= \frac{4.21 \times 10^{12}}{1000 \times 14.07}$ $= 2.99 \times 10^8\,\text{m s}^{-1}$ $= 0.997c$

Worked example: Try yourself 8.4.1

RELATIVISTIC MOMENTUM

a Calculate the momentum, as seen by a stationary observer, provided to an electron with a rest mass of 9.11×10^{-31} kg, as it goes from rest to a speed of 0.985*c*. Assume *Gedanken* conditions exist in this example.

b If three times the relativistic momentum from part (a) is applied to the electron, calculate the new final speed of the electron in terms of *c*.

RELATIVISTIC ADDITION OF VELOCITIES

Consider a situation in which a spaceship moving at 0.90*c* fires a rocket forwards at a speed of 0.20*c*. Classical physics would say that a stationary observer would see the rocket moving at 1.10*c*, which is a speed faster than light. This is an impossible situation.

Binh and Amaya are in a train carriage moving at a velocity v, while Clare is standing still beside the track. Amaya is standing at the back of the carriage and throws a ball forwards, towards Binh, at a velocity of u' (Figure 8.4.3). The term u' indicates that it is the velocity of the ball that Amaya and Binh calculate in their frame of reference.

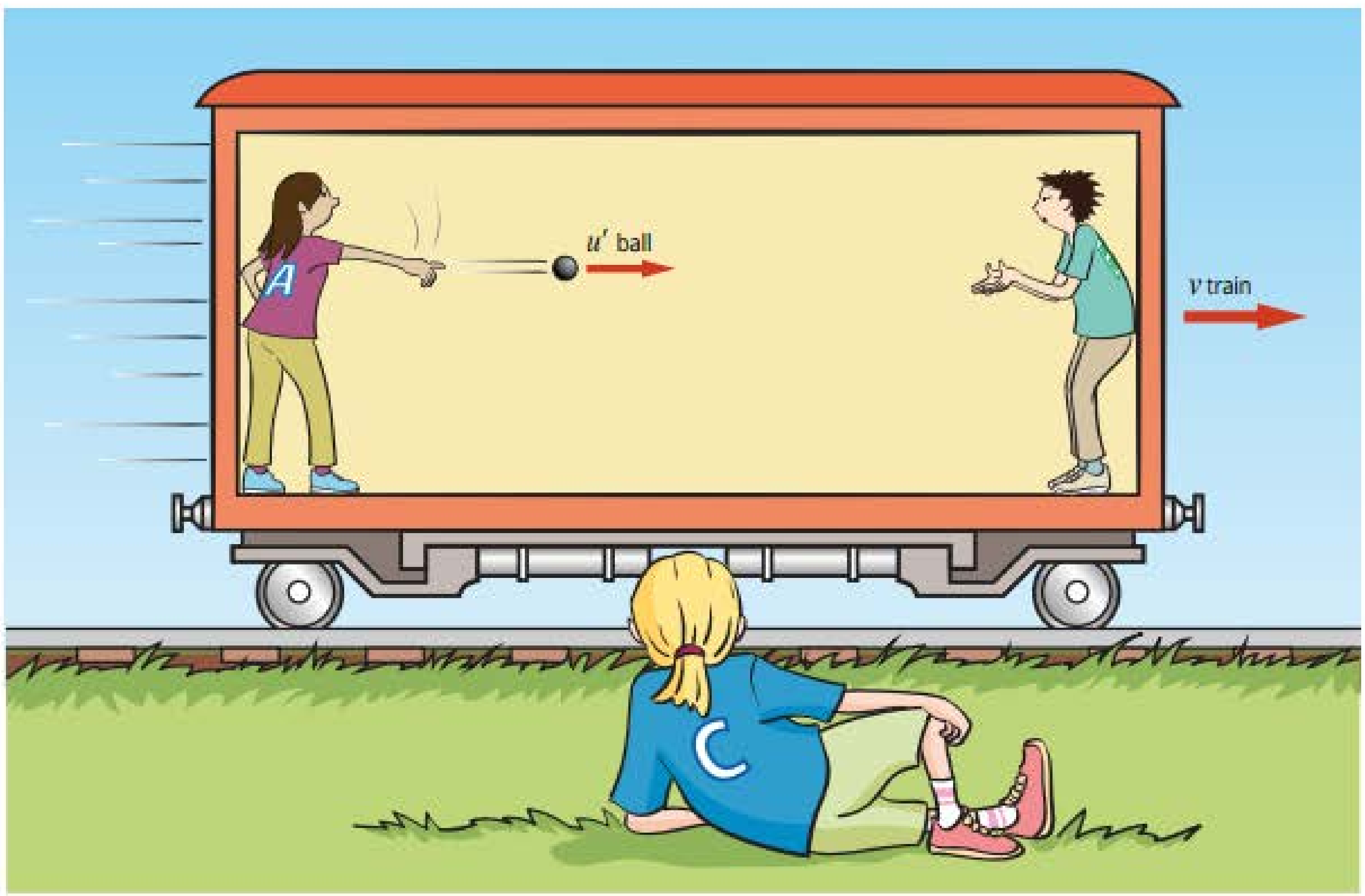


FIGURE 8.4.3 The velocity of the ball relative to the train carriage is u', the velocity of the ball for the stationary observer (Clare) is u, and the velocity of the train carriage is v.

The velocity is calculated using the distance the ball travels and the time it takes.

$$u' = \frac{x'}{t'}$$

$$x' = u't'$$

Suppose you are interested in the velocity u at which Clare sees the ball travelling. The solution to the problem relies on two equations, that we will not derive here, called the Lorentz transformation equations for x' and t'.

$$x' = \gamma(x - vt)$$

and

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

Setting the two equations for x' equal to each other gives the following result:

$$u't' = \gamma(x - vt)$$

Substituting the equation for t' into this equation yields:

$$u'\gamma\left(t - \frac{vx}{c^2}\right) = \gamma(x - vt)$$

Cancel the term γ and rearrange the equation to make x the subject:

$$u'\left(t - \frac{vx}{c^2}\right) = x - vt$$

$$u't - \frac{u'vx}{c^2} = x - vt$$

$$x + \frac{u'vx}{c^2} = u't + vt$$

$$x\left(1 + \frac{u'v}{c^2}\right) = (u' + v)t$$

$$x = \frac{(u' + v)}{\left(1 + \frac{u'v}{c^2}\right)}t$$

In Clare's frame of reference the velocity of the ball is given by:

$$u = \frac{x}{t}$$

Substituting the equation for x above into the equation for velocity u of the ball in Clare's frame of reference, you get the equation for calculating the relativistic addition of velocities:

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}}$$

u is the velocity of the moving object in the stationary frame of reference
u' is the velocity of the moving object in the moving frame of reference
v is the velocity of the moving frame of reference
c is the speed of light

Rearranging this equation to make u' the subject gives us:

$$u' = \frac{u - v}{1 - \frac{u'v}{c^2}}$$

Worked example 8.4.2

RELATIVISTIC ADDITION OF VELOCITIES

Assume *Gedanken* conditions exist in this example. An observer on a train platform sees a train carriage passing by travelling at 0.90c. Inside the carriage a ball is thrown towards the front of the carriage at 0.20c. Calculate the velocity of the ball as seen by the stationary observer, in terms of c.

Thinking	Working
Identify the variables for the velocity for the stationary observer *u*, the velocity of the ball relative to the carriage *u'* and the velocity of the carriage *v*.	$u = ?$ $u' = 0.20c$ $v = 0.90c$
Use the relativistic velocity addition formula.	$u = \dfrac{u' + v}{1 + \dfrac{u'v}{c^2}}$
Substitute the values for *u'* and *v* into the equation and calculate the answer *u*.	$u = \dfrac{(0.20c) + (0.90c)}{1 + \dfrac{(0.20c)(0.90c)}{c^2}}$ $= \dfrac{(1.10c)}{1 + \dfrac{(0.18c^2)}{c^2}}$ $= \dfrac{1.10c}{1 + 0.18}$ $= \dfrac{1.10c}{1.18}$ $u = 0.932c$

Worked example: Try yourself 8.4.2

RELATIVISTIC ADDITION OF VELOCITIES

Assume *Gedanken* conditions exist in this example. An observer on Earth sees a spaceship travelling at 0.99c pass by. Inside the spaceship a laser beam pulse is directed towards the front of the spaceship at c. Calculate, in terms of c, the velocity of the laser beam pulse as seen by the stationary observer.

The result of Worked example: Try yourself 8.4.2 indicates Einstein's postulate that the speed of light is always *c*, no matter from which frame of reference it is viewed.

EINSTEIN'S FAMOUS EQUATION

As the momentum of an object increases, so does its kinetic energy. The classical relationship between the two can be written as:

$$\begin{aligned} E_k &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(mv) \times v \\ &= \frac{1}{2}pv \end{aligned}$$

This form of the equation shows that the kinetic energy of an object is related to the object's momentum as well as its velocity.

Einstein showed, however, that the classical expression for kinetic energy was not correct at high speeds. The mathematics involved is beyond the scope of this course, but Einstein, working from the expression for relativistic momentum and the usual assumptions about work, forces and energy, was able to show that the kinetic energy of an object was given by the expression:

$$E_k = (\gamma - 1)mc^2$$

Although it is not very obvious from this expression, if the velocity is small this expression actually reduces to the classical equation for E_k of $\frac{1}{2}mv^2$. A small velocity in this context means small in comparison to c. But even for speeds of up to $0.10c$, the classical expression is accurate to better than ±1%.

Einstein's expression can be expanded to:

$$E_k = \gamma mc^2 - mc^2$$

This kinetic energy equation, in turn, can be rearranged as:

$$\gamma mc^2 = E_k + mc^2$$

Einstein interpreted the left-hand side of this expression as being an expression for the total energy of the object:

$$E_{tot} = \gamma mc^2$$

The right-hand side appeared to imply that there were two parts to the total energy: the kinetic energy, E_k, and another term that only depended on the rest mass, m. The second term, mc^2, he referred to as the rest energy of the object, as it does not depend on the speed of the object. This appeared to imply that somehow there was energy associated with mass (Figure 8.4.4). An astounding proposition to a classical physicist but, as you have seen, in relativity mass increases as you add kinetic energy to an object. The conservation of energy relationship is therefore:

$$E_{tot} = E_k + E_{rest}$$

$$E_{tot} = \gamma mc^2 \quad \text{or} \quad E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

where γm is the relativistic mass (kg)
c is the speed of light ($m\,s^{-1}$) and
E_{tot} is the total energy (J)

You will have seen part of this equation before.

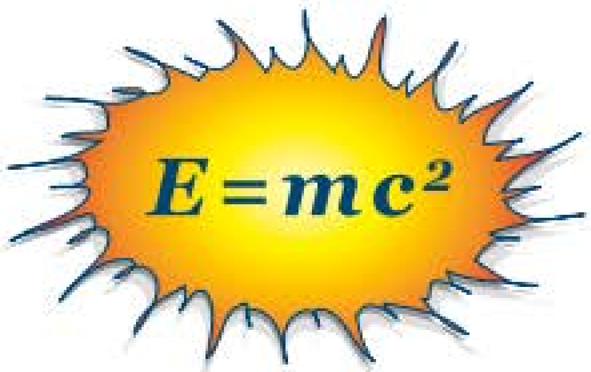


FIGURE 8.4.4 Einstein's famous equation.

This equation tells you that mass and energy are totally interrelated. In a sense, you can say that mass has energy, and energy has mass.

8.4 Review

SUMMARY

- Relativistic momentum includes the Lorentz factor, γ, and hence as the impulse is increased the mass seems to increase towards infinity as the speed gets closer to, but never equals, c. The relativistic momentum equation is:

  $p = \gamma mv = \gamma p_0$
- A term called relativistic mass, γm, may be used to indicate the mass of an object that is moving.
- Einstein found that the total energy of an object was given by:

  $E_{total} = E_k + E_{rest} = \gamma mc^2$
- The kinetic energy is given by:

  $E_k = (\gamma - 1)mc^2$
- The rest energy, which is the energy associated with the rest mass of an object, is given by:

  $E_{rest} = mc^2$
- Mass and energy are seen as different forms of the same thing. This means that mass, m, can be converted into energy, and energy can be converted into mass.
- Relativistic velocity addition differs from classical velocity addition, in that no two velocities can be added that exceed the speed of light.
- The relativistic velocity addition equation provides the velocity u of an object viewed by a stationary observer, when the object is moving with a velocity u', in a frame of reference that is moving at a velocity v, and is given by the formula:

$$u' = \frac{u - v}{1 - \frac{u'v}{c^2}}$$

KEY QUESTIONS

1 Calculate the relativistic momentum of the *Rosetta* spacecraft as observed by the scientists at the European Space Agency. *Rosetta's* rest mass is 1230 kg and its speed was 775 m s^{-1}.

2 Calculate the relativistic momentum of a carbon-12 nucleus in a linear accelerator if its rest mass is $1.99264824 \times 10^{-26}$ kg and it is travelling at 0.850c.

3 Calculate the relativistic momentum of another carbon-12 nucleus in the solar wind if its rest mass is $1.99264824 \times 10^{-26}$ kg and it is travelling at a speed of 800 m s^{-1}.

The following information relates to questions 4–6.

A very fast arrow has a rest mass of 12.3 g and a speed of 0.750c.

4 Calculate the relativistic kinetic energy of the arrow.

5 Calculate the kinetic energy of the arrow according to the classical equation.

6 What accounts for the difference between the kinetic energy of the arrow in the relativistic calculation and the kinetic energy in the classical calculation?

A the difference in the arrow's velocity in the two calculations

B the difference in the arrow's momentum in the two calculations

C the difference in the arrow's rest mass in the two calculations

D the presence of the Lorentz factor in the relativistic calculation

7 Assume *Gedanken* conditions exist in this question. A high-speed spaceship travels at 0.98c past a stationary space station. An astronaut on the space station notices that a subatomic particle is moving through the spaceship at a velocity of 0.95c relative to the space station. Calculate the velocity of the subatomic particle in the spaceship in terms of c.

Chapter review

08

KEY TERMS

aether	length contraction	proper length
classical physics	Lorentz factor	proper time
frame of reference	medium	simultaneous
Gedanken	momentum	spacetime
inertial frame of reference	postulate	time dilation

1 Prove that for an object travelling at any possible velocity, the value of the term below must be less than 1.

$$\sqrt{1-\frac{v^2}{c^2}}$$

2 One of the fastest objects ever made on Earth was the Galileo Probe which, as a result of Jupiter's huge gravity, entered its atmosphere in 1995 at a speed of nearly 50 000 m s^{-1}. Give an estimate of the Lorentz factor for the probe to nine decimal places. (You may use the expression $\gamma \approx 1+\frac{v^2}{2c^2}$.)

3 In 1905 Einstein put forward two postulates. Which two of the following best summarise them?

A All observers will find the speed of light to be the same.

B In the absence of a force, motion continues with constant velocity.

C There is no way to detect an absolute zero of velocity.

D Absolute velocity can only be measured relative to the aether.

4 Where on the Earth's surface are we closest to an inertial frame of reference?

5 Which of the following is closest to Einstein's first postulate?

A Light always travels at 3×10^8 m s^{-1}.

B There is no way to tell how fast you are going unless you can see what's around you.

C Velocities can only be measured relative to something else.

D Absolute velocity is that measured with respect to the Sun.

6 Very briefly explain why Einstein said that we must use four-dimensional spacetime to describe events that occur in situations where high speeds and large distances are involved.

7 Imagine that Amaya is at the front end of a train carriage moving forwards at 10.0 m s^{-1}. She shines a laser towards Binh, who is at the other end of the carriage. Clare is watching all this from the side of the track. At what velocity does Clare see the light travelling?

8 Which one or more of the following conditions is sufficient to ensure that we will measure the proper time between two events? We must:

A be in the same frame of reference.

B be in a frame of reference that is travelling at the same velocity.

C be stationary.

D not be accelerating with respect to the frame of the two events.

9 Spaceships A and B leave the Earth and travel towards Vega, both at a speed of 0.9*c*. Observer C back on Earth sees the crews of A and B moving in 'slow motion'. Describe how the crew of A see the crew of B, and how they see C and the Earthlings moving.

A B will appear normal, C will be sped up.

B B will appear normal, C will be slowed down.

C B will appear slowed down, C will be normal.

D B will appear sped up, C will be slowed down.

E None of these.

10 If you were riding in a very smooth, quiet train with the blinds drawn, how could you tell the difference between the train (i) being stopped in the station, (ii) accelerating away from the station, (iii) travelling at a constant speed?

11 You are in a spaceship travelling at very high speed past a new colony on Mars. Do you notice time going slowly for you; for example, do you find your heart rate is slower than normal? Do the people on Mars appear to be moving normally? Explain your answers.

12 An observer sitting in a very fast jet plane is looking out of the window at a clock placed on top of a mountain. The passenger, using the clock on the mountain, notes that it takes a goat 20.0 s to run along a rocky slope. If the plane is flying at a speed of 2.00×10^8 m s^{-1}, calculate how much time has passed on the passenger's clock.

13 A spectator is standing next to the pool clock and watching as a swimmer races at a speed of 2.25×10^8 m s^{-1}. The spectator times on the pool clock that the swimmer completes one stroke in 1.50 s.

a Calculate how much time the spectator sees pass on the swimmer's wristwatch.

b Calculate how much time the swimmer sees has passed on the pool clock, during which time her own wristwatch shows that 1.50 s have passed.

14 In the Twin Paradox explanation, when can you say the twin that stays at home ages faster than the twin who goes on the journey?

A during the acceleration phase

B during the deceleration phase

C during both the acceleration and deceleration phases

D during the constant velocity portion of the journey

15 a At what speed would a rocket be travelling if it is seen by a stationary observer to be half its normal length?

b The rocket ship is then observed to accelerate to a certain speed so that its length halved again. Did that mean that it doubled its speed? To what speed did it accelerate?

16 Binh and Amaya are playing table tennis in their spaceship. They rush past Clare in her space station at a relative speed of 240 000 km s^{-1}. Binh says that after he hits the ball it returns to his bat after 1.00 s. Their table is 3.00 m long in the direction of motion of their spaceship and is 1.00 m high.

a Calculate the time between hits, as measured by Clare.

b Calculate the length and height of the table, as measured by Clare.

17 Star Xquar is at a distance of 5 light-years from Earth. Space adventurer Raqu heads from Earth towards Xquar at a speed of 0.9c.

a For those watching from Earth, how long will it take for Raqu to reach Xquar?

b From Raqu's point of view how long will it take her to reach Xquar?

c Explain why it is that, although Raqu knew that Xquar was 5 light-years from Earth, and that she was to travel at 0.9c, it took much less time than might be expected from these figures.

18 The space shuttle travelled at close to 8000 m s^{-1}. Imagine that as it travelled east–west it took a photograph of Australia, which is close to 4000 km wide. Because of its speed, the space camera will see everything on Earth slightly contracted.

a About how much less than 4000 km wide will Australia appear to be in this photograph?

b Will the north–south dimension of Australia be smaller as well?

19 Imagine that as we watch a traveller from Earth travelling at 99.5% of the speed of light to the star Vega, we see that their clock slows down by a factor of about 10 times.

a Explain how this factor of 10 was arrived at.

b Does this mean that the traveller will experience this slowing down of time?

c Vega is about 25 light-years from Earth, so in our frame of reference it takes light from Vega 25 years to reach us. How long will it take our space traveller to reach Vega?

d How long will the traveller find that it takes to travel to Vega?

e Does your answer to part (d) imply that they were able to get to Vega in less time than light? Explain your answer.

20 Muons are high-speed particles that are created some 15 km above the Earth's surface. Classical physics dictates that due to their short life spans, muons should not ever reach the Earth's surface even though they travel at incredible speeds (approx. 0.992*c*). However, they do reach Earth. Explain how this is possible, by referring to the frames of reference of an observer on Earth and the muon itself.

21 The starship *Enterprise*, while travelling at 0.85c relative to an observer on a stationary starship, fires a missile forwards at 0.58*c* in the same direction as it is travelling. At what speed does the observer on the stationary starship see the missile moving?

22 A Starfighter with a rest mass of 950 kg is travelling into battle at a speed of 0.65c.

a Calculate its Newtonian momentum.

b Calculate its relativistic momentum.

c Comment on why these values are different.

23 Two spacecraft are travelling directly towards each other. Their relative velocity as measured by a stationary observer is 0.765c.

a If one is travelling at 0.67*c*, determine the velocity of the other.

b Are these two spacecraft in the same inertial reference frame?

24 During a dog-fight between intergalactic Starfighters, a piece of material with a rest mass of 1.90 kg breaks off one of the fighters when it is travelling at 0.85c.

a Calculate the relativistic momentum of this piece.

b What is its relativistic energy at this speed?

CHAPTER

09 The Standard Model

The Standard Model of particle physics and the Big Bang theory are the products of almost 100 years of development of mathematical theory and experimental observations. Particle accelerators are the main tools of the physicists who study the universe on the smallest scales. They are the tools required to test theoretical predictions made by the complex mathematical models that are part of the Standard Model of particle physics. These machines also contribute to the development of the Big Bang theory by creating the conditions that were present in the early universe.

In this chapter you will apply the physics concepts you have learnt in this course to understand the evidence that support these current theories.

Science as a Human Endeavour

The Big Bang theory describes the early development of the universe, including the formation of subatomic particles from energy and the subsequent formation of atomic nuclei. There is a variety of evidence that supports the Big Bang theory, including Cosmic Background Radiation, the abundance of light elements and the red shift of light from galaxies that obey Hubble's Law. Alternative theories exist, including the Steady State theory, but the Big Bang theory is the most widely accepted theory today.

Science Understanding

- the Standard Model is based on the premise that all matter in the universe is made up from elementary matter particles called quarks and leptons; quarks experience the strong nuclear force, leptons do not
- the Standard Model explains three of the four fundamental forces (strong, weak and electromagnetic forces) in terms of an exchange of force-carrying particles called gauge bosons; each force is mediated by a different type of gauge boson
- Lepton number and baryon number are examples of quantities that are conserved in all reactions between particles; these conservation laws can be used to support or invalidate proposed reactions. Baryons are composite particles made up of quarks
- high-energy particle accelerators are used to test theories of particle physics, including the Standard Model

  This includes deriving and applying the relationship

  $$\frac{mv^2}{r} = qvB$$
- the expansion of the universe can be explained by Hubble's Law and cosmological concepts, such as red shift and the Big Bang theory
- the Standard Model is used to describe the evolution of forces and the creation of matter in the Big Bang theory

9.1 Particles of the Standard Model

You will be familiar with the idea that the matter around us, and within us, is made of particles, and that the atoms in your body are comprised of protons, neutrons and electrons. Although electrons are still considered to be fundamental particles, what you may not know is that protons and neutrons are actually made of even smaller particles called quarks (Figure 9.1.1). These quarks experience a force called the strong nuclear force and this force holds the nucleus of atoms together.

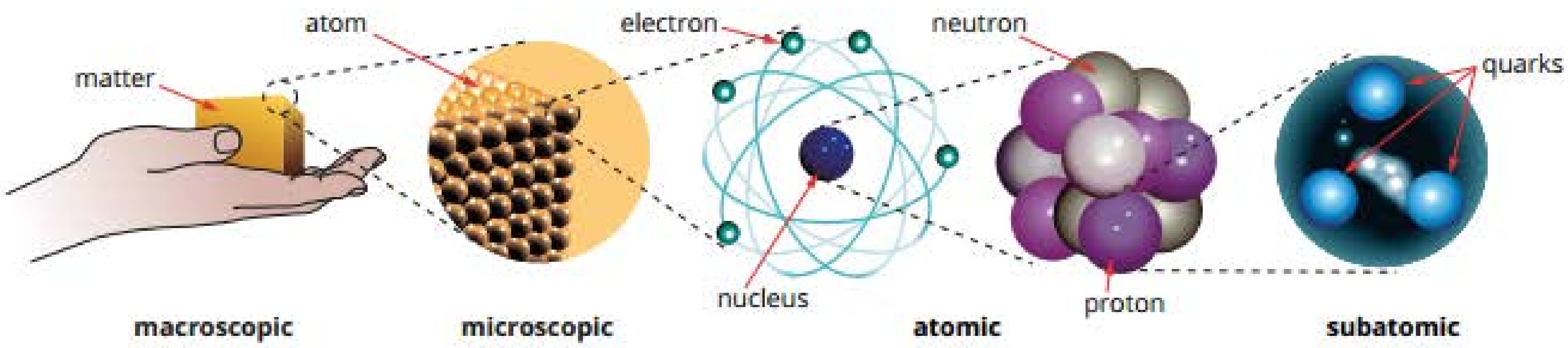


FIGURE 9.1.1 The structure of matter is shown, starting on the left, with the macroscopic, then microscopic, atomic and subatomic levels.

The existence of quarks is just the beginning. The universe is a vastly more complex and beautiful place than most people realise. The existence of quarks and the strong nuclear force are just two pieces of the puzzle that belong to the discipline of particle physics. The current understanding of what exists and occurs on the smallest scales in the universe is contained in the Standard Model of particle physics. This collection of experimentally supported theories has developed over more than a century and through the life work of many brilliant people.

THE STANDARD MODEL OF PARTICLE PHYSICS

The **Standard Model** of particle physics is currently the most successful theory for predicting the behaviour and properties of the particles that exist in nature. It is a mathematical description of all known particles and three of the forces acting on them.

The Standard Model is based on the theories of quantum mechanics and **special relativity** and the assumption that interactions between particles are caused by exchange particles called **gauge bosons**. These theories, and this assumption, have enabled theoretical physicists to develop the mathematical theories that together are the components of the Standard Model.

Figure 9.1.2 shows the three fundamental forces that are described by the quantum field theories of quantum electrodynamics and **electroweak theory**: the **strong nuclear force**, the **weak nuclear force** and the **electromagnetic force**.

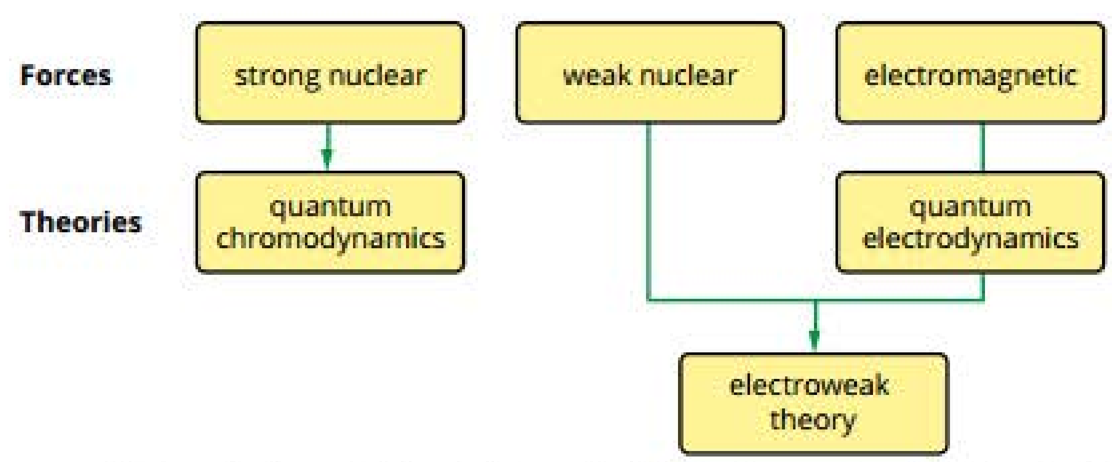


FIGURE 9.1.2 The three fundamental forces that are described by the quantum field theories of **quantum chromodynamics** and the electroweak theory.

The development of the Standard Model has been a mixture of theoretical prediction, followed by subsequent confirmation by experimentation, and experimental evidence leading to refinements in the theory. Experiments using **particle accelerators** have identified hundreds of subatomic particles apart from protons and neutrons. Physicists have been able to classify the particles they have discovered, on the basis of their interactions, into three groups: **leptons**, gauge bosons and hadrons.

The particles of the Standard Model

The huge array of particles that can be found in nature were once commonly known as the 'particle zoo'. This expressed the large variety of particles that had been discovered as being like different animals in a zoo. Physicists had found more than 100 particles by the early 1960s and gave them names such as the kaon, the lambda and the omega.

At the time, it was not known that all these particles had an underlying commonality. They were all thought to be fundamental at the time and to not be composed of other particles. After the discovery of the **quark**, it became clear that most of the known particles were made of quarks. All particles composed of quarks are called **hadrons**.

Quarks are never found by themselves in nature; they are only found in combinations of two or three, which results in the hundreds of other particles. The most common of these combinations are the proton and neutron, which are each comprised of three quarks. Figure 9.1.3 illustrates how the known particles are classified under the Standard Model.

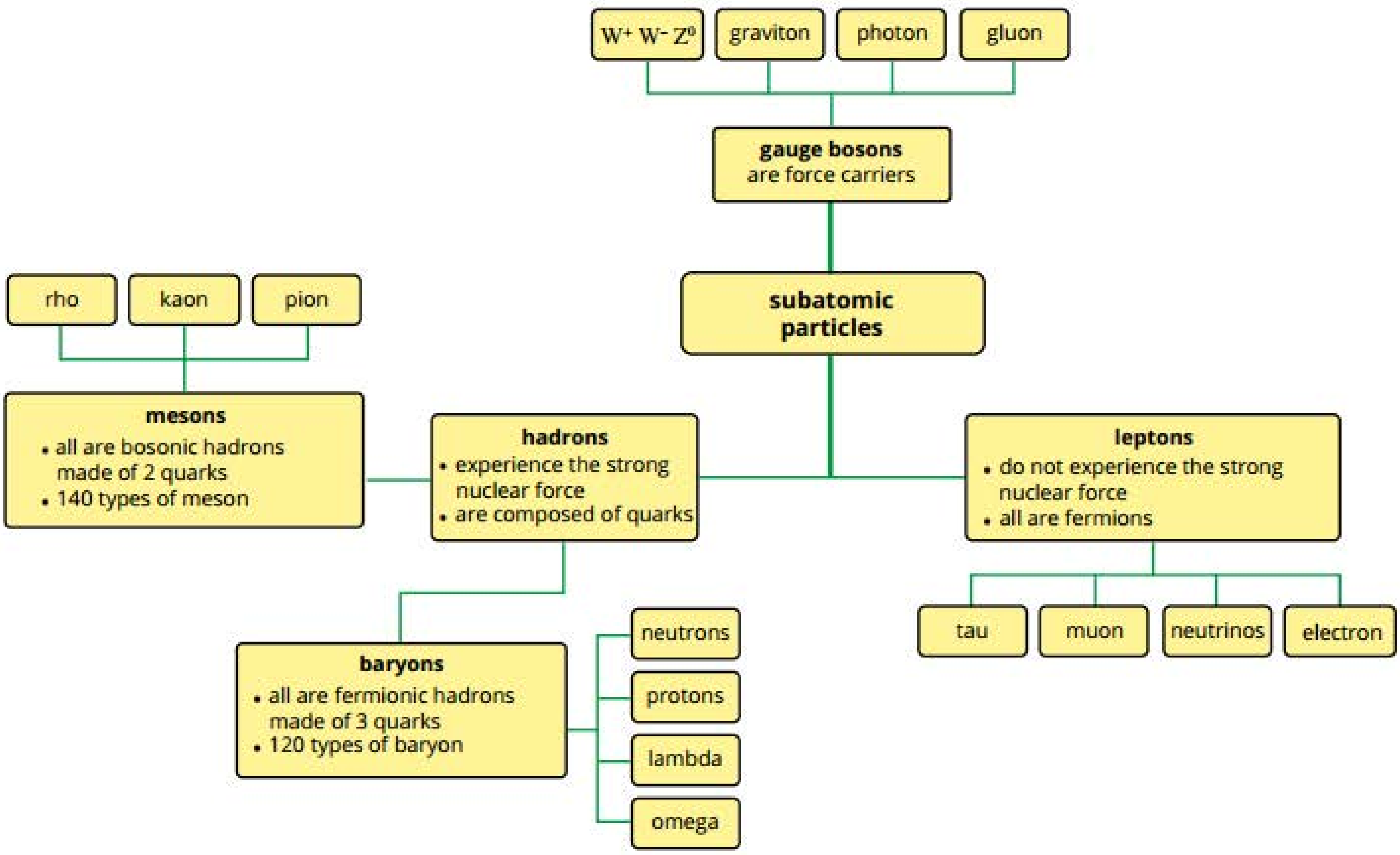


FIGURE 9.1.3 This mind map summarises the classification of particles in the Standard Model of particle physics.

FUNDAMENTAL FORCES AND THE GAUGE BOSONS

There are four forces that can act on particles and govern their behaviour. The fundamental assumption in the Standard Model is that forces arise through the exchange of particles called gauge bosons (or just bosons). Bosons are often called force-carrying, force-mediating or exchange particles.

Table 9.1.1 provides a summary of the nature of these particles, their strength, and the range over which they can exert a force.

TABLE 9.1.1 Comparing the four fundamental forces in nature

Force	Nature	Relative strength	Range (m)	Force carrier (gauge bosons)
strong nuclear	holds the nucleus of the atom together and acts between quarks	1	10^{-15} (~ diameter of an atomic nucleus)	gluon
electromagnetic	responsible for both electric and magnetic fields exerting forces of attraction or repulsion	$\frac{1}{137}$	infinite	photon
weak nuclear	causes radioactive decay	10^{-6}	10^{-18} (much less than the diameter of a proton)	W^+, W^- and Z
gravity	a force of attraction between any two objects with mass	6×10^{-39}	infinite	graviton (theoretical and unobserved)

EXTENSION

Why gravity is not part of the Standard Model

The Standard Model of particle physics successfully describes the strong nuclear, electromagnetic and weak nuclear forces. Unfortunately, there is no well-developed and experimentally proven quantum theory of gravity. If this existed it might be able to be combined with the electroweak theory and quantum chromodynamics (the theory that describes the interactions of the strong force) to create a grand unified theory (GUT) that could explain the fundamental behaviour of all matter in the universe.

The behaviour of objects due to the force of gravity is described with spectacular accuracy by Einstein's general theory of relativity. This theory basically says that mass tells space how to bend, and bent space tells mass how to move. It applies on the largest scales in the universe and even predicts how binary black holes should emit gravitational waves.

Quantum mechanics describes mathematically the behaviour of subatomic particles. This theory covers the interactions of matter and energy that occur on the smallest scales and is one of the foundations of the Standard Model of particle physics.

Einstein's theory of general relativity does not work for small-scale dimensions and high energies. Subatomic particles exist on very small scales and in many cases, are at very high energies. Therefore, the Standard Model of particle physics cannot accommodate gravity and its theoretical exchange particle, the graviton.

Forces through exchange of particles

Previously it was thought that forces were exerted on particles by fields. For example, there would be a region around a charged particle in which another charged particle would experience a force. This might seem quite puzzling, as the force is applied without any direct interaction by the two particles. In the Standard Model this is resolved. The mechanism explaining how forces are exerted involves the exchange of other particles.

An analogy can be used to show how a force can be exerted on two particles through the exchange of a third particle. On the left of Figure 9.1.4, two stationary inline skaters begin to pass footballs back and forth to each other. As they do this they will begin to move away from each other. This is due to the conservation of momentum each time they throw and catch a ball. This situation would be analogous to two particles experiencing a repulsive force.

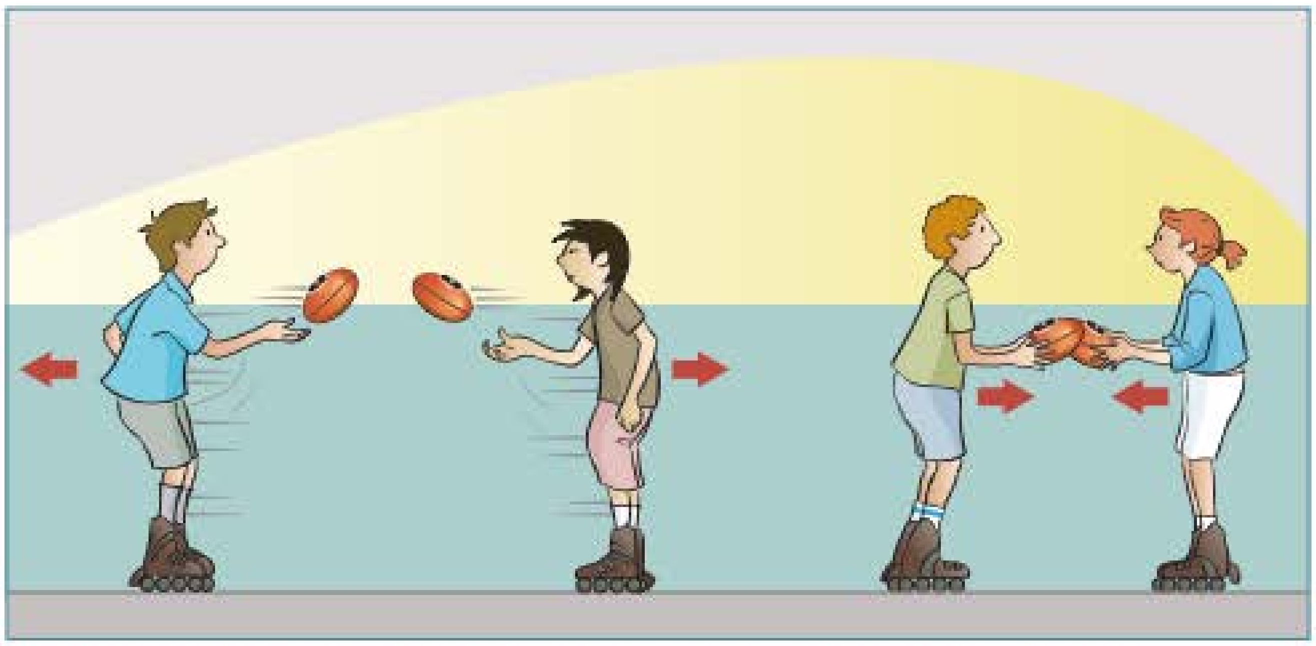


FIGURE 9.1.4 Inline skaters exchanging footballs act as an analogy for particles experiencing repulsive and attractive forces due to the exchange of gauge bosons.

A force of attraction can also be illustrated using this analogy. If the two skaters on the right of Figure 9.1.4 now exchange the footballs by grabbing them out of the other's hands, they exert a force of attraction on each other. This causes the skaters to move together, and is analogous to two particles experiencing an attractive force.

On a particle level, Figure 9.1.5 represents an electron approaching another electron. Each electron emits a photon that is absorbed by the other. This causes each electron to experience a force of repulsion, and the photons are responsible for the electromagnetic force acting on the two electrons.

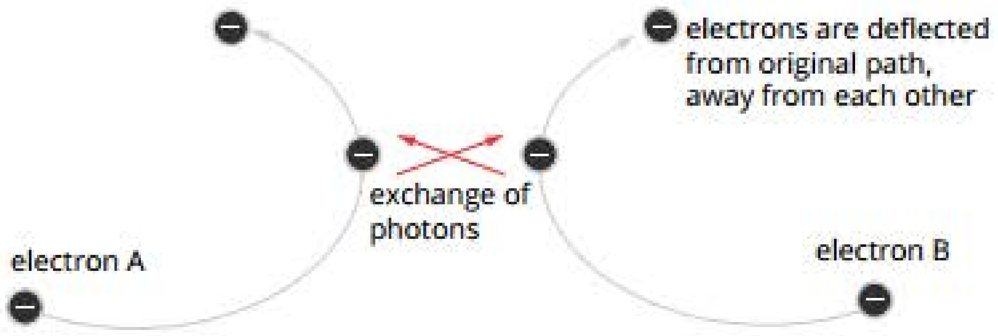


FIGURE 9.1.5 Two electrons approach each other, are repelled and then move away from each other. The two electrons exchange a photon, which is the carrier of the electromagnetic force.

PHYSICSFILE

The Higgs boson

Until the 1960s, physicists had no answer to the question of how particles get their mass. British physicist Peter Higgs, and others, proposed an answer to this question in 1964. Using the Standard Model of particle physics, they predicted that there was an as yet undiscovered particle that would interact with other particles and give them what we measure as mass. This particle would later be known as the Higgs boson.

In 2008, more than 10 000 physicists and engineers collaborated to build the Large Hadron Collider (LHC) near Geneva, Switzerland. This is the largest and most complex experimental facility ever built and one of its primary goals was to test the prediction of the existence of the Higgs boson.

The search began and in 2012 the discovery of a candidate for the Higgs particle was announced. On 10 December 2013, two of the original researchers, Peter Higgs and François Englert, were awarded the Nobel Prize in Physics for their work and prediction.

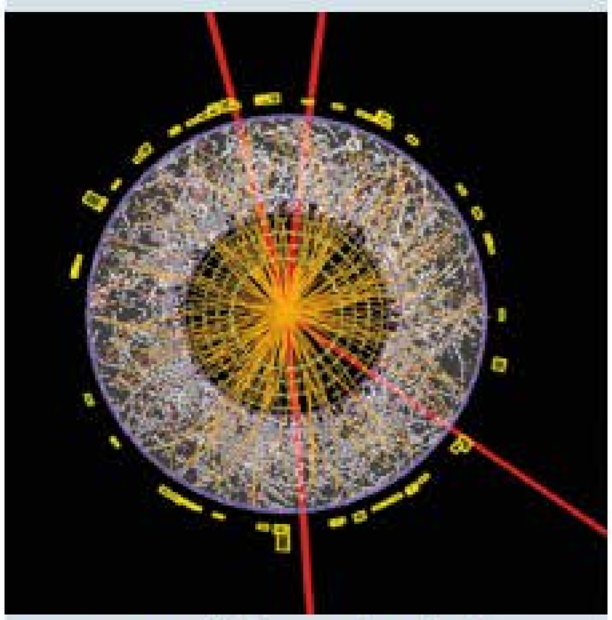


FIGURE 9.1.7 The image shows tracks of particles and measurements of their energies in the ATLAS detector at the LHC. The nature and energies of the particles produced are consistent with predictions of the formation of a Higgs boson.

FERMIONS

The Standard Model states that all particles of matter are comprised of one or more of the 12 fundamental or elementary particles. A fundamental particle is one that, to the best of our knowledge, it is not comprised of other smaller particles. All these particles are called **fermions** and are divided into two groups of six particles called quarks (blue) and leptons (green), shown in Figure 9.1.6.

	Charge			
Quarks	$+\frac{2}{3}$	up u 0.004	charm c 1.5	top t 176
Quarks	$-\frac{1}{3}$	down d 0.008	strange s 0.15	bottom b 4.7
Leptons	–1	electron e^- 5×10^{-4}	muon μ^- 0.1	tau τ^- 1.8
Leptons	0	electron neutrino ν_e $< 1 \times 10^{-8}$	muon neutrino ν_μ $< 1 \times 10^{-4}$	tau neutrino ν_τ $< 1 \times 10^{-2}$

Note: The number below the symbol for each particle is the mass and is given relative to the mass of a proton, ~1 GeV/c^2.

FIGURE 9.1.6 Fermions are the fundamental particles found in the universe.

All fermions (quarks and leptons) obey **Pauli's exclusion principle**, which states that no two particles can occupy the same quantum state at the same time. For example, no two electrons can be in the same shell or orbital around an atom and have the same energy. All fermions have the same property, called spin angular momentum, abbreviated to just 'spin'. All fundamental fermions have a spin of $+\frac{1}{2}$.

All quarks experience the strong nuclear force, which separates them from leptons, which do not interact via this force. Charged leptons exist as individual particles and interact via the electromagnetic force, while neutral leptons such as neutrinos do not interact much at all. Quarks also have non-integer (fractional) charges, unlike leptons, which have whole-number charges of −1 or 0.

All hadrons (baryons and mesons) are made of quarks held together by the strong nuclear force.

You are made of quarks, gluons and electrons

Most of the particles described by the Standard Model are not observed in everyday situations. They are made in high-energy particle collisions within machines called particle accelerators. The particles of which you are made consist of only a few fundamental ingredients.

Within the nucleus of all atoms, the protons and neutrons are made of three quarks that are held together by the strong nuclear force. All hadrons made of three quarks are called **baryons**. A proton consists of two up quarks and one down quark. A neutron consists of two down quarks and one up quark, as shown in Figure 9.1.8. The fractional charges of the quarks combine to give the proton a charge of +1 and the neutron a charge of 0.

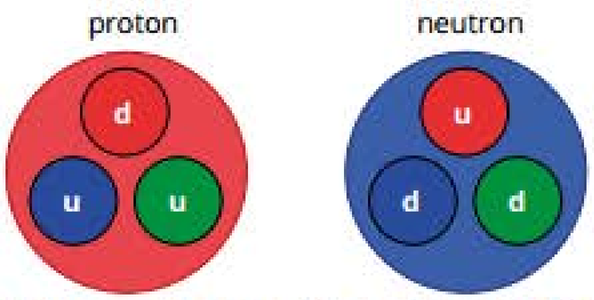


FIGURE 9.1.8 Protons and neutrons are baryons, which are all made of three quarks.

It is interesting to note that if you add the masses of the quarks in a proton (or a neutron), they add up to much less than the mass of the proton. This is because most of the mass of the proton is stored in **gluons**, the elementary particles that act as exchange particles for the strong nuclear force between quarks.

All hadrons that are made up of two quarks, a real quark and an antiquark, are called mesons. For example, the pion-plus (π^+) meson consists of an up quark and an anti-down quark, while the kaon-plus (K^+) meson consists of an up quark and an anti-strange quark.

MATTER VS ANTIMATTER

Most particles discussed so far have what is called an **antiparticle**. Each antiparticle has the same properties such as mass, spin and lifespan as the particle. However, their electric charges and another characteristic called the **quantum number** have the same magnitude but the opposite sign. For example, a positron has the same mass as an electron but a positive charge. Similarly, antiquarks have the same mass as a real quark but have an opposite fractional charge.

There are two conventions used to indicate that a particle is an antiparticle.

- The antiparticles of uncharged particles are indicated by placing a bar above the symbol for the normal matter particle. For example, an electron neutrino has the symbol ν_e and the antielectron neutrino has the symbol $\bar{\nu}_e$.
- The antiparticle of a charged particle is given the symbol of the particle but with the opposite sign. For example, the antiparticle of the muon (μ^-) is the antimuon (μ^+).

9.1 Review

SUMMARY

- The Standard Model of particle physics explains three of the four fundamental forces in the universe (electromagnetic force, the strong nuclear force and the weak nuclear force) in terms of an exchange of particles called gauge bosons.
- The gauge bosons (or just bosons) for these three forces are the photon, the gluon and the Z, W^- and W^+ for the electromagnetic force, the strong nuclear force and the weak nuclear force, respectively.
- All matter in the universe is made of fundamental particles called quarks and leptons.
- There are six quarks that all experience the strong nuclear force mediated by gluons. These combine to form hadrons and cannot exist alone. The hadrons include baryons, made of three quarks, and mesons, made of two quarks.
- Matter–antimatter pairs have similar properties, such as mass, spin and lifetime, but their electric charge and other characteristics called quantum numbers have the same magnitude but the opposite sign.

KEY QUESTIONS

1 Choose the correct responses from those given in bold to complete the sentences about the particles of the Standard Model.
The particles of the Standard Model have been classified into three main groups.
- Force-carrier particles such as the gluon are called **hadrons/gauge bosons/leptons**.
- Fundamental particles, such as the electron, that can be found individually and do not experience the strong nuclear force are called **hadrons/gauge bosons/leptons**.
- The particles that are largest in number, and are made of quarks, are called **hadrons/gauge bosons/leptons**.

2 Describe the Standard Model of particle physics using the following terms: leptons, quarks, the strong nuclear force, gauge bosons, electromagnetic force, gluons, photons, force-carrier particles, gravitational force, weak nuclear force, and Z, W^- and W^+ particles.

3 Which of the following statements about subatomic particles is correct?
A The particles you are made of are composed of quarks, leptons and gauge bosons.
B The particles you are made of are composed of quarks.
C The particles you are made of are composed of quarks and leptons.
D The particles you are made of are composed of leptons and gauge bosons.

4 A neutron in an evacuated container decays to produce a proton and two other particles. The proton is then attracted to an electron and becomes the nucleus of a hydrogen atom. The atom then slowly drifts to the base of the container.
Describe the forces that have been involved in the events described above in order of the sequence in which they occurred. State which forces (if any) are not involved at all.

5 Which of the following statements correctly compares the two groups of particles within the fermions (quarks and leptons)?
A Particles in one of these groups are charged and those in the other are not.
B Particles in one of these groups must exist in groups of two or three and those in the other group can exist individually.
C Both groups of particles are conserved in the same way in decays and reactions.
D The particles in both groups experience the strong nuclear force.

6 Classify each of the following as a gauge boson, a lepton or a hadron.
a gluon
b neutrino
c neutron
d photon
e electron
f muon
g proton
h tau

7 A common analogy used to explain how forces are mediated by the exchange of particles involves two skaters passing balls to each other. Which of the following correctly describes what a ball represents in this analogy?

A The ball represents a force of attraction or repulsion.
B The ball represents the particle being exchanged.
C The ball represents a particle experiencing a force.
D The ball represents a particle exerting a force.

8 Complete the following table to summarise the two types of fundamental matter particles.

<table>
<tr><td></td><th colspan="3">Fermions</th></tr>
<tr><td>Type of fermion</td><td colspan="2"></td><td></td></tr>
<tr><td>Gauge boson</td><td colspan="2"></td><td></td></tr>
<tr><td>Charge</td><td colspan="2"></td><td></td></tr>
<tr><td>Hadrons formed</td><td></td><td></td><td></td></tr>
<tr><td>Number of particles in the hadron</td><td></td><td></td><td></td></tr>
</table>

9 Which of the following correctly indicates the quarks that make up protons and neutrons?

A Proton (up, up, down) and neutron (down, down, down)
B Neutron (up, down, down) and proton (up, up, up)
C Neutron (up, up, down) and proton (up, down, down)
D Proton (up, up, down) and neutron (up, down, down)

10 Write the correct symbols for the following particles.

a positron
b tau antineutrino
c tau neutrino
d muon
e antimuon

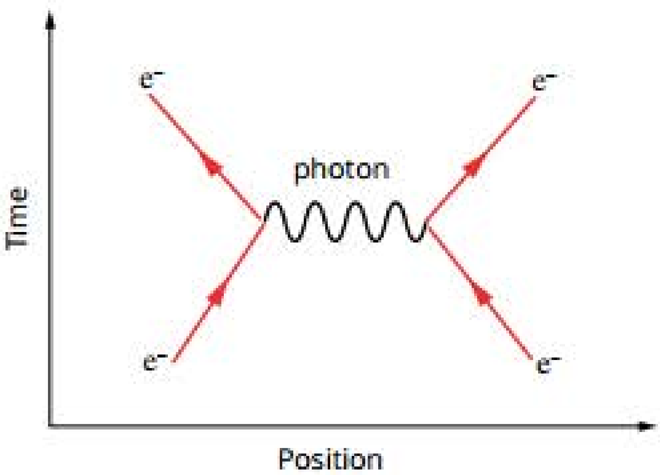


FIGURE 9.2.1 The Feynman diagram for the interaction between two electrons.

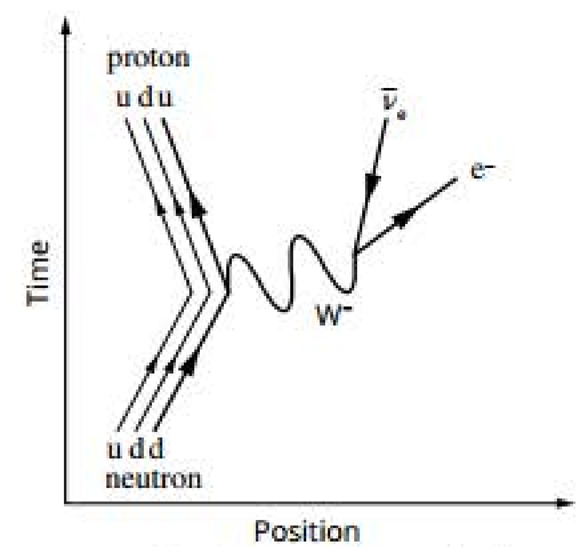


FIGURE 9.2.2 The Feynman diagram for the decay of a neutron.

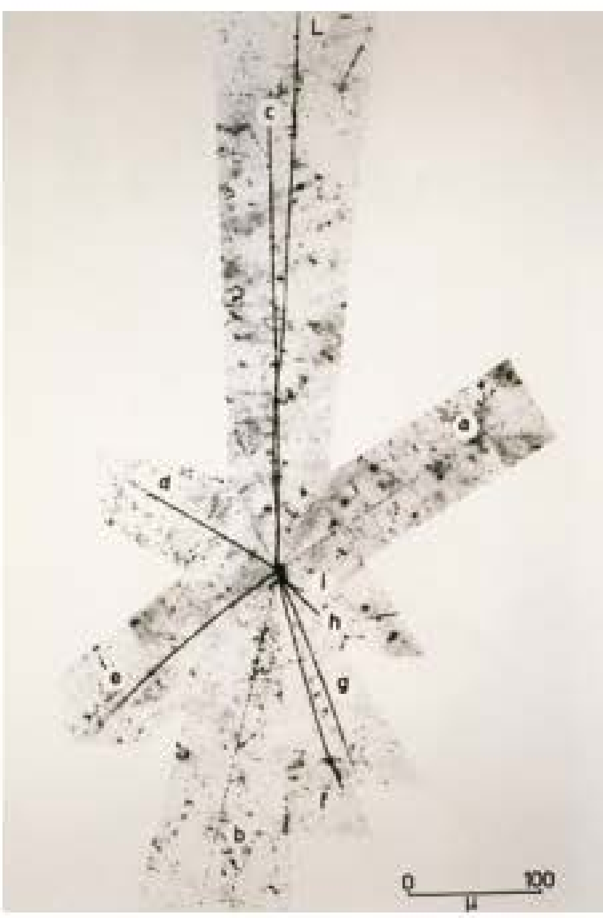


FIGURE 9.2.3 This image records a proton–antiproton annihilation. This event was recorded in 1955 in a photographic emulsion at the Bevatron accelerator at the Lawrence Berkeley Laboratory, California.

9.2 Interactions between particles

Particle physicists have been able to develop the Standard Model by observing countless interactions between particles. They use special relativity and conservation laws that particles must obey to analyse these observations and test their predictions.

In developing an understanding of an aspect of the universe, such as the behaviour of particles, physicists are not happy with just observing and recording the behaviour. The goal of the physicist is to discover the 'rules of the game'. Once the rules are known, they can then be used to explain what is observed and to predict what else is possible. These rules may come from a large number of observations (i.e. they can then be determined from watching for long enough), or a theoretical idea can be proposed, usually derived mathematically, and experiments conducted to verify the prediction.

INTERACTIONS, EVENTS AND EXCHANGE PARTICLES

Interactions between particles that produce some change, or are observed by a particle detector, are commonly known as events. Those interactions that result in the formation of new particles are sometimes referred to as reactions. Interactions can involve attractive or repulsive forces, decay or **annihilation**.

An example of a repulsive interaction involving the electromagnetic force is the scattering of one electron by another. This example is illustrated using a diagram called a **Feynman diagram** (Figure 9.2.1). It shows two electrons approaching each other (on the horizontal axis) over time (the vertical axis) and exchanging a photon. This causes each electron to be repelled and therefore change direction.

An example of a decay is the decay of a neutron into a proton. This involves the weak nuclear force and is brought about by the conversion of a down quark in the neutron into an up quark. This requires the emission of a W^- boson, which decays into an antielectron neutrino and an electron (Figure 9.2.2). This is commonly known as a beta-minus radioactive decay.

It is important to note that the directions of the arrows in Figures 9.2.1 and 9.2.2 do not indicate the direction in which a particle travels, but rather how a particle travels in time; for example, an antiparticle moves backwards in time. The simplest information you can gather from these diagrams is of the particles involved. You can also tell from the time axis which particles existed originally and which resulted from the decay or reaction. There are complex rules that govern these diagrams, as they are a graphical representation of mathematical expressions.

MATTER MEETS ANTIMATTER: ANNIHILATION

When physicist Paul Dirac first proposed antimatter, he did more than just predict its existence. He also predicted that matter and antimatter will annihilate when they collide. This is observed and exploited on a daily basis in experiments and applications involving antimatter.

Figure 9.2.3 shows an antiproton entering along the track marked L (top), before colliding with a proton. When the proton and antiproton mutually annihilate, the huge amount of energy released produces several new particles whose tracks form the 'star' pattern in this image.

Another example is the annihilation of an electron and a positron. This produces two photons that carry away the energy initially contained in the electron and positron. These events are a verification of Einstein's famous equation $E = mc^2$, where mass (m) and energy (E) are equivalent and can be converted from one form to another.

The opposite of an annihilation is also observed. This is called particle–antiparticle **pair production**. Energy in the form of a photon can create a particle–antiparticle pair if the photon has energy greater than or equal to the mass of the particle–antiparticle pair. Pair production also illustrates the relationship between mass and energy in Einstein's equation.

PARTICLE ENERGIES: THE ELECTRON-VOLT

To unravel the details of any particle interaction it is essential to determine the properties of the particles involved, such as their charge, and to measure their energies. A convenient unit when dealing with the energies of particles is the electron-volt (eV). One electron-volt is the energy given to an electron when it moves through a potential difference of 1 V, as shown in Figure 9.2.4.

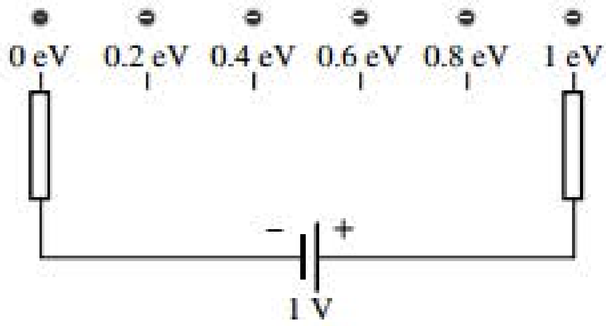


FIGURE 9.2.4 An electron is shown at equal time intervals as it is accelerated across a potential difference of 1 V, giving it kinetic energy of 1 eV. Kinetic energy is lowest at the negative (left) end and gradually increases as the electron moves towards the positive (right) end.

> By definition, voltage (V) is an amount of work (W) per unit charge (q).
>
> $$V = \frac{W}{q}$$
>
> So then the work done accelerating an electron is equal to the product:
>
> $$W = Vq$$
>
> For a potential of 1 V, the amount of work done on an electron is:
>
> $$W = 1 \times 1.60 \times 10^{-19}\,\text{C}$$
> $$= 1.60 \times 10^{-19}\,\text{J}$$
>
> Therefore $1\,\text{eV} = 1.60 \times 10^{-19}\,\text{J}$.

In particle accelerators, it is common to accelerate a particle through thousands of volts. An electron accelerated through 10^6 V gains 10^6 eV of kinetic energy or 1 MeV. Modern accelerators achieve particle energies in the 1 GeV (10^9 eV) and 1 TeV (10^{12} eV) range.

Einstein's famous equation $E = mc^2$ tells us that mass and energy are equivalent. Essentially, they are two forms of the same thing and are sometimes referred to as mass–energy to indicate their close connection. It is common in particle physics to express both the energy given to an accelerated particle and the mass of a particle in electron-volts, as this is an amount of energy.

Worked example 9.2.1

ELECTRON-VOLTS AND JOULES

Calculate the energy given to an electron accelerated across a potential difference of 10 000 V. Give your answer in joules.	
Thinking	**Working**
Determine the number of electron-volts of energy that the particle gains.	$1\,\text{eV} \times 10\,000 = 10\,000\,\text{eV}$ or 10 keV
Convert the value in eV to J.	$10\,000 \times 1.60 \times 1^{-19}\,\text{J} = 1.60 \times 10^{-15}\,\text{J}$

Worked example: Try yourself 9.2.1

ELECTRON-VOLTS AND JOULES

> Calculate the energy given to an electron accelerated across a potential difference of 2.5×10^6 V. Give your answer in joules correct to two significant figures.

CONSERVATION LAWS: MASS-ENERGY

Some of the rules that govern the interactions of particles have already been covered in this course. These are the laws of conservation of momentum, and conservation of energy.

Energy is conserved in a particle interaction. It is now accepted that mass and energy are equivalent, so when considering the conservation of energy, you must consider two things:

- the total energy contained in the rest masses of the particles, and
- the kinetic energy before and after the event.

A simple example is the decay of a stationary neutron to produce a proton, an electron, and an antielectron neutrino, as shown in Figure 9.2.5.

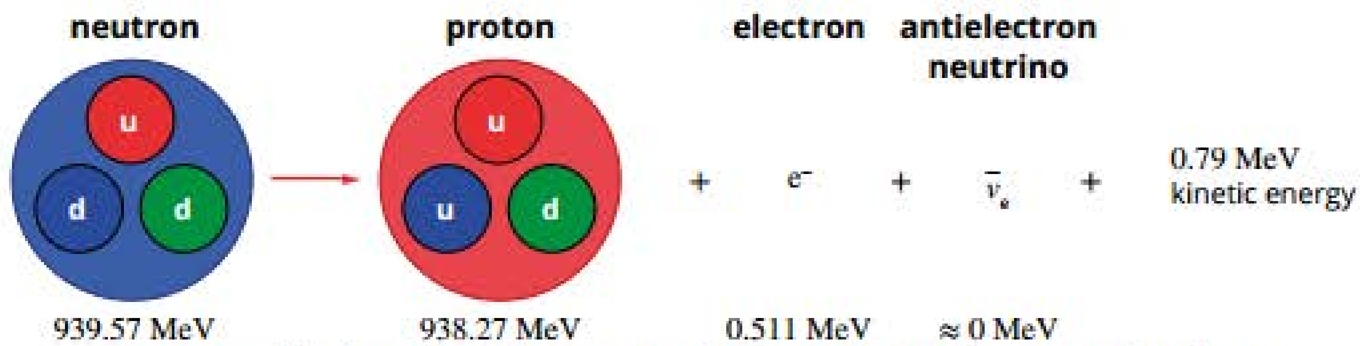


FIGURE 9.2.5 The decay of a stationary neutron is shown to illustrate the conservation of mass–energy. The total energy of the system is conserved.

The products of the decay of a stationary neutron have a total mass less than that of the original neutron. Since the neutron is originally stationary, the balance of the energy must be carried away as the kinetic energy of the products. The energy in the mass of the particles is determined using $E = mc^2$, where m is the rest mass. So the mass of a particle is often expressed in units of energy.

This decay will obey the law of conservation of energy, as the energy stored within the mass of the neutron before the decay is equal to the sum of the energies stored within the mass of the products, plus their kinetic energies, after the decay. The momentum carried away by the products would have to be in different directions, so it adds to equal zero. This would mean that the **law of conservation of momentum** is obeyed and the total momentum before (zero) is equal to the total momentum after (zero overall).

Worked example 9.2.2

CONSERVATION OF MASS–ENERGY

A 2 MeV gamma-ray photon interacts with an atomic nucleus and an electron–positron pair is produced. Each particle created has a mass of 0.511 MeV. How much kinetic energy is carried away by the particle–antiparticle pair produced?

Thinking	Working
The energy of the gamma-ray photon is conserved and is equal to the total mass–energy of the particles produced plus the kinetic energy (E_k) of those particles.	$2\text{ MeV} = 2 \times 0.511\text{ MeV} + E_k$
Solve for the kinetic energy.	$2\text{ MeV} = 1.022\text{ MeV} + E_k$
Calculate the answer.	$E_k = 0.978\text{ MeV}$

Worked example: Try yourself 9.2.2

CONSERVATION OF MASS–ENERGY

A 1.20 MeV gamma-ray photon interacts with an atomic nucleus and an electron–positron pair is produced. Each particle created has a mass of 0.511 MeV. How much kinetic energy is carried away by the particle–antiparticle pair produced?

CONSERVATION OF CHARGE AND QUANTUM NUMBERS

There are other quantities that must be conserved in particle interactions, together with energy and momentum. This includes electric charge; the total charge present before an event is equal to the total charge after the event.

For example, in the decay of a stationary neutron, the total charge of the products must balance to produce an overall charge of zero. The charge on the reactant (a neutron) is zero. A proton has a charge of +1 and an electron has a charge of −1, which means the charges cancel each other and the total charge of the products is zero. Therefore, charge is conserved.

The values that express the charge of a particle are an example of what is known as a quantum number. You have just seen that the quantum number for charge is conserved in particle decays.

Worked example 9.2.3

CONSERVATION OF CHARGE

A physicist proposed that the products of an observed collision of a proton and a neutron were two protons and an antiproton. Is this reaction allowed according to the law of conservation of charge?	
Thinking	**Working**
To find out if this reaction is allowed, calculate the total charge present before and after the interaction.	Proton's charge = +1 Neutron's charge = 0 Antiproton's charge = −1 Total charge before = +1 + 0 = +1 Total charge after = (+1) + (+1) + (−1) = +1
Compare the total charge before and after the interaction. If they are equal then the reaction is allowed; if they are different then the reaction is forbidden.	The total charge before and after is the same. Therefore, this proposed reaction is allowed according to the law of conservation of charge.

Worked example: Try yourself 9.2.3

CONSERVATION OF CHARGE

A physicist proposed that the products of an observed collision of a proton and a neutron were two protons, a neutron and an antiproton. Determine whether this reaction is allowed according to the law of conservation of charge.

Even though the reaction proposed in Worked example 9.2.3 is allowed by the conservation of charge, it may not be allowed under other **conservation laws**. You will see some of these in the next section. Revisit Worked example: Try yourself 9.2.3 after you have completed this section and work out which conservation laws forbid that proposed reaction.

OTHER CONSERVED QUANTUM NUMBERS

Particle physicists strike a serious problem when they study new particle interactions. Even with the laws of conservation of energy, momentum and charge, a multitude of products are theoretically possible with the energy available.

Observations undoubtedly show that only a small number of the possible outcomes of an interaction actually occur. Once a large number of events have been observed, it may become clear there are other 'rules' that are important. This means that other quantum numbers are needed to summarise the rules.

These rules are determined by observing a large number of events and testing the proposed rules against them. Once a large enough number and range of interactions have been observed that all obey the newly proposed rules, the new quantum number is assigned to those particles. This quantum number can then be used to predict the outcomes of interactions involving those particles.

Examples of conserved quantum numbers that have been determined in this way are **baryon number** and **lepton number**.

Conservation of baryon number

Baryons are all composite particles containing combinations of three quarks. You will recall from Section 9.1 that two of the common baryons are neutrons and protons.

Baryons are given a quantum number, B, of +1. For antibaryons $B = -1$, and for all particles other than baryons involved in the interaction or decay $B = 0$.

The law of conservation of baryon number states that the sum of the baryon numbers before an interaction must equal the sum of the baryon numbers after the interaction. Another way of stating this is that the net number of baryons remains constant in any process.

Worked example 9.2.4

CONSERVATION OF BARYON NUMBER

A physicist proposed that the products of an observed collision of a proton and a neutron were two protons and an antiproton. Is this reaction allowed according to the law of conservation of baryon number?

Thinking	Working
The law of conservation of baryon number states that the net number of baryons remains constant in any process. To find out if this reaction is allowed, calculate the total baryon number before and after the interaction.	Proton $B = +1$ Neutron $B = +1$ Antiproton $B = -1$ Total B before = (+1) + (+1) = +2 Total B after = (+1) + (+1) + (−1) = +1
Compare the total baryon number before and after the interaction. If they are equal then the reaction is allowed according to the law of conservation of baryon number. If they are different then the reaction is forbidden, as baryon number is not conserved.	The total baryon number before the reaction is different from the total baryon number after the reaction. Therefore, this proposed reaction is forbidden according to the law of conservation of baryon number.

Worked example: Try yourself 9.2.4

CONSERVATION OF BARYON NUMBER

A physicist proposed that the products of an observed collision of a proton and a neutron were two protons, a neutron and an antiproton. Is this reaction allowed according to the law of conservation of baryon number?

Conservation of lepton number

Leptons are the fundamental particles shown in Figure 9.2.6.

Charge	Leptons		
−1	electron e^-	muon μ^-	tau τ^-
0	electron neutrino ν_e	muon neutrino ν_μ	tau neutrino ν_τ

FIGURE 9.2.6 The table shows the six leptons and their charges.

Lepton number, L, is a quantum number with value of +1 for leptons, −1 for antileptons and 0 for other particles that are not leptons. This number has been introduced to represent the experimental observation that the net number of leptons in any reaction is unchanged (conserved).

The law of conservation of lepton number is more complicated than that for baryons. There are actually three types of lepton numbers: one type for the electron and the electron neutrino, L_e; one type for the muon and the muon neutrino, L_μ; and one type for the tau and the tau neutrino, L_τ. Each of these three lepton numbers must be considered and conserved independently.

Worked example 9.2.5

CONSERVATION OF LEPTON NUMBER

A physicist proposed that the products of an observed decay of a muon (μ^-) were an electron (e^-), an antielectron neutrino ($\overline{\nu}_e$) and a muon neutrino (ν_μ). Is this reaction allowed according to the law of conservation of lepton number?

Thinking	Working
Identify which types of leptons are involved in the interaction. These need to be considered separately.	Electron/electron neutrino, L_e: yes Muon/muon neutrino, L_μ: yes Tau/tau neutrino, L_τ: no
To find out if this reaction is allowed, calculate the total electron lepton number L_e before and after the interaction.	Electron $L_e = +1$ Antielectron neutrino $L_e = -1$ Total before $L_e = 0$ Total after $L_e = (+1) + (-1) = 0$
Compare the total electron lepton number before and after the interaction. If they are equal then the reaction is allowed. If they are different then the reaction is forbidden.	The total electron lepton number before the reaction is the same as the total electron lepton number after the reaction. Therefore, this proposed reaction is allowed according to the law of conservation of electron lepton number.
To find out if this reaction is allowed, calculate the total muon lepton number L_μ before and after the interaction.	Muon $L_\mu = +1$ Muon neutrino $L_\mu = +1$ Total before $L_\mu = +1$ Total after $L_\mu = +1$
Compare the total muon lepton number before and after the interaction. If they are equal then the reaction is allowed. If they are different then the reaction is forbidden.	The total muon lepton number before the reaction is the same as the total muon lepton number after the reaction. Therefore, this proposed reaction is allowed according to the law of conservation of muon lepton number.

Worked example: Try yourself 9.2.5

CONSERVATION OF LEPTON NUMBER

A physicist proposed that the observed decay of a meson (made of two quarks) called a pion (π^-) produced an electron neutrino (ν_e), a muon neutrino (ν_μ) and an antimuon (μ^+). Is this reaction allowed according to the law of conservation of lepton number?

It is important to remember that if an interaction breaks any one of the conservation laws, then according to the Standard Model that interaction is not allowed, even if it meets all the other conservation laws.

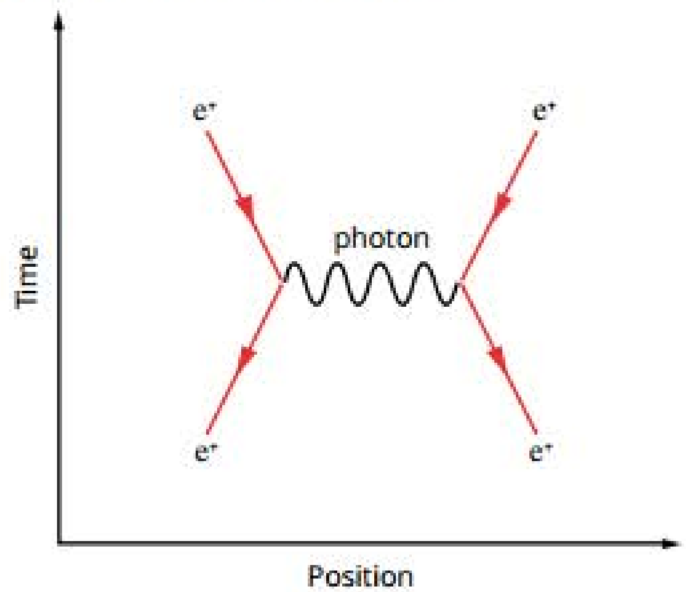
9.2 Review

SUMMARY

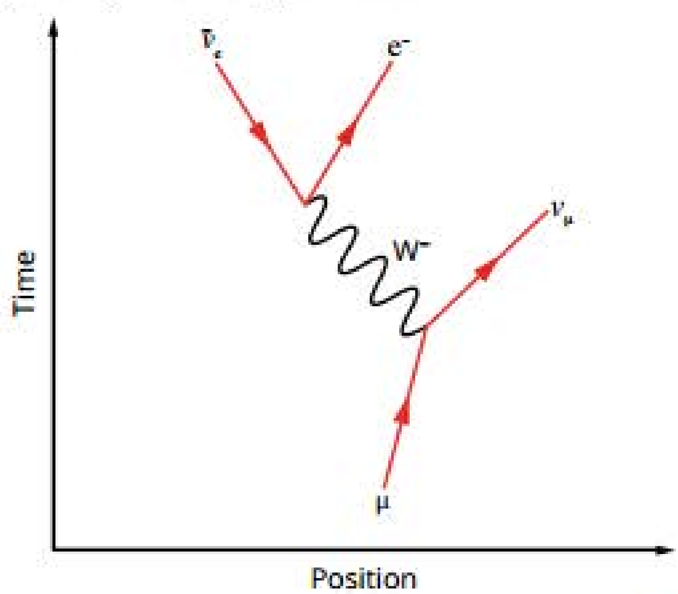
- The development of the Standard Model has been a mixture of theoretical prediction, followed by confirmation by experimentation, and experimental evidence leading to refinements in the theory.
- Mass and energy are interchangeable according to Einstein's famous equation $E = mc^2$. The masses of particles are often quantified as an equivalent amount of energy and the available energy in an interaction includes the rest mass and the kinetic energy.
- Particle energies are often expressed in electron-volts, but calculations should always be done in the SI unit of joules.
- Energy, momentum, charge, baryon number and lepton number are some of the quantities that must all be conserved in particle interactions according to the Standard Model.

KEY QUESTIONS

1 Describe the common particle interaction represented by the Feynman diagram below.



2 Describe the common particle interaction represented by the Feynman diagram below.



3 An electron is accelerated between two charged plates across a potential difference of 1.4×10^3 V. How many electron-volts of energy will the electron have gained when it is half way between the plates?

4 A decay is proposed in which a stationary neutron decays into a proton, a muon and an antimuon. This is summarised below including the masses of the particles. Use the laws of conservation of energy and charge to determine whether this decay is allowed or forbidden.

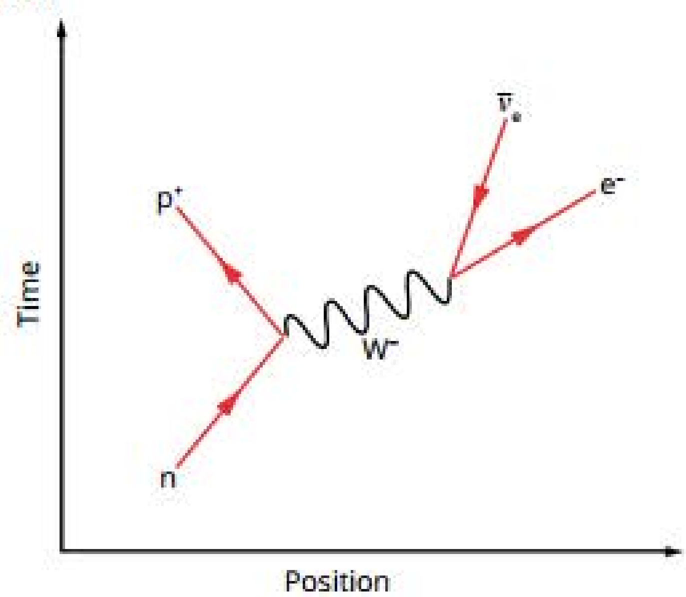
neutron		proton		muon		antimuon
n	→	p^+	+	μ^-	+	μ^+
939.57 MeV		938.27 MeV		105.7 MeV		105.7 MeV

5 Compare annihilation and pair-production, then state what happens when an electron meets a positron.

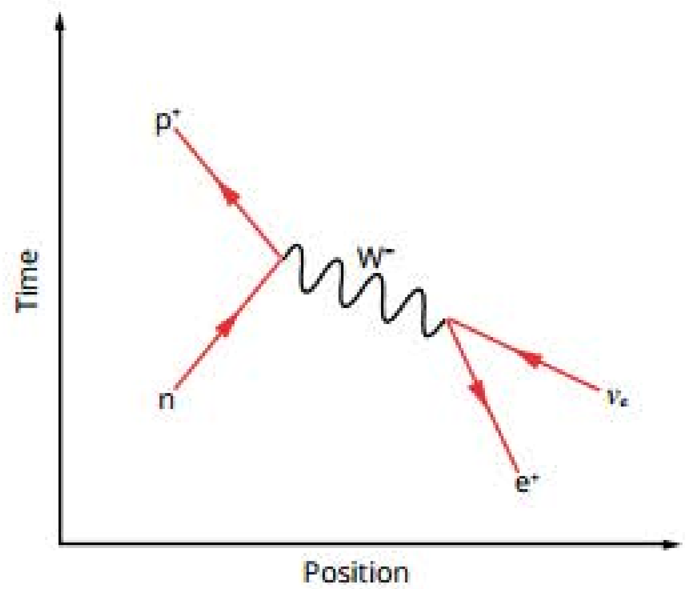
6 Describe how conservation laws help to predict interactions or decays.

7 Which of the Feynman diagrams A–D shows a decay in which a proton is converted into a neutron, and the other products are a positron and an electron neutrino?

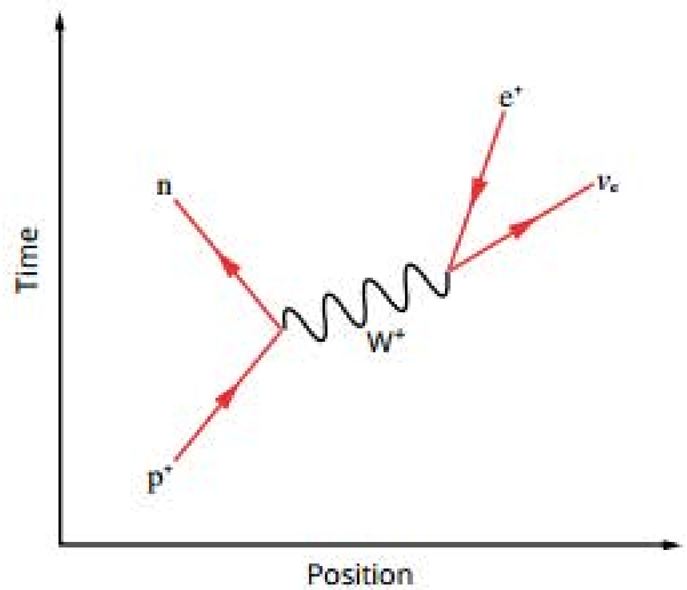
A



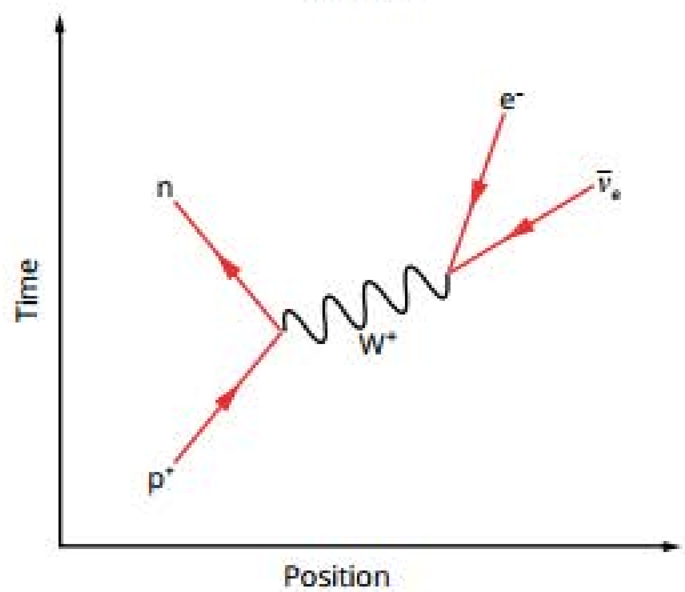
B



C



D



8 Determine which conservation law is violated in each reaction below, making the reaction forbidden.

a $p^+ + p^- \rightarrow \mu^+ + e^-$

b $p^+ + p^+ \rightarrow p^+ + \pi^+$

c $\pi^- + p^+ \rightarrow p^+ + \pi^+$

9 Use conservation laws to determine the name of the missing product X in the following reaction:

$n \rightarrow p^+ + e^+ + X$

10 Your classmate proposed that a neutron could decay into a proton, an electron and one other particle. Use the law of conservation of lepton number to determine what this particle would be if the decay was allowed.

9.3 Particle accelerators

Particle accelerators are the main tools of the physicists who study the universe on the smallest scales. They are the tools required to test theoretical predictions made by the complex mathematical models that are the parts of the Standard Model of particle physics.

Discoveries in particle physics are often unexpected, and these surprises are just around the corner with every new accelerator that pushes our experiments to higher and higher energies. Even now, new particles are being found, made of more and more exotic combinations of the fundamental particles.

Making particles move at near the speed of light and collide head-on is very challenging. It requires the most complex machines that have ever been built. These machines are designed, built and operated by thousands of scientists and engineers from around the world. No one country has the resources to overcome the expense and difficulties that are involved.

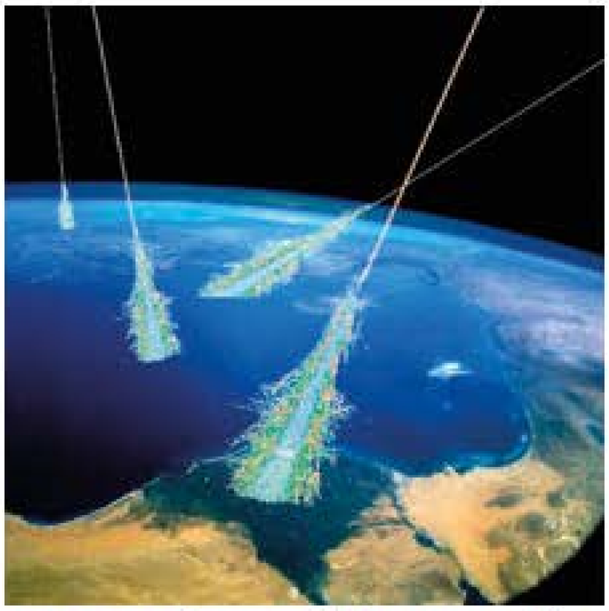


FIGURE 9.3.1 Showers of high-energy particles occur constantly as energetic cosmic rays strike Earth's atmosphere.

PARTICLE EXPERIMENTS

Many of the earliest experiments studied particles that occur naturally due to cosmic rays. An example of this was the discovery of the positron. Cosmic rays are high-energy particles that come from space and produce a cascade of other high-energy particles that reach the Earth's surface (Figure 9.3.1). Experiments using cosmic-ray-induced particles are limited to the naturally available energies and by the number of events.

To test theories about high-energy physics, particle accelerators must be constructed to accelerate matter to high velocities under controlled conditions. Particles often need to be accelerated to very high velocities to initiate reactions by overcoming repulsive forces between the nuclei. They can also achieve high enough energies to produce particles more massive than themselves. Another consequence is that with higher velocities comes greater resolution to study finer structure.

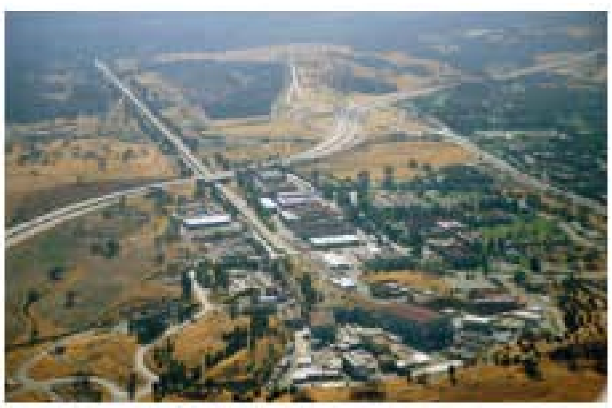


FIGURE 9.3.2 Aerial photograph of the Stanford Linear Accelerator Center, California, showing the 3.2 km linear accelerator housed in the world's longest building. Particles are accelerated from the top left to the detectors in the foreground.

The resolution of a scattering microscope is described by de Broglie's wavelength formula:

$$\lambda = \frac{h}{mv} = \frac{h}{p}$$

This illustrates that the higher the momentum (p) of a particle, the shorter its wavelength (λ). The shorter the wavelength used to study a substance, the finer the detail that can be seen. Accelerators have steadily increased in energy to allow particles to be given more and more velocity, and therefore momentum, to study the fine details within atoms and composite particles such as hadrons.

Experimental evidence for quarks

Particle accelerators have been used to test the Standard Model and provided the first experimental evidence supporting the existence of quarks. Experiments were performed in the 1960s and 1970s to study electrons scattering off protons and neutrons. These were performed using the Stanford Linear Accelerator (SLAC), in California (Figure 9.3.2), which accelerated electrons to energies of between 4 GeV and 21 GeV.

FIGURE 9.3.3 The End Station, an experimental hall at the Stanford Linear Accelerator Center, contains three giant particle spectrometers that detect particles of different energies and scattered at different angles.

In this experiment, giant particle spectrometers, such as the one shown in Figure 9.3.3, observed the scattering angle and energy loss of the electrons after the collision. Earlier investigations of the proton at low energies had shown that there should be a relatively even internal distribution of electric charge. The results of these new investigations gave the surprising result that the electric charge within the proton is concentrated in smaller components of negligible size. This was soon interpreted as a verification of the quark model. Analysis of the results also gave the first hint of the existence of the gluons, which are responsible for the strong nuclear force that holds quarks together.

The physicists in charge of these experiments (Professors Jerome I. Friedman, Henry W. Kendall and Richard E. Taylor) were awarded the Nobel Prize in Physics in 1990 for 'their pioneering investigations concerning deep inelastic scattering of electrons on protons and bound neutrons, which have been of essential importance for the development of the quark model in particle physics'.

Linear accelerators

The SLAC consists of a long, evacuated cylinder that contains a set of hollow metal drift tubes. Every drift tube is electrically connected to an alternating power supply, which provides an alternating potential difference of constant frequency. A charged particle is injected into the accelerator and is attracted towards the first drift tube. Once the particle is in the tube, the electric field drops to zero and the polarity of the tube is reversed. When the particle emerges from the first drift tube, it is the field between the first and second drift tube that accelerates the particle. The particle enters the next tube and the process is repeated. The operation of the SLAC is shown in Figure 9.3.4.

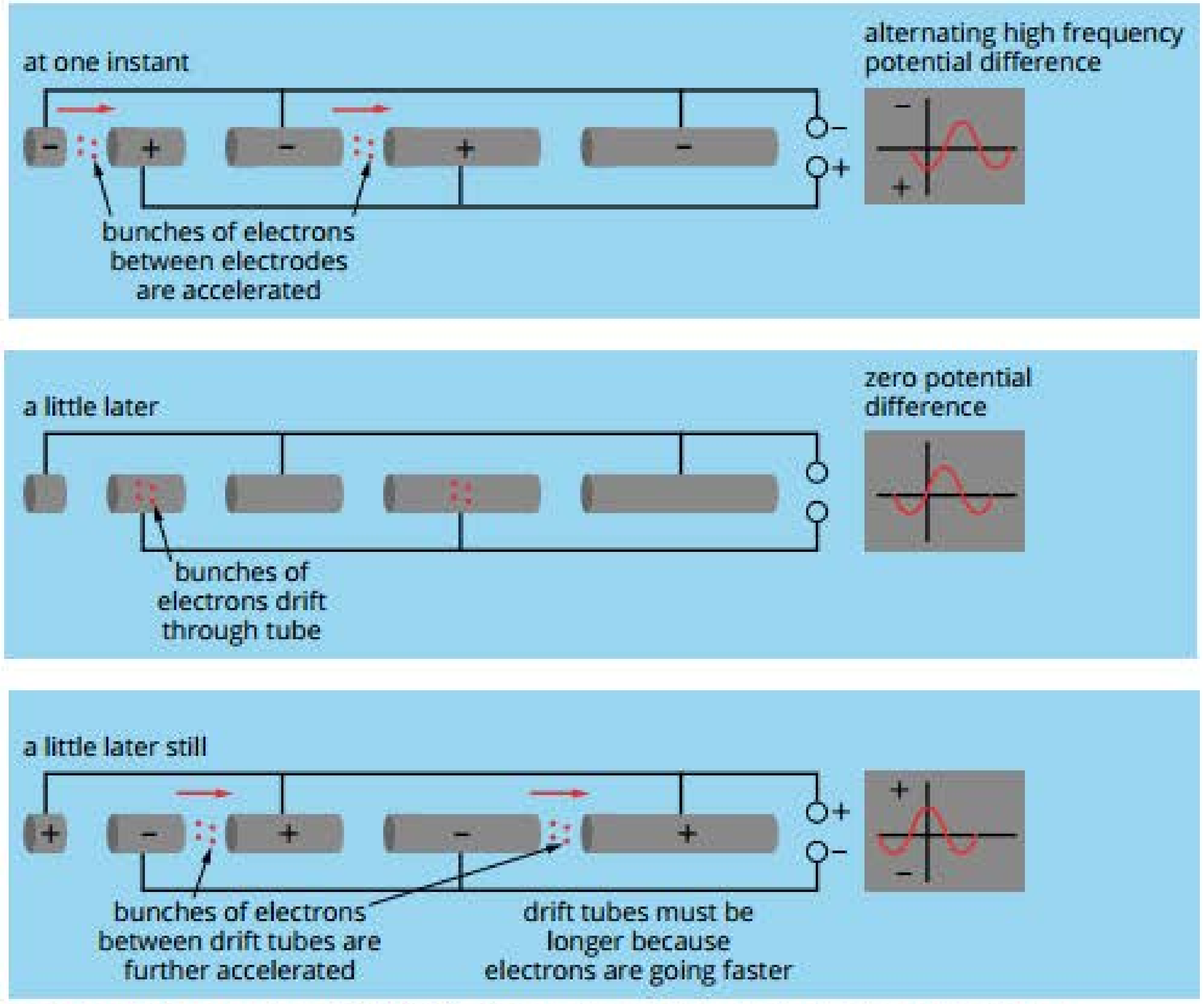


FIGURE 9.3.4 The operation of the SLAC relies on alternating electric fields along its length to continually accelerate electrons.

Linear accelerators played a large role in testing the Standard Model of particle physics, but the quest for higher collision energies led to the use of another type of accelerator, the synchrotron.

SYNCHROTRONS

A large portion of the energy of a particle hitting a fixed target, such as in a traditional linear accelerator, is converted into the momentum of the products. This energy is mostly wasted. In the quest to produce more massive particles using collisions, the goal is to use that energy to produce the mass of the particle in question.

One solution to this problem is to collide two particles of identical mass and velocity, but moving in opposite directions. This means the total momentum before the collision is zero. According to the law of conservation of momentum, the total momentum after the collision should also be zero. This means that only a relatively small amount of the collision energy is converted into the momentum of the products.

A synchrotron accelerator is the design of choice for many modern experiments, including the search for the Higgs boson at the Large Hadron Collider (LHC).

The Large Hadron Collider synchrotron

The LHC, the world's largest and highest-energy particle accelerator, was built by the European Organization for Nuclear Research (also known as CERN). The collider is located inside a circular tunnel with a circumference of 27 km. This tunnel is at a depth ranging from 50 to 175 m underground, near Geneva, Switzerland.

This type of synchrotron has four main components: the evacuated beam pipes, the magnet system, the radio frequency (RF) accelerating cavities, and detectors (Figure 9.3.5). Particles are confined to a circular path by using bending magnets, and an electric field in the acceleration cavities exerts a net force on the particles during each orbit, causing acceleration. The particles moving in opposite directions are then collided and the products of the collision are identified and quantified by the detectors.

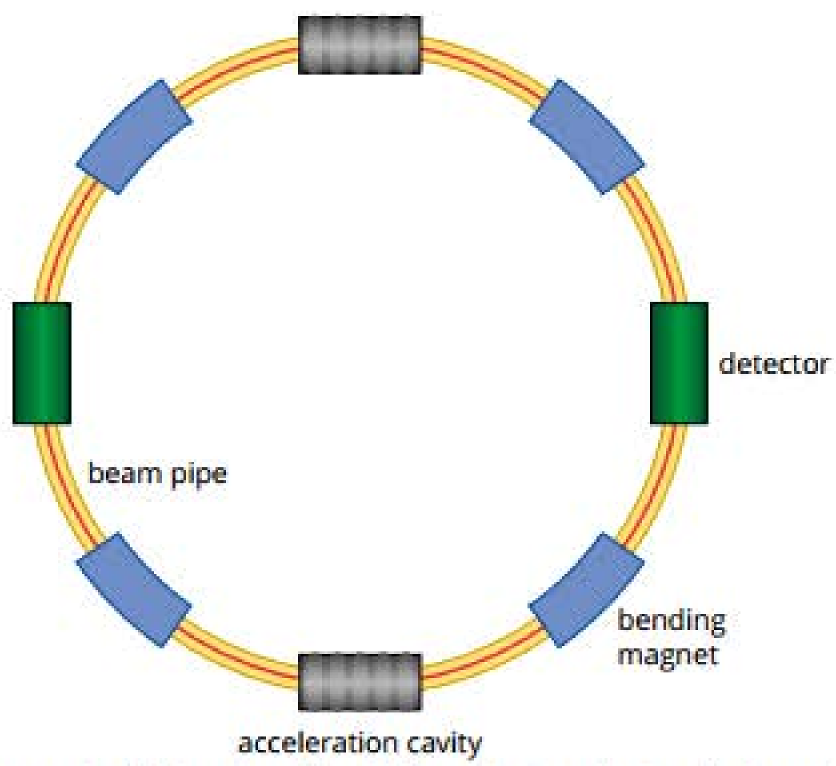
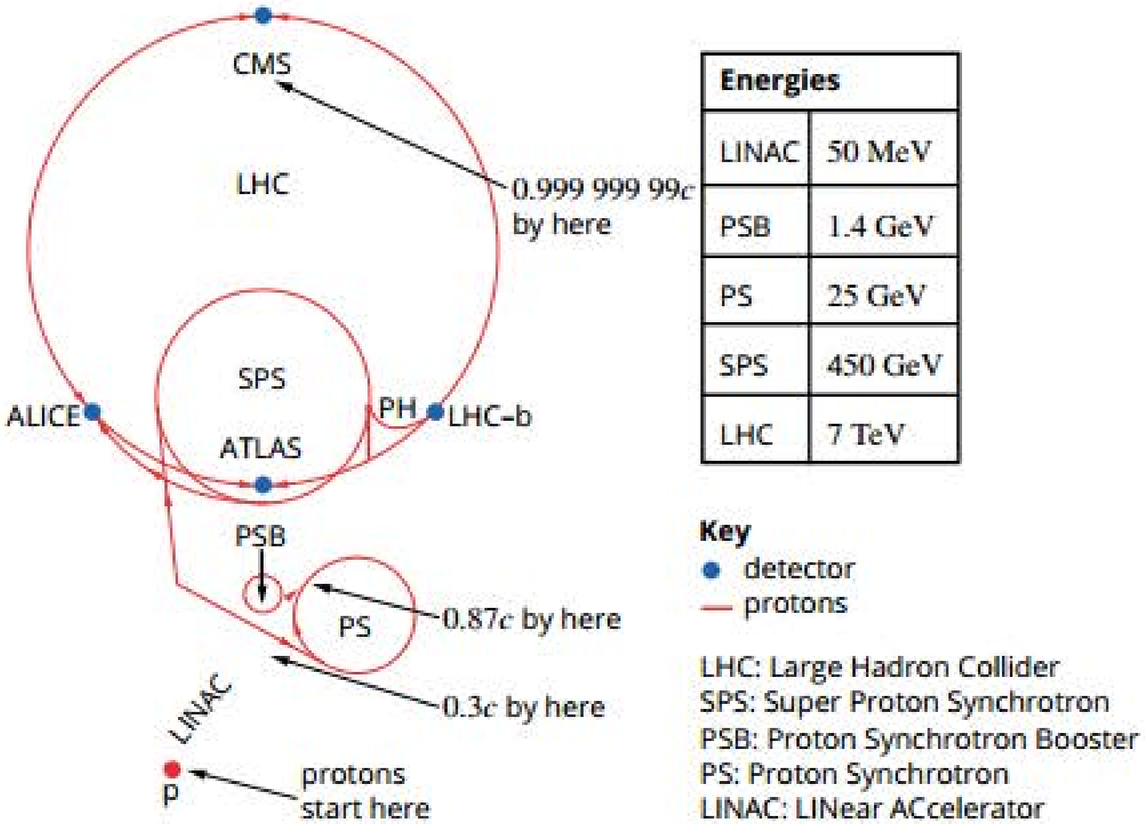


FIGURE 9.3.5 A very simplified version of the LHC synchrotron showing the beam pipes (red), the magnet system (blue), the acceleration cavities (grey) and the detectors (green).

The protons bound for the LHC are first put into motion in a linear accelerator. They then pass through three synchrotrons before reaching the LHC itself. An overview of all the accelerators that are the parts of the LHC is shown in Figure 9.3.6. The velocities of particles at several stages and the energies involved when colliding protons are also shown.

Bunches of protons are injected into the main synchrotron after being accelerated to about 10% of their final energy. These bunches are travelling at very close to the speed of light at the time of injection, but, even so, a very large amount of energy is required to give them their final velocity of $0.999\,999\,99c$.

As of 2017, the LHC accelerates protons to an energy of 7 TeV, which results in a total collision energy of 14 TeV. These high energies have already resulted in several new particle discoveries, as the LHC can explore collision energies not previously available.



Energies	
LINAC	50 MeV
PSB	1.4 GeV
PS	25 GeV
SPS	450 GeV
LHC	7 TeV

FIGURE 9.3.6 The CERN accelerator complex, including the LHC and its four detectors.

ENERGIES OF PARTICLES IN ACCELERATORS

Modern particle physics involves accelerating particles to velocities very close to the speed of light. This means that any understanding of how much energy a particle would have must be based on special relativity. The energy of a particle at high speeds is a combination of its rest energy and kinetic energy.

Einstein's equation, $E = m_0c^2$, allows us to calculate the energy contained within a particle with a rest mass m_0. When Einstein developed this equation, he showed that the classical equation for kinetic energy, $E_k = \frac{1}{2}mv^2$, does not hold at relativistic speeds. The total energy of a particle is then its rest energy plus its relativistic kinetic energy.

A simple method for interpreting this increase in a particle's energy at relativistic speeds is as an increase of the mass of the particle. Therefore, the mass of a relativistic particle is given by:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

where m_0 is the rest mass and m is the larger, **relativistic mass** (also called dilated mass). The graph shown in Figure 9.3.8 illustrates the fact that, as a particle approaches the speed of light, the mass of the particle increases significantly.

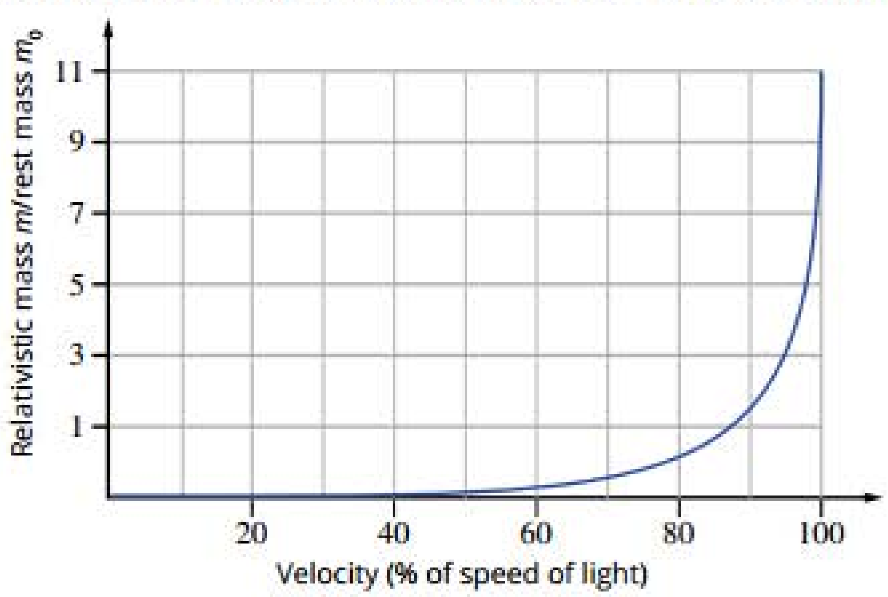


FIGURE 9.3.8 The graph shows that the increase in the mass of a particle becomes very significant once the particle exceeds about 40% of the speed of light.

PHYSICSFILE

A little energy in a small space is a big deal

Two protons colliding inside one of the LHC's detectors have a combined energy of 14 TeV. A simulation of this interaction is shown in Figure 9.3.7.

You may be surprised to know that when you clap your hands the collision probably involves more energy than 14 TeV. A collision of 1 TeV is about the energy of motion of a flying mosquito.

The thing that makes the collisions at the LHC so impressive is that the energy of 14 TeV is squeezed into a space about a million times smaller than a mosquito.

Energy density is the important thing. Clap your hands and the worst you might do is hurt your hands or your ears. Collide two protons at 14 TeV and you might even make a tiny black hole. That's impressive.

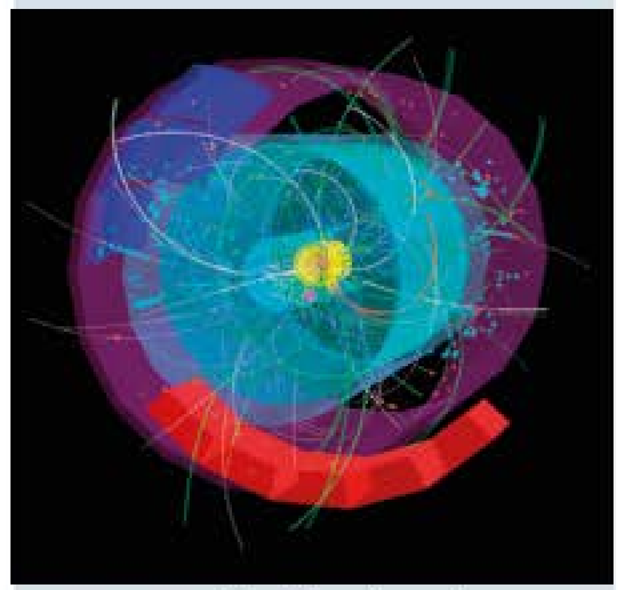


FIGURE 9.3.7 Simulation of a proton–proton collision at 14 TeV centre of mass energy at CERN LHC.

From the frame of reference of a particle accelerator, the mass of the particles in the accelerator increases as the speed approaches the speed of light. According to Newton's second law, a larger mass requires a larger force to achieve the same acceleration. This has the practical implication that as particles approach the speed of light, they require an increasing amount of energy to accelerate them closer to the speed of light.

The total energy of a relativistic particle would be equal to $E = mc^2$, where m replaces m_0. Substituting the relativistic mass equation into the mass–energy equation gives an expression for the total energy of a particle:

$$E = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

Using this equation gives an energy value in J if the mass is in kg and velocities in $\mathrm{m\,s^{-1}}$.

Worked example 9.3.1

ENERGIES OF PARTICLES

An electron with a rest mass of 9.109×10^{-31} kg is accelerated to 0.95c. Calculate the total energy of the particle in J and MeV. Give your answer to two significant figures.	
Thinking	**Working**
Identify the equation you need to use, the information you know already and the unknown you need to calculate.	The equation that is needed is: $E = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}}$
Identify relevant known data.	$m_0 = 9.109 \times 10^{-31}$ kg $v = 0.95c$ $c = 3.0 \times 10^8\ \mathrm{m\,s^{-1}}$ $E = ?$
Convert any values to SI units if needed and substitute into the equation.	$E = \frac{9.109 \times 10^{-31} \times (3.0 \times 10^8)^2}{\sqrt{1-\frac{0.95^2c^2}{c^2}}}$ $= 2.6 \times 10^{-13}$ J
Convert J to MeV.	$1\ \mathrm{eV} = 1.6 \times 10^{-19}$ J $1\ \mathrm{MeV} = 1.6 \times 10^{-13}$ J $\therefore E\ (\mathrm{MeV}) = \frac{2.6 \times 10^{-13}}{1.6 \times 10^{-13}}$ $= 1.625$ MeV $= 1.6$ MeV

Check your answer. This amount of energy should be greater than the rest energy of the electron.	$E = m_0c^2$ $= 9.109 \times 10^{-31} \times (3.0 \times 10^8)^2$ $= 8.198 \times 10^{-14}$ J 1 eV = 1.6×10^{-19} J 1 MeV = 1.6×10^{-13} J $\therefore E\,(\text{MeV}) = \dfrac{8.198 \times 10^{-14}}{1.6 \times 10^{-13}}$ $= 0.51$ MeV The answer is 3.2 times this value, so significant dilation of mass has occurred at 0.95c. The answer seems reasonable.

Worked example: Try yourself 9.3.1

ENERGIES OF PARTICLES

A proton with a rest mass of 1.672×10^{-27} kg is accelerated to 0.95c. Calculate the total energy of the particle in J and MeV. Give your answer to two significant figures.

BENDING A PROTON BEAM

The two beams of proton bunches travel in the beam pipes of the LHC within a high vacuum, like the vacuum of space. The beams are directed around the accelerator by many strong magnets that deflect the path of the protons to maintain a radius of 4.3 km (Figure 9.3.9).

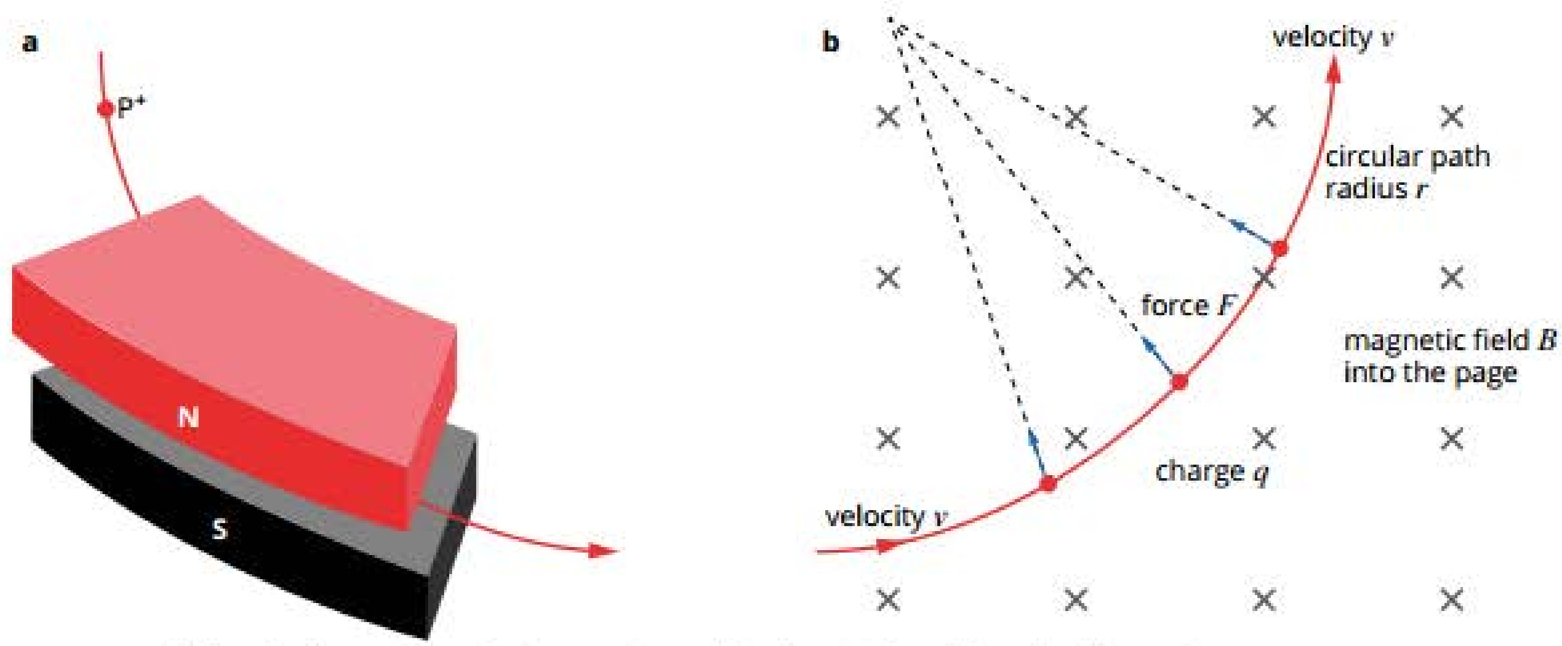


FIGURE 9.3.9 (a) A pair of very strong electromagnets provide a force to bend the path of the protons around the accelerator. (b) The path of the particle from above the accelerator and the centripetal force acting towards the centre of the circular path due to the magnetic field.

The Lorentz force, F (in newtons), on the particles with charge q (in coulombs) due to their motion through a magnetic field B (in teslas) at a velocity v (in m s^{-1}) is given by:

$$F = qvB$$

The direction of this force can be predicted by the right-hand palm rule. This force causes a circular motion that can be described by the centripetal force equation:

$$F = ma = \frac{mv^2}{r}$$

where r is the radius of curvature of the circular path of the particle (in m), a is the centripetal acceleration (in m s^{-2}) due to the change in the direction of the velocity of the particle, and m is the mass (in kg).

Equating the two equations above gives:

$$F = qvB = \frac{mv^2}{r}$$

This equation can be used to calculate the variables involved at non-relativistic speeds, where m is the rest mass of the particle m_0. At relativistic speeds the equation seen previously

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

indicates that the mass of the particle would increase when viewed from the rest frame of the accelerator. By inspecting the equation above, you can see that as v increases m also increases. The previous relationship then shows that the strength of the magnetic field between the bending magnets must increase as velocity increases. This exerts a larger force to maintain the same radius within the beam pipe.

Worked example 9.3.2

BEAM BENDING

A proton travels at 1.0×10^8 m s^{-1} in a direction perpendicular to a uniform magnetic field of strength 0.5 T. The rest mass of a proton is $m_0 = 1.672 \times 10^{-27}$ kg and the charge on the proton is $+1.6 \times 10^{-19}$ C. Assume that the increase in mass of the particle at this velocity is negligible. Calculate the radius of the circular path of the proton. Give your answer to one significant figure.

Thinking	Working
Rearrange the equation that represents the relationship between the Lorentz force and the centripetal force to make r the subject.	$F = qvB = \frac{mv^2}{r}$ $r = \frac{mv}{qB}$
Identify relevant known data.	$m_0 = 1.672 \times 10^{-27}$ kg $v = 1.0 \times 10^8$ m s^{-1} $B = 0.5$ T $q = 1.6 \times 10^{-19}$ C
Substitute the values into the equation.	$r = \frac{1.672 \times 10^{-27} \times 1 \times 10^8}{1.6 \times 10^{-19} \times 0.5}$ $= 2.09$ $= 2$ m

Worked example: Try yourself 9.3.2

BEAM BENDING

A proton travels at 0.9×10^8 m s^{-1} in a direction perpendicular to a uniform magnetic field of strength 0.7 T. The rest mass of a proton is $m_0 = 1.672 \times 10^{-27}$ kg and the charge on the proton is $+1.6 \times 10^{-19}$ C. Assume that the increase in mass of the particle at this velocity is negligible. Calculate the radius of the circular path of the proton. Give your answer to one significant figure.

Acceleration of protons in the LHC

As they travel around the ring, the protons in each beam pipe are accelerated and bunched using eight radio frequency (RF) cavities. Each bunch contains billions of protons and there are more than 2000 bunches circulating in the LHC when it is in operation. Every time a bunch passes through an RF cavity it is accelerated by the electric field (Figure 9.3.10). This electric field oscillates at the correct frequency to impose an accelerating force on each bunch when it arrives after each circuit of the accelerator.

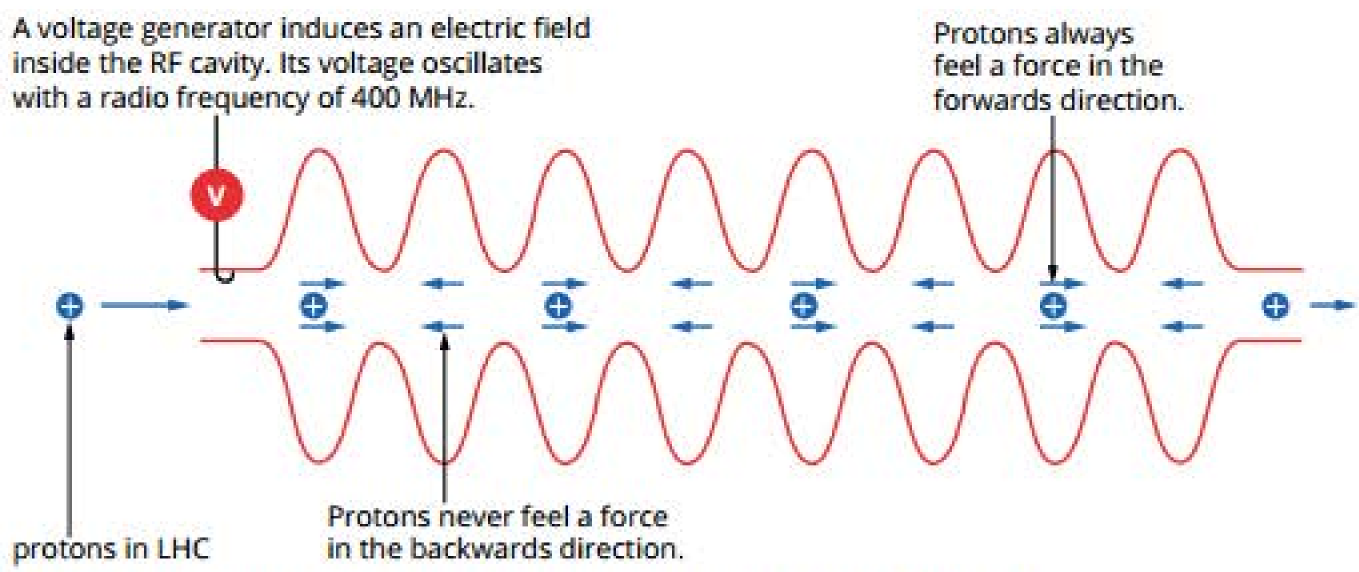


FIGURE 9.3.10 The basic concepts behind the operation of the RF cavities in the LHC.

The final result in the LHC is two beams circulating in opposite directions, each containing about 50×10^9 protons in 2800 discrete bunches (one every 7 m). After 20 minutes, they reach their final energy of 7 TeV, while doing 11 245 circuits of the LHC ring per second. In those 20 minutes, the protons cover a distance further than from the Earth to the Sun and back.

Once the protons are at these enormous energies they need to be brought together in a collision. Their energies are converted into new particles and these events are detected and measured using particle detectors such as ATLAS and CMS at the LHC.

SYNCHROTRON COLLIDERS AND THE DISCOVERY OF THE W AND Z BOSONS

The discovery of W^+, W^- and Z gauge bosons is considered one of the greatest achievements in the history of science. Their prediction in the 1970s, and subsequent discovery in the 1980s, confirmed the main concepts of the Standard Model. It was now accepted that there are three fundamental forces (excluding gravity) and that these forces are mediated by force-carrying particles.

Specific predictions were made of the existence of heavy bosons (80.4 GeV and 91.2 GeV) for the weak nuclear force, and that weak interactions sometimes occurred without any change in electric charge. The theory on which these predictions were made implies that the electromagnetic force and weak force are two manifestations of the **electroweak force**. These two forces appear different because the force-carrying boson for the weak nuclear force has mass, but the photon does not. The large mass of the boson for the weak nuclear force also explains its short range compared to the electromagnetic force.

What came to be known as the electroweak theory predicts that at high energies the electromagnetic force and the weak nuclear force are unified as the electroweak force. This unification implies that at the high energies that existed in the early universe, the forces that were acting between particles were very different.

It was the development of higher-energy accelerators that made confirmation of these predictions possible. For the first time, particles could be collided with sufficient energy to produce heavy bosons. This was done by colliding protons and antiprotons in the Super Proton Synchrotron (SPS) at CERN (Figure 9.3.11). The SPS first accelerated protons in 1976, achieving energies of 400 GeV. Since 1981, the machine has also operated as a proton–antiproton collider, with the same bending magnets (red) and focusing magnets (blue) guiding protons and antiprotons around the ring in opposite directions. The accelerator forms an underground ring 7 km in circumference.

FIGURE 9.3.11 The tunnel of the Super Proton Synchrotron (SPS) particle accelerator at CERN, the European particle physics laboratory outside Geneva.

PHYSICSFILE

Confirmation of existence of W^+, W^- and Z bosons

In a series of experiments, the existence of W^+, W^- and Z bosons was finally confirmed in 1983 and 1984. The 1984 Nobel Prize in Physics was awarded jointly to Carlo Rubbia and Simon van der Meer for 'their decisive contributions to the large project, which led to the discovery of the field particles W and Z, communicators of weak interaction'. These experiments involved detecting the products from particle collisions using huge particle detectors. These detectors measured the trajectories and energies of the particles produced. Careful analysis of the data and comparison to theoretical predictions allowed the identification of W and Z bosons and for several components of the Standard Model to be verified. A computer-graphic display, such as that in Figure 9.3.12, illustrates a Z particle decay.

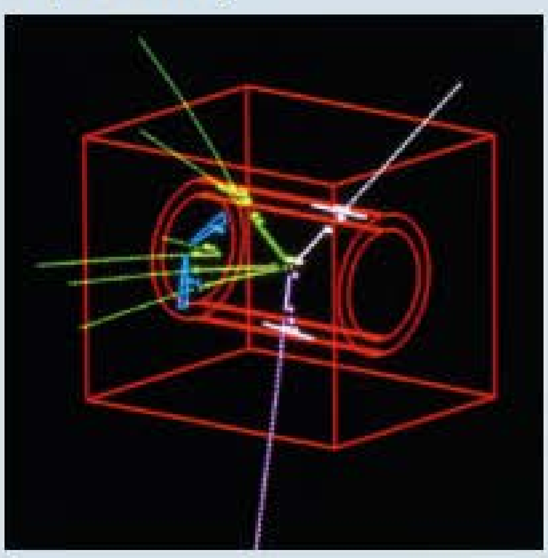


FIGURE 9.3.12 A computer-graphic display of a Z particle decay event detected by the UA1 detector is shown. The Z particle lived too briefly to be visible itself. It decayed into two electrons (long white and lilac lines) and some other particles (green) that were detected.

ANALYSING PARTICLE EVENTS

Many particle physicists are still searching for theoretically predicted, but currently undiscovered particles. In modern particle physics, this typically involves colliding particles moving in opposite directions at extremely high energies. Measurements are made of the products of these collisions and these measurements are analysed, looking for evidence (or not) that the predicted particle was produced.

The process described above may sound simple but it is very challenging. The particle that physicists are searching for may only exist for a tiny fraction of a second. This is not enough time for the particle itself to be detected after the collision. Physicists must instead detect the particles that have been formed in the decay of the particle they are interested in, and then work backwards.

Particle physicists observe the collision event and make careful measurements. They then use the known rules governing all interactions of the particles involved (such as the conservation laws discussed in Section 9.2) to work backwards and determine exactly what happened. The energies and paths of some of the particles observed may indicate that they were produced after the initial collision by the decay of other short-lived particles.

The LHC's ATLAS detector

When the LHC began operation, its first job was to rediscover the particles of the Standard Model. This allowed physicists to calibrate their detectors and gain a precise knowledge of what is already known so they can clearly separate it from something new.

The beams of the LHC create up to one billion collisions per second and can produce hundreds of particles with each collision. The components of the ATLAS detector identify and measure the multitude of particles that are produced in a collision. The data produced by this detector are generated for a team of about 2900 physicists.

The volume of raw data generated by ATLAS, about 1000 TB every second, cannot all be stored. It is analysed in real-time by customised computer programs using patterns predicted by mathematical models to indicate, for example, the formation of hypothetical particles. Events of interest are saved for further analysis by particle physicists; the rest of the data generated is discarded.

Layers of ATLAS and their functions

The inner tracking chamber of the ATLAS detector (Figure 9.3.13) detects the curved path taken by charged particles in the strong magnetic field of the solenoid magnet. This allows the momentum and charge of these particles to be determined. In a similar way, the muon spectrometer identifies and measures the momentum of muons. These muons require a separate tracking system, as they are weakly interacting.

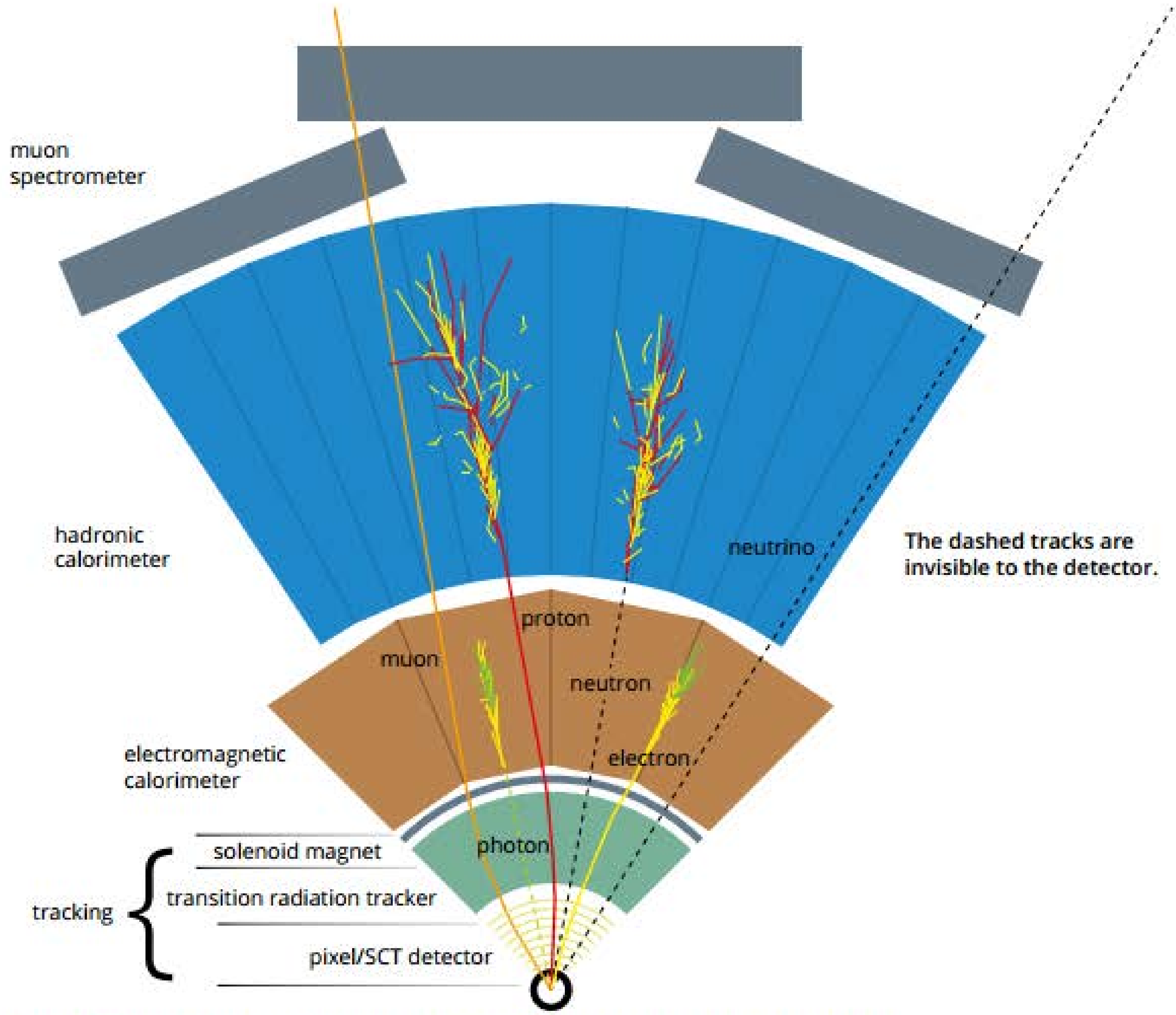


FIGURE 9.3.13 This is a cross-section of the detector, showing its layers and examples of where different particles would be detected after an event.

The calorimeters measure the energies carried by photons, electrons, positrons and hadrons. The interactions inside both types of calorimeters cause cascades or 'showers' of new particles. With each interaction or collision, the momentum and energy is shared between more particles, until eventually all are slowed and stopped. To accurately measure the total energy requires the detector to stop nearly all particles, and thus explains the huge overall size of the detectors used.

Particle identification detectors are also used to identify the type of particle by various techniques. One commonly used method to detect the nature of charged particles is by analysing Cherenkov radiation, which is produced when charged particles travel faster than the normal speed of light in that medium. When a charged particle travels at a speed faster than that of light in given medium, it emits Cherenkov radiation at an angle that depends on its velocity. The particle's velocity can be calculated from this angle. Velocity can then be combined with a measure of the particle's momentum to determine its mass, and therefore its identity.

9.3 Review

SUMMARY

- Particle accelerators were invented to produce collisions with enough energy to produce new particles heavier than those being collided.
- The Stanford Linear Accelerator (SLAC) was used to collide accelerated electrons with a stationary target containing protons and neutrons. In the 1970s, these experiments resulted in the first evidence for quarks inside protons and neutrons from observations of the way the electrons were scattered in these collisions. This evidence gave support to the quark model of hadrons, which is now part of the Standard Model.
- A linear accelerator accelerates particles using an alternating electric field. The field between successive drift tubes accelerates the particles in a straight line. The drift tubes are successively longer to ensure the alternating field is the correct polarity, as the particles take less time to move from one drift tube to the next.
- As a particle is accelerated to near the speed of light, it takes increasingly more energy to accelerate the particle. This can be interpreted by a stationary observer as an increase in mass, and can be calculated using equations of special relativity. The total energy can then be determined using Einstein's equation for mass–energy, where the mass is the relativistic mass:

$$E = \frac{m_0 c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

- Modern particle experiments require higher energies that are difficult to achieve with linear accelerators. Synchrotron colliders are used, which collide two bunches of particles travelling in opposite directions. Synchrotrons are composed of a circular beam pipe, bending magnets to keep the particles moving in a circular path, RF cavities to accelerate the particles, and particle detectors to detect the results of collisions.
- The variables involved in maintaining the circular paths of particles within a synchrotron can be understood by applying the relationship:

$$F = qvB = \frac{mv^2}{r}$$

- The Super Proton Synchrotron at CERN confirmed the existence of the W^+, W^- and Z bosons in the 1980s. These were predicted by the electroweak theory, a component of the Standard Model. The discovery of these particles as predicted supports the Standard Model.

KEY QUESTIONS

1 Which of the following would allow a particle physicist to study the finest detail within a hadron? Explain the reason for your choice.

A accelerate a lepton with less mass than the electron
B increase the momentum of the accelerated particles
C decrease the velocity of the accelerated particles
D increase the wavelength of the accelerated particles

2 Describe the function of the parts of the synchrotron listed below.

a bending magnets
b beam pipe
c accelerating cavities
d detectors

3 The data from a particle physics experiment revealed that a particle had been produced that was initially travelling at 0.999c. When it was travelling at this velocity, it had a total energy of 3.78×10^{-10} J. Calculate the rest mass of this particle.

4 A proton travels at 0.999c in a direction perpendicular to a uniform magnetic field of 7.0T. The rest mass of a proton is $m_0 = 1.672 \times 10^{-27}$ kg and the charge on the proton is 1.6×10^{-19} C. Calculate the radius of the circular path of the electron. Give your answer to one significant figure.

5 Which of the following is thought to be a reason the weak nuclear and electromagnetic forces appear different? Explain the reason for your choice.

A The gauge bosons responsible for the weak nuclear force have no mass and those responsible for the electromagnetic force have large masses.
B The gauge bosons responsible for the weak nuclear force have large masses and those responsible for the electromagnetic force have no mass.
C The electromagnetic and weak nuclear forces were unified as the electroweak force in the early universe.
D The electromagnetic and weak nuclear forces are unified as the electroweak force in the Standard Model.

6 What was the main technological development that was responsible for the ability to discover the W and Z bosons?

7 Which of the following best describes the main physical principles used to analyse the results of particle collisions?

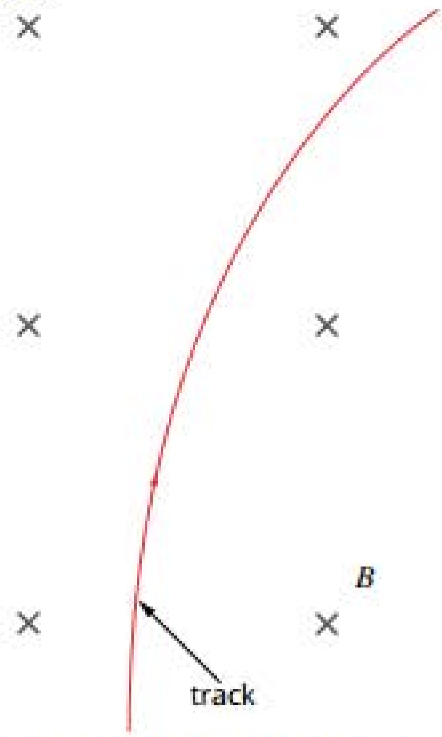
A the conservation of charge

B the conservation of lepton number

C the conservation of linear momentum

D all the conservation laws that govern particle interactions, including those above

8 A fundamental particle is detected by a particle detector and its track in a strong magnetic field is shown below.



The strength of the field was 5.0 T, the particle was moving at a velocity of 0.4*c* and the radius of its circular path was 10 m. Determine the charge and mass of the particle. Give your answers to two significant figures.

9 A proton ($m_0 = 1.672 \times 10^{-27}$ kg) is injected into an accelerator at 0.99*c*. How many times would you have to increase its velocity to double the total energy of the proton? Give your answer to two significant figures.

10 At non-relativistic speeds, if you double the velocity of the particles within the bending magnets, what is the radius of the particle beam?

9.4 Expansion of the universe

Perhaps the biggest question of all is 'Where do we come from?'. Humans have always asked and often tried to answer this question, but today our understanding is based on evidence, in other words, on science.

The understanding of the history of the universe comes from the Big Bang theory. This theory is built on the theories of general relativity and the Standard Model of particle physics, and also relies on other theories such as the theory of thermodynamics. It is supported by a large body of evidence, including the observations of Hubble's law, the cosmic microwave background radiation and the abundance of different elements in the universe.

You live in an expanding universe in which energy from the Big Bang was converted into matter, where the four forces were once one, and those forces, once distinct, built the matter and structure in the universe we see today.

DEVELOPING A PICTURE OF THE UNIVERSE

For most of human history the Earth was considered the centre of the universe, and the Sun and stars all just in the Earth's 'heavens'. Galileo proposed that the Sun was at the centre, but it was Newton who realised that the Sun was in fact a star, like the thousands of others in the night sky, and our planet one of six known at that time revolving around the Sun. Newton deduced that if the other stars were like our Sun they must be something like a million times farther away—a distance far too large to measure by any means known at that time.

With the development of better telescopes in the 19th century, the tiny apparent **parallax movement** (which results from the Earth's motion around the Sun) was detected for a few stars, and the first real measurements of distances made. It turned out that while the closest stars were about 300 000 times as far as the Sun, the vast majority of them were well over a million times as far.

There are several hundred billion stars in our galaxy, but there are about a hundred billion galaxies. Figure 9.4.1, an image that was taken deep in space by the Hubble Space Telescope, shows just some of these many galaxies. But where did all these stars and galaxies come from? Have they always existed? Are they changing? Are they made of the same sort of stuff that our star and planet are made of?

FIGURE 9.4.1 The night sky looks dark, but the Hubble Space Telescope reveals that the deeper we look the more galaxies we see.

Distances in space

By the end of the 19th century larger telescopes had revealed many new mysteries. Beautiful **nebulae** with spiral structures had been observed, but it was unclear whether these were relatively close objects or something much more distant.

It was Edwin Hubble who resolved the mystery. After careful examination of his photographs of spiral nebulae, he discovered that some contained stars called Cepheid variables. These were known to be very bright stars that slowly changed in brightness over a period ranging from a day or two to several weeks. It had been found that their **intrinsic brightness** (actual brightness) was related to the period of this variation (Figure 9.4.2).

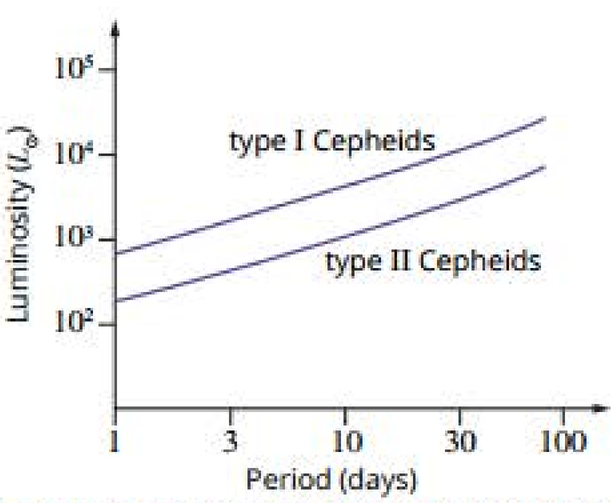


FIGURE 9.4.2 The period–luminosity relationship for Cepheid variables. There are two types that can be distinguished by the metal content of their spectra. $L_\odot$ is the luminosity of the Sun. Fortunately, Cepheids are very much brighter than the Sun, enabling them to be seen even in distant galaxies.

Given the intrinsic brightness, it is possible to work out the distance by comparing the intrinsic brightness to the apparent brightness. If a star that is intrinsically bright appears to be faint, it is clearly a long way away.

In this way, Hubble could estimate the distance to some of these spiral nebulae. His results clearly indicated that these nebulae were indeed a very long way away, in fact they were well outside our Milky Way galaxy. The closest, the Andromeda galaxy, is around 750 kiloparsecs (kpc) away. Compare that with the diameter of the Milky Way, which is about 50 kpc.

PHYSICSFILE

Units of distance in space

It is hard to picture the vast range of distances in space. Astronomers use two measurements of interstellar distances in particular.

The 'astronomical unit', or AU, is the distance from the Earth to the Sun. It is equivalent to 1.5×10^8 km. It is useful when discussing the distances within the solar system or to nearer stars. For example, the distance to Pluto is about 40 AU, while that to the nearest star (Proxima Centauri) is about 270 000 AU.

For larger distances the **parsec** is used. This is the distance you would have to be from the Earth so that the radius of the Earth's orbit around the Sun would make an angle of one second of arc (shown in Figure 9.4.3). It is equivalent to 206 265 AU. The kiloparsec (kpc, 10^3 pc) and megaparsec (Mpc, 10^6 pc) are also used for galactic distances.

Modern telescopes have been used to measure the distances of galaxies containing Cepheids out to about 30 Mpc. Using newer techniques, astronomers have now found galaxies out to about 1000 Mpc.

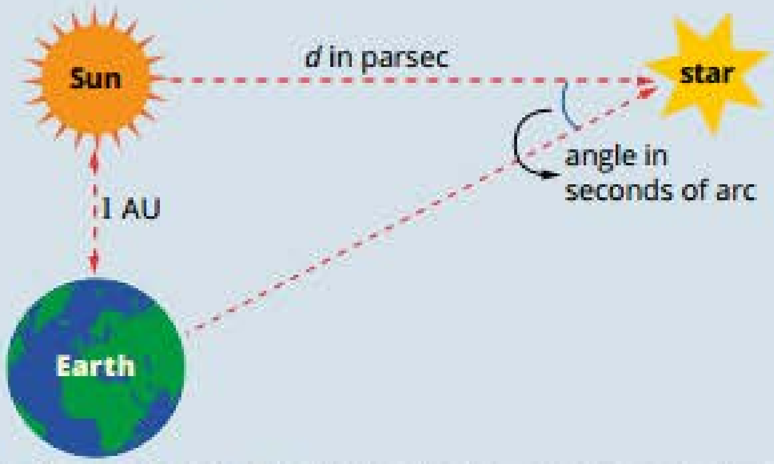


FIGURE 9.4.3 This diagram illustrates the relationship between astronomical units (AU) and parsecs (pc).

REDSHIFT AND HUBBLE'S LAW

The spectrum of the light from a star or a galaxy reveals a lot about the makeup of the object. When astronomers looked at the spectra of spiral nebulae, they found something very strange. Although they contained the familiar sets of lines that revealed the presence of various elements, all the lines seemed to be shifted towards the red end of the spectrum (Figure 9.4.4). Furthermore, the more distant the galaxy, the greater this so-called **redshift**.

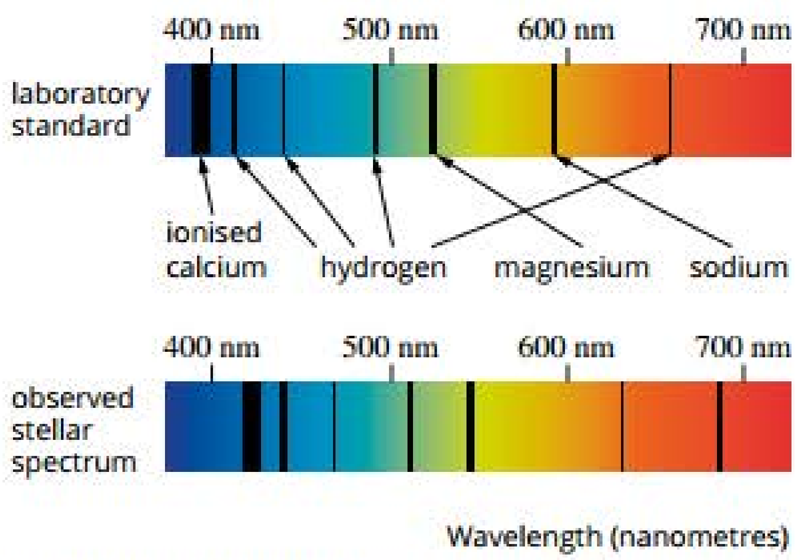


FIGURE 9.4.4 The spectrum of a star in our galaxy is redshifted, indicating it is moving away from the Earth in its motion around the galaxy.

The obvious interpretation of the redshift was that it was a due to the motion of the galaxies away from us. In fact, Hubble found that the speed of recession, v, was directly proportional to the distance away, d (Figure 9.4.5). This relationship became known as **Hubble's law**.

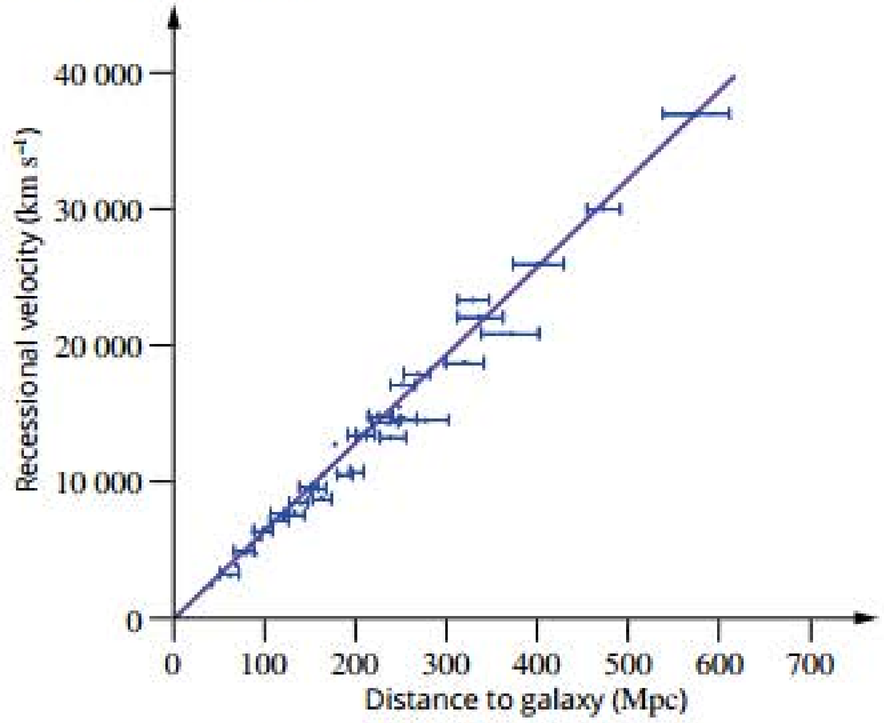


FIGURE 9.4.5 Hubble discovered that the galaxies were receding from us at velocities proportional to their distance. The slope of the line is known as the Hubble constant, H_0.

Hubble's law and can be expressed as:

$v = H_0 d$

H_0 is the Hubble constant.

The slope of the graph in Figure 9.4.5 represents the **Hubble constant**. It shows that the Hubble constant is about 70 (km s^{-1}) Mpc^{-1}.

So, for example, a 'typical' galaxy at a distance of 100 Mpc would be rushing away from us at 7000 km s^{-1}. (That's travelling from Perth to Sydney in approximately half a second.)

EXPANSION OF THE UNIVERSE

If the galaxies are all rushing away from us, then does that mean that the Earth is in the centre of the universe after all? To the astronomers of the early 20th century this seemed highly unlikely. The universe seemed to be basically uniform and infinite, and therefore without a centre. Also, the idea that there could be some sort of 'edge' to the universe didn't make sense. After all, if there was an edge, what was beyond it?

Even Einstein, when he produced his general theory of relativity in 1915, was so dismayed to find that it did not allow for an infinite universe that he amended it by adding what he called a 'cosmological constant' so that an infinite universe could exist. (Later he said this was his 'greatest blunder', because if he had believed his own equations, it would have been he who discovered the expanding universe.)

Hubble explains expansion

Hubble eventually proposed that because all the distant galaxies were receding from us, space itself was expanding rather than the galaxies rushing away from us. This is just what Einstein's equations had predicted, but even he had not wanted to believe such an outrageous idea.

Hubble suggested that at a very large distance from the Earth, space will be expanding faster than the speed of light and therefore the light from stars would never reach us. Although this distance is not the 'edge of the universe', it is a distance beyond which humans will never see—the 'edge of the visible universe'. This doesn't violate Einstein's special relativity because it is not that anything is travelling through space at faster than the speed of light.

PHYSICSFILE

Modelling the expanding universe

The idea that space itself is expanding is complicated. A helpful analogy is to imagine an ant on a balloon that is being blown up, as in Figure 9.4.6. This is like the universe expanding:

- The stars represent the galaxies in the universe.
- As the balloon expands, the ant would see all the paper stars moving away from it.
- Wherever the ant wanders on the balloon's surface it would see the same thing.
- The stars remain the same size rather than expanding also. It is space that is expanding, not the stars and galaxies in it.
- Unlike a conventional explosion, as space/the balloon expands it will appear uniform in whatever direction you look and from wherever you look. There is no 'centre' of the three-dimensional space of the universe.

The stars on the balloon were stuck on and not drawn on for a good reason. As the balloon expands, drawn stars would expand also, but the stick-on stars stay the same size. This is how astrophysicists picture the expansion of space. It is space that is expanding, not the stars and galaxies in it.

FIGURE 9.4.6 As the balloon is blown up the stars all move away from the ant. The more distant stars will move away faster than those that are closer. Remember that this is a two-dimensional analogy for what is really a three-dimensional universe.

EXTENSION

Cosmological redshift

It is worth noting here that although the redshift seen in distant galaxies was originally thought to be a Doppler shift—the result of motion away from us—it is more correctly interpreted as the direct consequence of expanding space, as shown in Figure 9.4.7. The wavelength of the radiation increased with space rather than in space. This type of redshift is more correctly called a 'cosmological redshift' and was predicted by Einstein's general theory of relativity.

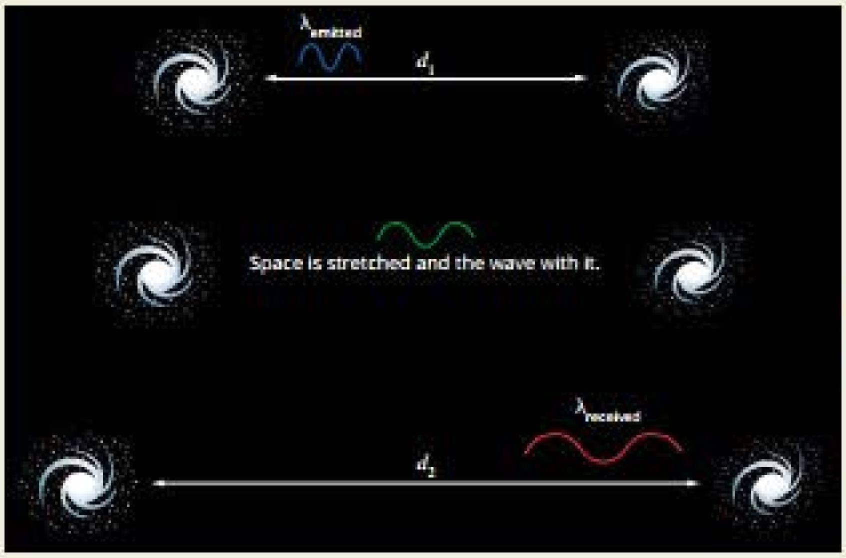


FIGURE 9.4.7 Light from distant galaxies is cosmologically redshifted during its journey through expanding space. The farther the light has travelled, the greater the redshift.

PHYSICS IN ACTION

The start of the universe

Throughout recorded history, scientists have always sought to find an answer to how the universe began, as well as if, when and how it will end. Astronomers have constructed hypotheses called cosmological models to try and find the answer. In the 1940s, the **Steady State theory** and the **Big Bang theory** were the two contenders.

Steady State theory

Logic tells us that if space is expanding, then at some time in the past it must have been just a dot. However, in the 1950s, astronomer Fred Hoyle didn't agree and he put forward what became known as the Steady State theory.

He suggested that the universe was:

- infinite—the 'outer' stars would never reach infinity and so could go on moving away from us forever
- expanding—matter is being created all the time at just the right rate to keep the density of the universe constant.

This was not really such an outrageous idea. Quantum mechanics had already suggested that matter was less 'permanent' than previously thought.

Big Bang theory

At the time, the alternative was Hubble's theory that galaxies outside the Milky Way were moving away from us. This therefore implied that at an instant in time, the universe and everything in it was contained in a single point. In other words, everything was created from nothing in one 'big bang'.

At the moment of this big bang, all of the matter and energy in our present universe was packed together in some infinitely small region, which rapidly expanded. Based on understandings from nuclear and particle physics, the Big Bang theory predicts that energy was converted into matter in the early universe. The universe rapidly expanded and cooled. Millions of years later this matter began to condense into galaxies as the universe continued to expand.

It was in fact Hoyle who first used the expression 'big bang' to describe a universe that started off from a tiny dot. At the time, he was intending to be derogatory, but the name stuck.

Steady State theory vs Big Bang theory

The astrophysicists of the day really were caught in a dilemma. Either way they had to accept the unacceptable—that matter just came into being out of nothing. It seemed impossible to resolve this dilemma. If Hoyle was right, then where should scientists look for the new matter being created? If the universe really started with a big bang, how could they possibly know? The observational evidence, an expanding universe, was consistent with both theories.

If matter was being created at a steady rate, it was not too difficult to calculate that about two or three atoms of hydrogen would need to be created every day in a volume about the size of a large sports arena to account for the observed density and expansion. That didn't seem all that impossible, but there was not much point in trying to look for it!

On the other hand, if the Big Bang idea was correct, what difference might it make to what you can see around you now? Not much, but the universe would have been denser at an earlier time. It seemed that only some means of looking back in time could resolve the problem.

FIGURE 9.4.8 In the 1950s, the British astronomer Sir Fred Hoyle proposed the Steady State theory in opposition to the Big Bang theory.

EXTENSION

Time, space, matter and ...

It is important to reiterate that the Big Bang is not to be seen as some sort of explosion from a small point in space. It was more an explosion of space. In fact, it was more correctly an explosion of spacetime.

Einstein had already showed that time and space were not the separate entities that one would normally think them to be. So, the Big Bang would not have occurred at some point in time any more than it would have occurred at a point in space. Time, space, and matter were all created together in the one 'big bang'.

Some physicists say that even the laws of physics may have been created in the Big Bang. Asking the question, 'What happened before the Big Bang?' is rather like asking, 'What is south of the South Pole?'. There is, in fact, no answer to the question because the question itself has no meaning—you simply can't go south from the South Pole.

COSMIC MICROWAVE BACKGROUND RADIATION

In the early 1960s, physicists Robert Dicke and P.J. Peebles at Princeton University calculated that in the Big Bang the temperature near the beginning would have had to have been hot enough to create a lot of helium from the fusion of hydrogen nuclei. Any hot object, including a hot new universe, gives off **blackbody radiation**. This would have been intense high-energy gamma radiation that would have filled all space in the early universe and would have been radiating out ever since.

If the radiation was still present it would have become very low energy, long-wavelength radiation. Not because somehow the energy had been lost, but because as space expanded so would the wavelength of the radiation. Unlike galaxies, which are held together by their gravity, the radiation would have expanded with space and so the wavelength would have increased by something like a hundred million times up to about a millimetre, corresponding to a temperature of just a few kelvins.

Now it happened that in 1964 two physicists, Arno Penzias and Robert Wilson, at the Bell Telephone Laboratories near Princeton, were trying out a new directional radio antenna designed to communicate with the new satellites that had just been placed in orbit. They found that wherever they pointed their antenna they seemed to pick up 'radio noise', microwave radiation with a wavelength of a few millimetres. After testing their apparatus for everything they could think of, they concluded that the radiation was actually coming from outer space.

Fortunately, they had heard of Dicke and Peebles' ideas and realised that the annoying noise they had found was probably this leftover radiation from the Big Bang. Needless to say, there was much celebrating, not to mention a great new area for research—and the death knell for Fred Hoyle's Steady State theory, which had no explanation for this radiation.

Although some of this **cosmic microwave background radiation** (CMBR) reaches Earth's surface, fortunately most is absorbed by the atmosphere, or it would create a lot more radio interference. In 1989 a special satellite called the Cosmic Background Explorer (COBE) was launched. It was designed to detect and measure radiation in the range of 1 μm to 1 cm from all directions in space. If it really was the leftover radiation from the Big Bang, it should come uniformly from all directions and appear to be at a temperature of only a few degrees above absolute zero.

As you can see in Figure 9.4.9, COBE showed that the spectrum of the CMBR corresponded exactly to the radiation expected from a so-called 'blackbody' at a temperature of just 2.725 K (i.e. just a little above absolute zero) and that it was uniform in all directions.

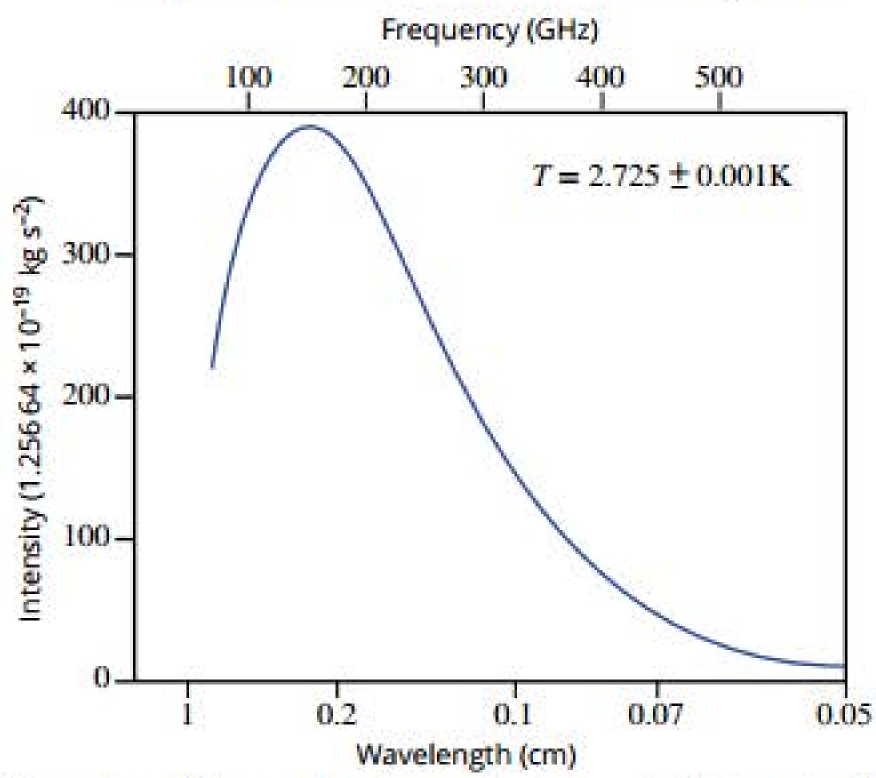


FIGURE 9.4.9 The spectrum of the cosmic microwave background radiation precisely fits the expected curve for blackbody radiation at 2.725 K. In fact, the experimentally measured points and error bars are so small they are within the thickness of the line. This was important support for the Big Bang theory.

Further evidence for the Big Bang

There was a difficulty with the Big Bang idea. If spacetime suddenly 'exploded' out of nothing, would it not expand uniformly in all directions, giving a completely bland universe? How could galaxies and stars form in this totally homogenous universe? So, was the early universe really so bland?

More recently another satellite, called WMAP, has mapped the radiation in more detail. It confirmed that there were very small variations in the temperature of the microwave background, as you can see in Figure 9.4.10. These were only of the order of thousandths of a degree, but they were very significant.

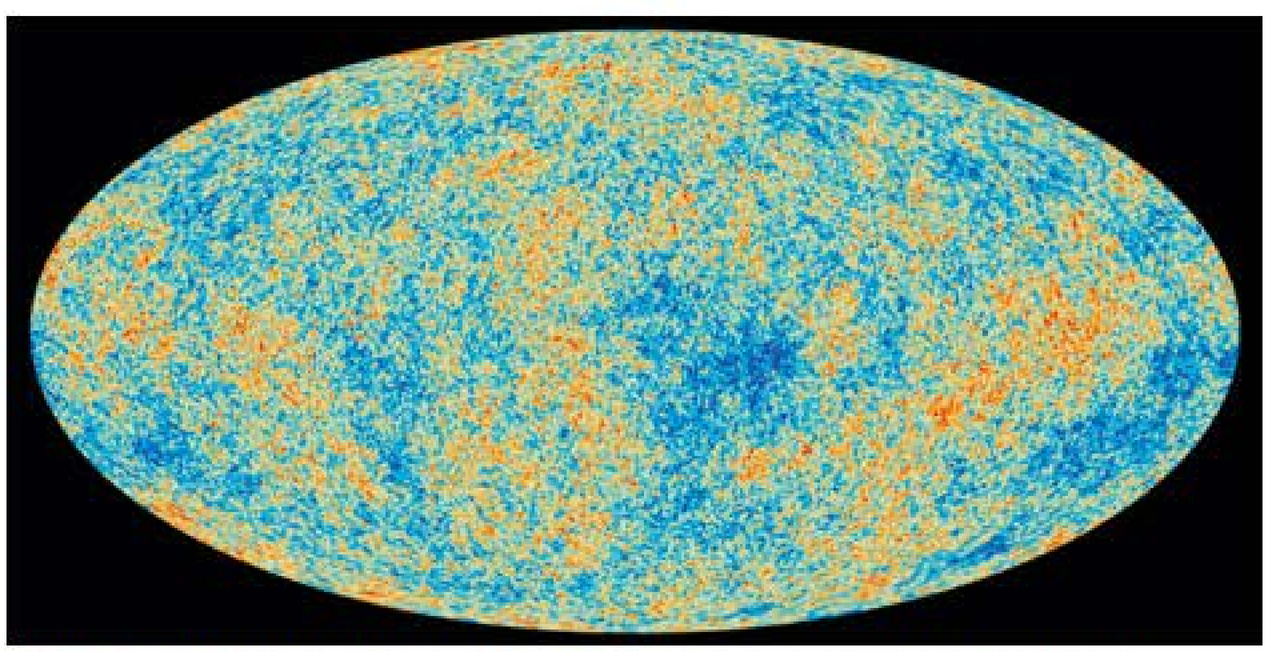


FIGURE 9.4.10 The WMAP satellite mapped the cosmic microwave background radiation and discovered tiny fluctuations in it, which probably marked the beginning of the formation of galaxies. (WMAP stands for Wilkinson Microwave Anisotropy Probe.)

For galaxies to form, there had to be some variation in the structure of the early universe. This would enable local clumps of matter to start to coalesce and form galaxies. Once a clump of matter starts to form, its gravity will accelerate the process. As the matter falls inwards, the temperature rises, creating the conditions for the formation of stars.

This early variation was just the sort of slight variation in the early universe that could have led to the formation of galaxies and stars. It would also have resulted in slight variations in the gamma radiation produced, and in the cooled-down form that we see as the CMBR. This is what WMAP was looking for and found, strongly supporting the Big Bang hypothesis.

THE AGE OF THE UNIVERSE

Given that the universe started with a big bang, physicists should be able to determine when it all started. In other words, how old is the universe? If Hubble's law is right, we simply need to run the rate of expansion backwards to see when all the galaxies were in the same place. This is not so difficult.

The Hubble constant, H_0, is the proportionality constant between the distance from us to a galaxy and the galaxy's recessional velocity ($v = H_0 d$). If the galaxy has been receding from us at a speed v and in that time has travelled a distance d, the time it has taken is $t = \frac{d}{v}$. Now the value of $\frac{d}{v}$ from Hubble's equation is $\frac{1}{H_0}$, so $t = \frac{1}{H_0}$. Notice that the distance, d, cancelled out in obtaining this expression, so all galaxies, no matter how far away they are now, must have started off at the same point in time—the beginning of the universe. This is when the Big Bang happened. The age of the universe is simply the reciprocal of Hubble's constant. Of course, you need to convert the units into units you recognise as follows.

The age of the universe is the reciprocal of the constant:

$$t = \frac{1}{H_0}$$

$$= \frac{1}{70} (\text{km s})^{-1}\,\text{Mpc}^{-1}$$

The expression for the constant includes two different units for distance, km and Mpc, so first you need to convert Mpc to km.

$$1\,\text{pc} = 3.086 \times 10^{16}\,\text{m}$$

$$\begin{aligned} 1\,\text{Mpc} &= 10^{6}\,\text{pc} \\ &= 10^{6} \times 3.09 \times 10^{16}\,\text{m} \\ &= 3.09 \times 10^{22}\,\text{m} \\ &= 3.09 \times 10^{19}\,\text{km} \end{aligned}$$

This means that 1 Mpc = 3.1×10^{19} km (to two significant figures) and so:

$$t = \frac{3.1 \times 10^{19}}{70}$$

$$= 4.4 \times 10^{17}\,\text{s}$$

This is equal to:

$$t = \frac{4.4 \times 10^{17}}{365 \times 24 \times 3600}$$

$$= 1.4 \times 10^{10}\ \text{years}$$

That is, the universe began 14 billion years ago.

This method of determining the age of the universe has assumed a constant rate of expansion and ignored the effects of gravity. Neither of these are totally valid assumptions, but astrophysicists now have more complex models that do take these effects into account. Curiously enough, these other effects tend to cancel each other out and the models still produce an age close to 14 billion years; in fact, the current best estimate is 13.7 billion years.

THE BIG BANG AND THE STANDARD MODEL

Progress in high-energy particle physics has allowed physicists to formulate the Standard Big Bang model. This combines the Standard Model of particle physics with the Big Bang theory. In a sense, particle accelerators act as time machines, taking us back to the conditions that prevailed when our universe was born. Perhaps one day scientists will construct an accelerator that will take us back to the beginning of time before matter formed, when the four forces of today's universe were one.

The Standard Model of particle physics has allowed us to piece together a picture of how matter formed and evolved. As the universe cooled after the Big Bang, fundamental particles and then composite particles formed. The strong nuclear force and the electromagnetic force were then able to create nuclei and atoms respectively. Gravity then constructed the universe as we know it by pulling matter together into structures such as galaxies made of stars and planets such as Earth.

Evolution of forces

Physicists believe that during the first 10^{-43} s, the four fundamental forces known today (gravity, weak nuclear force, strong nuclear force, and electromagnetic force) were unified as one single force. The fundamental particles known today are thought to have been created when the four forces separated. It is hoped that particle accelerators such as the LHC may soon provide evidence for the unification of forces. Figure 9.4.11 shows the strength of the four forces at different energies and at what energies each force was unified with others.

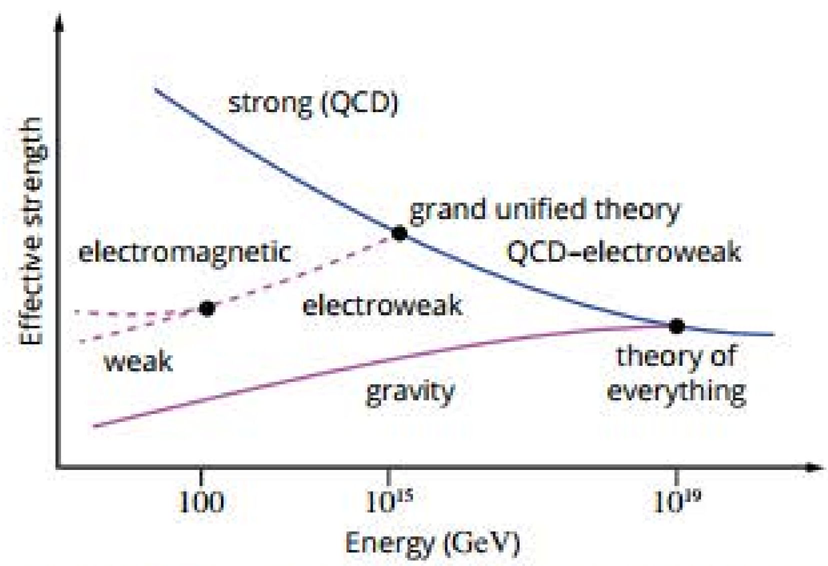


FIGURE 9.4.11 The graph shows the energy at which the four forces split from one another and their relative strengths. It also shows the theories of the Standard Model of particle physics that describe them when unified. We currently do not have a grand unified theory, or theory of everything. (QCD: quantum chromodynamics)

Although no complete physical description of this earliest period exists yet, it is thought that all that existed was a huge amount of energy and the size of the universe was less than the size of a proton. The temperature was 10^{32} K. At 10^{-43} s, gravity separated from the other forces (weak nuclear, strong nuclear and electromagnetic) but those three forces were still unified.

After 10^{-35} s, the strong nuclear force separated from the electroweak force and the energy released caused a period of time called inflation.

Inflation

In the initial stages of the Big Bang, there must have been an incredible period of inflation. This period only lasted about 10^{-24} s, but during that time the size of the universe expanded to about 10^{50} times its original size. It was now about the size of a basketball—it had started very small. Inflation was what prevented a very rapid collapse of the initial universe back into a black hole; but, had it lasted any longer, the expansion would have been so great that atoms would have never formed from the very thin particle zoo.

Matter and antimatter

To understand where the matter in the early universe came from, you need to be aware of a very important principle of modern physics—the Heisenberg uncertainty principle. Briefly, it says that there is a fundamental uncertainty in the position and momentum of any object.

It can also be interpreted as an uncertainty between mass and time. So, in any very short time there is a fundamental uncertainty about mass; the shorter the time, the greater the uncertainty. The real meaning of this is rather astonishing. It means that in a very short time, mass can come in and out of existence from nothing.

Furthermore, when particles are created, they always come in pairs. For every 'matter' particle there is an 'antimatter particle' (as discussed in Section 9.1). But because they are so close when they pop into existence, the matter and antimatter immediately **annihilate** (Figure 9.4.12). So, matter (and antimatter) are actually popping in and out of existence everywhere all the time. It's just that we can't see it because it doesn't last.

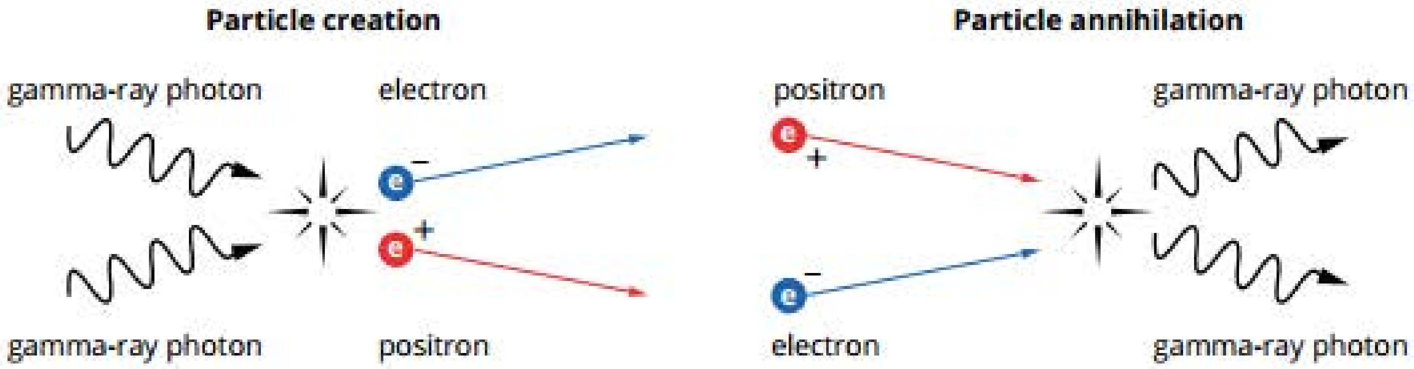


FIGURE 9.4.12 Electron–positron pairs were rapidly going in and out of existence in the early universe. As they met they annihilated, producing more gamma-ray photons, but the photons could also produce the electron–positron pairs.

Creation of lasting matter

To return to the question of the origin of our universe, you need to combine the two ideas just discussed—the inflation at the very beginning and the continual creation and annihilation of matter due to uncertainty. You will recall from Section 9.2 that because two opposite particles are always produced in this creation, the process is called pair production. Normally pair production doesn't result in the creation of any lasting matter. However, in the period of inflation, because of the extremely rapid expansion of space, the pairs of particles rapidly separated and didn't get a chance to annihilate. And so, in that tiny fraction of a second of inflation, huge amounts of matter were created.

In those first moments after inflation, the universe was chaotic with particles and antiparticles annihilating each other and producing high-energy gamma photons. Those photons themselves again collided with others and their energy formed new particles. But after about 0.0001 s the temperature of the expanding universe fell to about 10^{12} K, and that was too low for the creation of new particles. There was, therefore, wholesale annihilation of matter with antimatter and a huge reduction in the total amount of matter. The annihilation also produced an enormous increase in the amount of radiation. This huge fireball of radiation filled all space and dominated the universe for the next few hundred thousand years. It was, of course, the origin of the CMBR we now see.

In the first few seconds, while the temperature remained over a billion K or so, quarks combined to form protons and neutrons, then protons and neutrons were forced close enough to fuse together, forming hydrogen, helium and lithium nuclei (Figure 9.4.13). A little while later the temperature dropped below that needed for fusion and no further nuclei were formed.

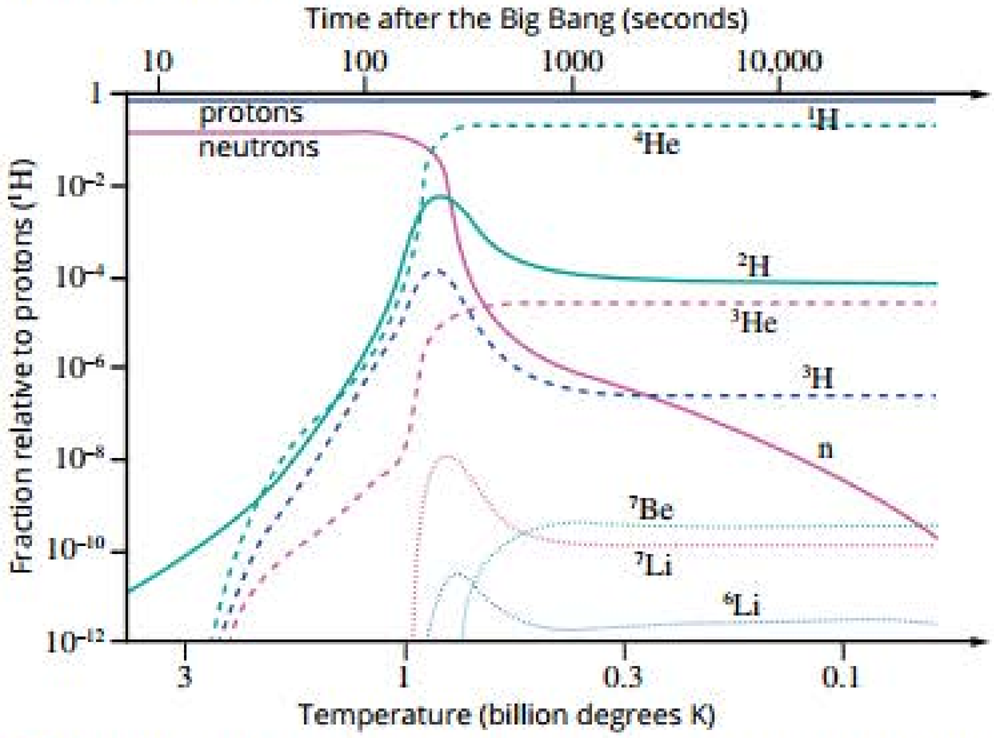


FIGURE 9.4.13 Most of the matter in the universe came into existence in the first few minutes after the Big Bang. Hydrogen (protons, ^{1}H) and helium (^{4}He) were by far the most dominant nuclei formed. Many neutrons were formed initially, but rapidly decayed into protons and electrons.

You might well ask how it was that any matter was left after all the annihilation of matter and antimatter. Strangely, there was not quite an even balance of matter and antimatter in the early universe, and so after the initial rapid annihilation, there was actually matter left over. The imbalance between matter and antimatter has been estimated as only about one extra matter particle in a billion particles—but it was just those 'one particle in a billion' that now make up our universe.

An observer would have found it immensely bright and very opaque because of the interactions of the high temperature photons with all the various particles. At about 300 000 years the expansion resulted in cooling to about 3000 K. At this temperature, the photon energy dropped into the infrared region of the spectrum; the photons no longer interacted with the nuclei and so the universe became completely dark. For this reason, it also became possible for the hydrogen and helium nuclei to hold on to electrons and form the first neutral atoms. This completely dark universe was extraordinarily thin, with just a few hydrogen and helium atoms scattered around in each cubic metre.

The photons left over at that time no longer interacted with anything and so have gone on flying through the universe ever since. They are what we see now as the CMBR, which give us a 'picture' of the universe at that time.

By looking into the far reaches of the universe, the Hubble Space Telescope is actually looking back in time, because of the time taken for light to reach us from that distance. We are seeing the galaxies as they were about 11 billion years ago (Figure 9.4.14). They are smaller and bluer than the older (closer) galaxies, which is consistent with the evolution of galaxies as expected from the Big Bang theory.

FIGURE 9.4.14 Small blue 'infant' galaxies are seen in the early universe when the Hubble Space Telescope peers back billions of years into the past.

The first stars

It wasn't until about a billion years after the Big Bang that things started to get more interesting. The average temperature had fallen to just 15 K (i.e. *very* cold) and, because of the slight unevenness in the early universe, matter started to clump together. Within those clumps, gravitational forces slowly but surely pulled vast numbers of atoms together in an almighty crunch. The collapse of all these atoms created the first stars. The gravitational energy released in the collapse raised the temperature to millions of degrees and hydrogen fusion started up again, raising the temperature even more and creating more helium and lithium atoms.

FROM THE BIG BANG TO THE EVOLUTION OF LIFE

The overwhelming abundance of hydrogen and helium in the universe is very good evidence that the Big Bang picture of creation is sound. However, a universe consisting of just hydrogen, helium and a little lithium could not produce life, which is dependent on a multitude of heavier elements, such as carbon. It took the supernova explosions of the early stars to produce the rest of the elements of the periodic table, elements from which life would finally start to evolve some 10 billion years after the Big Bang (Figure 9.4.15).

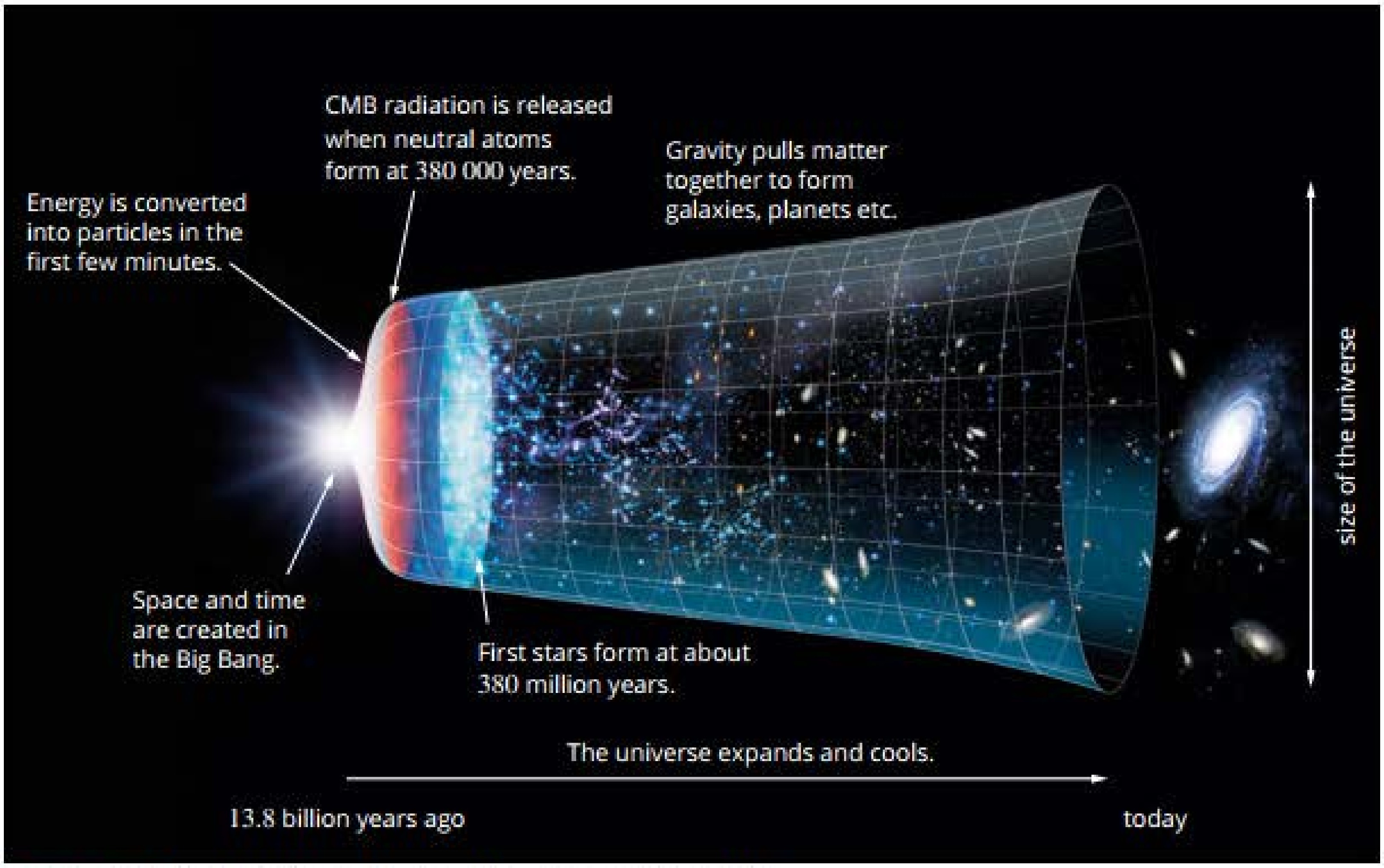


FIGURE 9.4.15 The 13.7 billion year history of the universe (not to scale).

9.4 Review

SUMMARY

- By studying Cepheid variable stars within spiral nebulae, Hubble was able to determine that these nebulae were actually other galaxies at vast distances.
- Most galaxies appeared to be 'redshifted'. This was taken to mean that they were moving away from us. The farther away, the faster this 'recessional velocity'—Hubble's law: $v = H_0 d$.
- As the universe appeared to be uniform and with no centre, the redshift was eventually interpreted not as a Doppler shift of galaxies moving through space but as the result of space itself expanding.
- Einstein's general theory of relativity had predicted this expansion, but he, like most physicists of the time, had felt that to be impossible.
- The implication seemed to be that the universe must, at some stage, either have started from an incredibly small point (the Big Bang) or it was infinite in time and space, with new matter being continuously created (the Steady State).
- The discovery of the cosmic microwave background radiation (CMBR) was strong evidence for the Big Bang theory, as it appeared to be the leftover radiation from the energy created in the Big Bang.
- Spacetime, matter and forces were all created at the Big Bang.
- Matter was created in the short time called inflation. The expansion was so rapid that matter and antimatter pairs of particles could not annihilate. After that the pairs did annihilate but there was slightly more matter than antimatter, and it is that remaining matter that makes up the universe today.
- In the early universe, only hydrogen and helium were formed before the universe cooled below the fusion temperature. This explains the present predominance of these elements in the universe and supports the Big Bang theory rather than the Steady State.
- The heavy elements necessary for our existence were formed later in the explosive supernovae of heavy stars.

KEY QUESTIONS

1 Which of the following is a correct description of redshift due to the Doppler effect? Explain your choice.

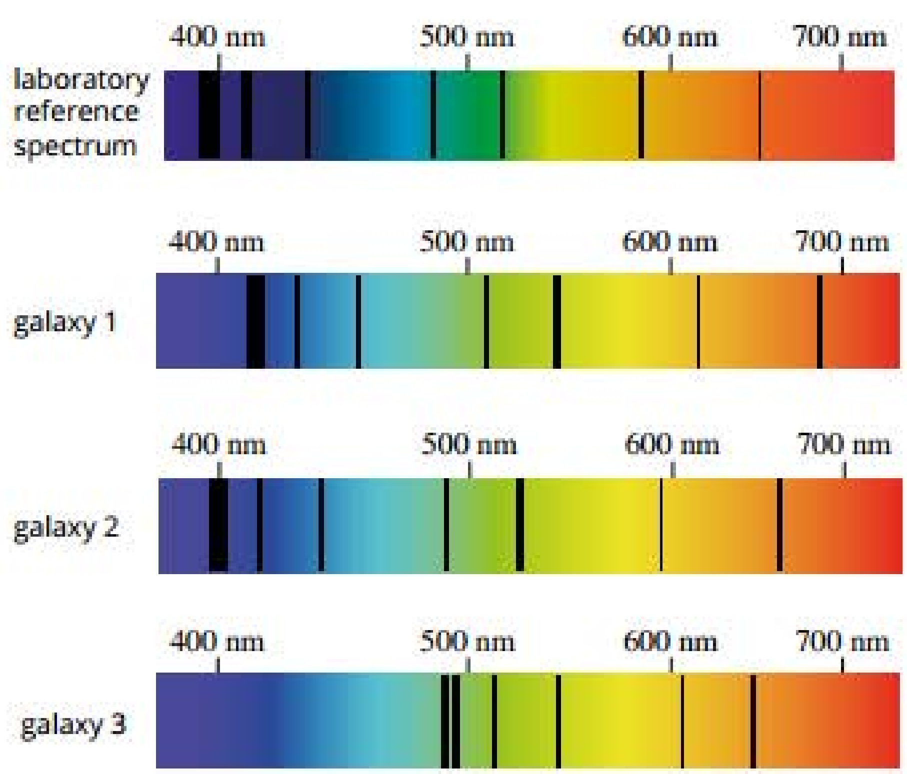
A The light from a source that is moving away from the observer will appear to have a longer wavelength and a lower frequency.

B The light from a source that is moving towards the observer will appear to have a longer wavelength and a lower frequency.

C The light from a source that is moving away from the observer will appear to have a shorter wavelength and a higher frequency.

D The light from a source that is moving towards the observer will appear to have a shorter wavelength and a higher frequency.

2 Sort the three spectra of galaxies in order of increasing distance of the galaxy from us, using the laboratory reference spectrum as your guide.



3 Explain what the value of the Hubble constant represents: $H_0 = 67.80\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$.

4 Use Hubble's law to determine the distance to a galaxy that is receding from us at $4 \times 10^7\,\text{m}\,\text{s}^{-1}$. Use $H_0 = 72\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$. Give your answer to one significant figure.

5 Outline how the expansion of the universe is different from a conventional explosion.

6 Which one or more of the statements below support the idea that space is expanding? Explain why your choice(s) is/are correct.

A All the stars and galaxies appear to be receding.

B The universe is expanding away from Earth.

C The stars and galaxies within space are not expanding.

D The same recession of stars and galaxies is seen throughout space.

7 What is the most likely cause of the cosmic microwave background radiation we now see?

8 Explain why cosmic microwave background radiation was seen as supporting the Big Bang theory over the Steady State theory.

9 Place the forces listed below into the order in which they became distinct from the other forces, from first to last.

weak nuclear, gravity, electromagnetic, strong nuclear

10 Most of the matter in the universe is hydrogen and helium. Explain how the Big Bang and Steady State theories account for the creation of hydrogen and helium.

Chapter review

09

KEY TERMS

annihilation
antiparticle
baryon
baryon number
Big Bang theory
blackbody radiation
conservation laws
cosmic microwave background radiation
electromagnetic force
electroweak force
electroweak theory
fermion
Feynman diagram
gauge bosons
gluon
hadron
Hubble constant
Hubble's law
intrinsic brightness
law of conservation of momentum
lepton
lepton number
nebula
pair production
parallax movement
parsec
Pauli's exclusion principle
particle accelerator
quantum chromodynamics
quantum number
quark
redshift
relativistic mass
special relativity
Standard Model
Steady State theory
strong nuclear force
weak nuclear force

1 Briefly outline two differences between leptons and quarks.

2 Which of the following correctly describes the strong nuclear force? Explain your choice.

- **A** The strong nuclear force attracts protons and neutrons.
- **B** The strong nuclear force attracts neutrons to protons.
- **C** The strong nuclear force attracts protons and neutrons to each other.
- **D** The strong nuclear force attracts electrons to the nucleus.

3 State the type of gauge bosons that mediate the forces below.

- **a** electromagnetic
- **b** strong nuclear
- **c** weak nuclear

4 Describe how the Standard Model explains three of the four fundamental forces. As part of your answer, identify the three forces and the associated particle(s).

5 Describe the groups into which the particles of the Standard Model are classified and highlight the differences between these groups.

6 Which of the fundamental forces is not currently considered to be part of the Standard Model.

7 How many joules of energy are given to an electron accelerated across a potential difference of 1×10^{10} V? Give your answer to two significant figures.

8 A particle event was thought to involve the decay of a meson called a pion (π^+) into a neutron (n) and an antimuon (μ^+). Determine if any of the conservation laws are violated by this reaction.

9 Suggest a reason why synchrotrons are used in modern particle physics experiments.

10 A proton travels perpendicular to a uniform magnetic field at 1.2×10^8 m s^{-1}. The rest mass of a proton is $m_0 = 1.672 \times 10^{-27}$ kg and the charge on the proton is $+1.6 \times 10^{-19}$ C. Assume that the increase in mass of the particle at this velocity is negligible. Calculate the strength of the magnetic field required to cause the radius of the circular path of the proton to be 2 m. Give your answer to one significant figure.

11 An electron ($m_0 = 9.109 \times 10^{-31}$ kg) is accelerated from rest to 0.98c. Calculate its increase in energy as a percentage of the rest energy.

12 The energy of colliding particles is commonly converted into mass in particle physics experiments using particle accelerators. Which two properties describe the types of energy that contribute to the total energy available in a collision at relativistic speeds?

13 Describe the operation of a linear accelerator such as the one at the Stanford Linear Accelerator Center.

14 Describe how Hubble could estimate the distances to spiral nebulae using stars called Cepheid variables.

15 An interaction between a proton and a neutron is thought to produce two protons and a meson (π^-), which has a negative charge. Is this reaction allowed or forbidden? Which conservation law applies if it is allowed, or is broken if it is forbidden?

16 Explain what happens in the process called annihilation.

17 A gamma ray photon interacts with an atomic nucleus and an electron–positron pair is produced. Each particle created has a mass of 0.511 MeV and kinetic energy of 0.250 MeV. How much energy did the gamma ray photon carry?

18 If the Hubble constant is 70 km s^{-1} Mpc^{-1}, estimate a value for the age of the universe in years.

19 Describe briefly how Edwin Hubble's observations of the redshifts of galaxies were used to formulate Hubble's law, and explain how Hubble's law is used to support the Big Bang theory.

20 Which of the following correctly describes the operation of a linear accelerator that produces electrons travelling at 0.9958c? Explain your choice.

A Work is done in the opposite direction to the field between the drift tubes to accelerate the electrons. The increasing length of the drift tubes ensures the field is in the opposite direction to the electrons' motion, as the frequency of the alternating potential difference is constant.

B Work is done in the direction of the field between the drift tubes to accelerate the electrons. The increasing length of the drift tubes ensures the field is in the same direction to the electrons' motion, as the frequency of the alternating potential difference is constant.

C Work is done in the opposite direction to the field between the drift tubes to accelerate the electrons. The increasing length of the drift tubes ensures the field is in the opposite direction to the electrons' motion, as the frequency of the alternating potential difference is varied.

D Work is done in the direction of the field between the drift tubes to accelerate the electrons. The increasing length of the drift tubes ensures the field is in the same direction to the electrons' motion, as the frequency of the alternating potential difference is varied.

21 One component of the Standard Model of particle physics predicted the existence of massive gauge bosons to mediate the weak nuclear force. A huge amount of money was spent and decades of experiments were conducted in the search for these bosons. Which of the following best describes these efforts in relation to the Standard Model? Explain why the other alternatives are incorrect.

A These experiments were done to prove the Standard Model was wrong.

B These experiments were done to test the predictions of the Standard Model.

C These experiments were done to prove the Standard Model was correct.

D These experiments were done to win a Nobel prize.

CHAPTER

10 Practical investigation

This chapter covers most of the skills needed to successfully plan and conduct a practical investigation.

Section 10.1 is a guide to designing and planning an investigation, including how to write a hypothesis, and how to identify the variables. It explains validity, reliability and accuracy, to assist in planning an investigation appropriately.

Section 10.2 is a guide to conducting investigations. It describes methods for accurately collecting and recording data to reduce errors. It explores presenting data using tables and graphs, to aid in selecting the most appropriate format for presenting the results.

Section 10.3 explains how to discuss an investigation and draw evidence-based conclusions that relate to the hypothesis and research question.

Practical investigation steps

The size and scope of a practical investigation can be initially quite daunting, but establishing a task list and timeline will help break it down into manageable steps. The entire task is expected to take between 7 and 10 hours.

Here are some steps that will need to be considered in a timeline:

- Determine the topic and type of investigation.
- Research and write down the theory on which the investigation is based.
- Determine an appropriate question to answer, and formulate a hypothesis.
- Identify the independent, dependent and controlled variables.
- Select equipment and resources needed for the investigation.
- Determine an appropriate procedure (methodology), considering validity, reliability and accuracy.
- Assess the risks and ethical issues and identify measures to address these.
- Conduct the investigation and record all data obtained.
- Analyse and evaluate the data.
- Evaluate your methods. Suggest ways of improving or extending the investigation.
- Write an evidence-based conclusion. Describe the limitations of the study.
- Write the final report or poster. (This should not be the focus of the investigation but rather an opportunity to communicate the investigation process and your conclusions.)

Some of these tasks are larger and will require more time than others. Many will overlap. Plan a realistic approach, consult with teachers to establish school-based time constraints and fix dates for the completion of each task. Allow time for reflection and to review your earlier work.

Key knowledge statements

Science Inquiry Skills

- identify, research, construct and refine questions for investigation; propose hypotheses; and predict possible outcomes
- design investigations, including the procedure(s) to be followed, the materials required, and the type and amount of primary and/or secondary data to be collected; conduct risk assessments; and consider research ethics
- conduct investigations, including the manipulation of force measurers and electromagnetic devices, safely, competently and methodically for the collection of valid and reliable data
- represent data in meaningful and useful ways, including using appropriate Système Internationale (SI) units and symbols, and significant figures; organise and analyse data to identify trends, patterns and relationships; identify sources of random and systematic error and estimate their effect on measurement results; identify anomalous data and calculate the measurement discrepancy between experimental results and a currently accepted value, expressed as a percentage; and select, synthesise and use evidence to make and justify conclusions
- interpret a range of scientific and media texts, and evaluate processes, claims and conclusions by considering the quality of available evidence; and use reasoning to construct scientific arguments
- select, construct, and use appropriate representations, including text and graphic representations of empirical and theoretical relationships, flow diagrams, nuclear equations, and circuit diagrams, to communicate conceptual understanding, solve problems and make predictions
- select, construct, and use appropriate representations, including text and graphic representations of empirical and theoretical relationships, vector diagrams, free body/force diagrams, field diagrams and circuit diagrams, to communicate conceptual understanding, solve problems and make predictions
- select, use, and interpret appropriate mathematical representations, including linear and non-linear graphs and algebraic relationships representing physical systems, to solve problems and make predictions
- communicate to specific audiences and for specific purposes using appropriate language, nomenclature, genres and modes, including scientific reports

10.1 Designing and planning the investigation

Taking the time to carefully plan and design a practical investigation before you begin will help you to maintain a clear and concise focus throughout. Preparation is essential. Ensure you understand the theory behind the investigation and prepare a detailed plan for the practical components of the investigation. This section is a guide to some of the key steps that should be taken when planning and designing a practical investigation (Figure 10.1.1).

FIGURE 10.1.1 There are many elements to a practical investigation, which may appear overwhelming to begin with. Taking a step-by-step approach will help the process and assist you in completing a worthwhile investigation.

DETERMINING THE TYPE OF INVESTIGATION

In your course, you will be required to undertake an investigation related to an area you have studied. Your teacher might suggest a particular topic or you may be required to suggest a topic that interests you. While your teacher may suggest the topic (such as circular motion) in most cases you will need to come up with the research question, hypothesis, variables, method, analysis and conclusion yourself.

Several types of research methods are used in science. Physics investigations typically fall into two methodologies:

- analysing the slope of a linear graph
- determining the relationship between two continuous **variables**.

From your research on your topic you may find that there is a mathematical relationship between the variables that you are exploring. It may be possible to use your algebra skills to write the mathematical relationship in the form of a linear equation, $y = mx + c$, where y is the dependent variable and x is the independent variable. A graph of the y and x variables will result in a linear graph with a slope equal to a constant, m, or collection of constants. By equating the value of the slope to these constants some meaningful analysis can be made, such as finding a value for the acceleration due to gravity or the specific heat capacity of a metal.

If, in your research, you find that there is no equation that links your variables, then you can investigate to determine what the relationship might be. By graphing your variables, you may find that the relationship is linear, or it may be inverse, exponential or even inverse squared. An investigation of this kind results in a calibration curve, which can be used to predict values of the dependent variable given particular values of the independent variable.

DEVELOPING RESEARCH QUESTIONS AND AIMS, FORMULATING HYPOTHESES AND IDENTIFYING VARIABLES

The research question, aim and hypothesis are interlinked. It is important to note that each of these can be refined as the planning of the investigation continues.

Formulating a question

A research question is a question that comes from an inquiring mind. When you are actively involved in developing your scientific understanding then you will want to know what factors affect a variable, or what is the relationship between two variables. The research question poses the question that the investigation seeks to answer. For example: What is the relationship between voltage and current?

Before formulating a question, it is good practice to conduct a literature review of the topic to be investigated. You should become familiar with the relevant scientific concepts and key terms.

During this review, write down questions, correlations, or equations as they arise.

Compile a list of possible ideas. Do not reject ideas that initially might seem impossible. Use these ideas to generate questions that are answerable.

Before constructing a hypothesis, formulate a question that needs an answer. This question will lead to a hypothesis when:

- the question is reduced to measurable variables
- a prediction is made based on knowledge and experience.

Evaluating your question

Once a question has been chosen, stop to evaluate the question before progressing. The question may need further refinement or even further investigation before it is suitable as a basis for an achievable and worthwhile investigation. A major planning point is to attempt something that it is possible to complete in the time available or with the resources on hand. It might be a little difficult to create a particularly complicated device with the facilities available in the school laboratory.

To evaluate the question, consider the following:

- Relevance: Is the question related to the appropriate area of study?
- Clarity and measurability: Can the question lead to a clear hypothesis?

  If the question cannot lead to a specific hypothesis, then it is going to be very difficult to complete the research.
- Time frame: Can the question be answered within a reasonable period of time? Is the question too broad?
- Knowledge and skills: Do you have a level of knowledge and a level of laboratory skills that will allow the question to be explored? Keep the question simple and achievable.
- Practicality: Are resources, such as laboratory equipment and materials, likely to be readily available? Keep things simple. Avoid investigations that require sophisticated or rare equipment. Equipment that is more-readily available includes timing devices, objects that could be used as projectiles, a tape measure and other common laboratory equipment.
- Safety and ethics: Consider the safety and ethical issues associated with the question you will be investigating. If there are issues, can these be addressed?
- Advice: Seek advice from your teacher about the question. Their input may prove very useful. Their experience may lead them to consider aspects of the question that you have not thought about.

Defining the aim of the investigation

An aim is a statement describing in detail what will be investigated to answer the research question. For example: The aim of the experiment is to investigate the relationship between the voltage and current in a circuit of constant resistance. Each aim should directly relate to the variables that will be referred to in the hypothesis. The aims do not need to include the details of the method.

Example

- Aim: The aim of the experiment is to investigate the relationship between mass and acceleration, when a constant force is applied.

Hypothesis

A hypothesis is a definite statement, based on previous knowledge and evidence or observations, that attempts to answer the research question. The hypothesis must relate the independent and dependent variables and describe the relationship between them. For example: Increasing the voltage supplied to a circuit of constant resistance increases the current in a linear way.

Here are some further examples of hypotheses:

- For a constant force, if the mass is increased, the acceleration is decreased as an inverse relationship.
- As the rotational speed of an electric motor increases, the current flowing into the motor will decrease as an inverse relationship.

- As the length of a pendulum increases, the square of the period of oscillation will increase as a linear relationship.
- As the height from which an object is dropped increases, the final velocity of the object will increase as a squared relationship.

There are no wrong or right hypotheses. You might formulate a hypothesis that a more experienced person will disagree with; however, the purpose of an investigation is to find the answer to a research question. If the answer to the question supports your hypothesis, then that is a positive result, as it will confirm your understanding of the concept. On the other hand, if your investigation does not support your hypothesis, you can now say that your original understanding was not correct and you can change your understanding to a more scientific one. Some of you might notice that the following hypothesis will not be supported by the investigation:

- The greater the mass of a marble, the faster it will hit the ground, when dropped from the same height.

This doesn't mean that the hypothesis is wrong, but it may indicate that there was some misconception that you had that was not exposed in your literature review.

Formulating a hypothesis

A good hypothesis should:

- be a definite statement of the relationship
- include an independent and a dependent variable that is continuous and measurable
- be worded so that it can be tested in the experiment

The hypothesis should also be falsifiable. This means that a negative outcome would disprove it. For example, the hypothesis that all apples are round cannot be proved beyond doubt, but it can be disproved—in other words, it is falsifiable. In fact, only one oval-shaped apple is needed to disprove this hypothesis. Unfalsifiable hypotheses cannot be tested by science. These include hypotheses on ethical, moral and other subjective judgements.

Variables

A good scientific hypothesis can be tested—that is, it can be supported or refuted through investigation. To be a testable hypothesis, it should be possible to measure both the change or treatment and the effect, or what will happen. The factors that can be changed, or are changed as a result of the experiment or investigation, are called the variables. An experiment or investigation determines the relationship between variables.

There are three categories of variables:

- The **independent variable** is the variable that is controlled by the researcher (the one that is selected and changed). You must test only one independent variable in any investigation, otherwise it cannot be stated that the changes in the dependent variable are the result of changes in the independent variable.
- The **dependent variable** is the variable that may change in response to a change in the independent variable. This is the variable that will be measured or observed. You should measure only one dependent variable in any investigation. If you want to measure another dependent variable then you will need to do another investigation with another hypothesis.
- **Controlled variables** are all the variables that must be kept constant during the investigation otherwise the test cannot be fair.

Read the following example of a hypothesis.

As the rotational speed of an electric motor increases, the current flowing into the motor decreases as an inverse relationship.

Identify the different variables.

- independent variable: rotational speed of the electric motor
- dependent variable: current flowing into the motor
- controlled variables: applied potential difference, temperature, load on the motor

Completing a table like Table 10.1.1 will assist in evaluating the question or questions.

TABLE 10.1.1 Break the question down to determine the variables.

Research question	How does the power of a kettle affect the time taken to boil water?
Independent variable	the power of the kettle
Dependent variable	the time the kettle takes to boil water
Controlled variables	mass of the water, purity of the water, starting temperature of the water and kettle
Potential hypothesis	The greater the power of a kettle, the less time it will take to boil water, as an inverse relationship.

Qualitative and quantitative variables

Variables are either qualitative or quantitative, with further subsets in each category.

- **Qualitative variables** can be observed but not measured; for example, describing a light globe as bright or dim. They can only be sorted into groups or categories such as brightness, type of material of construction or type of device.
  - Nominal variables are categorical variables in which the order is not important; for example, the type of material or type of device.
  - Ordinal variables are categorical variables in which order is important and groups have an obvious ranking or level; for example, brightness (Figure 10.1.2).
- **Quantitative variables** can be measured. Length, area, weight, temperature and cost are all examples of quantitative data.
  - Discrete variables consist of only integer numerical values, not fractions; for example, the number of pins in a packet, the number of springs connected together, or the energy levels in atoms.
  - Continuous variables allow for any numerical value within a given range; for example, the measurement of temperature, length, weight, and frequency.

In physics, you should ensure that you choose quantitative variables for both the independent and dependent variables. This will allow you to construct a line graph, and therefore determine the slope of the line, or the relationship between the variables.

FIGURE 10.1.2 When recording qualitative data, describe in detail how each variable will be defined. For example, if recording the brightness of light globes, light meters are a quantitative way to gather data.

WRITING THE METHODOLOGY

The methodology, or method, of your investigation is a step-by-step procedure. When detailing the method, ensure it enables you to conduct a valid, reliable and accurate investigation.

Validity

Validity refers to whether an experiment is in fact testing the hypothesis. Is the investigation obtaining data that is relevant to the question, or is it flawed?

To ensure an investigation is valid, it should be designed so that only one variable is being changed at a time. The remaining variables must remain constant, so that meaningful conclusions can be drawn about the effect of the independent variable alone.

To ensure validity, you must carefully determine:

- the independent variable—the variable that will be changed, and how it will change
- the dependent variable—the variable that will be measured
- the controlled variables—the variables that must remain constant.

Reliability

Reliability refers to the idea that the experiment can be repeated many times and will obtain consistent results. You can maintain the investigation's reliability by:

- listing and defining the control variables and how they will be kept constant
- listing the detailed steps that you will take to conduct the experiment, describing what you will do and how you will measure and record data
- ensuring that there are enough changes of the independent variable. Typically, five changes over a wide range of the independent variable are considered sufficient.
- ensuring there are enough trials conducted for each value of the independent variable. Typically, you should conduct at least three trials repeating the experiment, then average the three measurements. This reduces random errors and allows systematic errors to be identified. If a reading differs too much from the rest (known as an outlier), discard it before averaging (Figure 10.1.3).

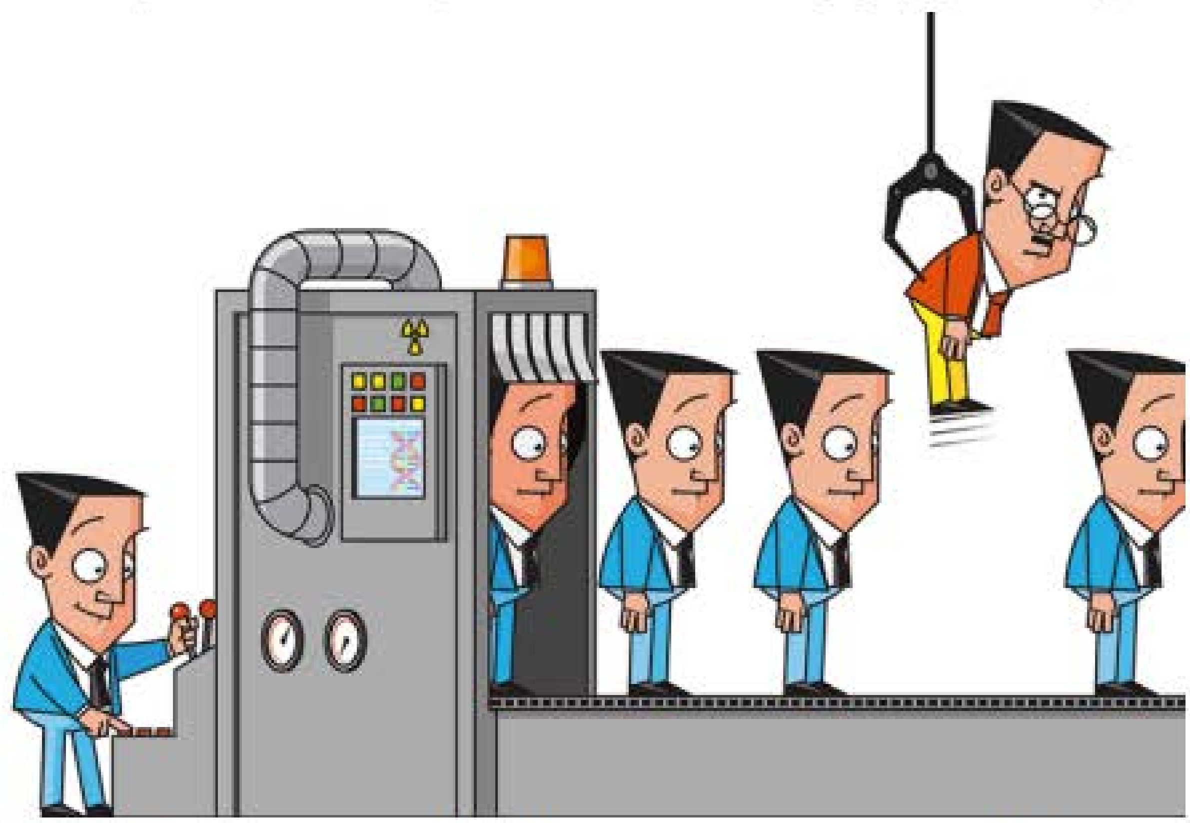


FIGURE 10.1.3 Replication increases the reliability of your investigation. It ensures that if anyone repeats the investigation they will obtain similar data.

Accuracy and precision

Precision refers to the extent to which the instrument can make repeated measures of the dependent variable that are the same for the same value of the independent variable. For example, if each measurement of the current in an electrical circuit is within 0.1 A of the others, then the device is more precise than a device for which there is a difference of 0.5 A. Accuracy refers to the ability to obtain the correct measurement.

You will need to consider if the instruments to be used are sensitive enough. Build some testing into your investigation to confirm the accuracy and reliability of the equipment and your ability to read the information obtained.

Reasonable steps to ensure the accuracy of the investigation include considering:

- the type of instrument that will be used to measure the independent and dependent variables.
- calibrating the measuring equipment by testing a standard.

Describe the materials and method in appropriate detail in your logbook. This should ensure that every measurement can be repeated and the same result obtained within reasonable margins of experimental error (less than 5% is reasonable).

Data analysis

Data analysis is part of the method. Consider how the data will be presented and analysed. A wide range of analysis tools are available. For example, tables can be used to organise data so that patterns can be established, and graphs can show relationships and comparisons. In fact, preparing an empty table showing the data that needs to be obtained will help in the planning of the investigation. See page 398 for more information on organising a data table.

Sourcing appropriate materials and technology

When designing your investigation, you will need to decide on the materials, technology and instrumentation that will be used to carry out your research. It is important to find the right balance between items that are easily accessible and those that will give you accurate results. As you move onto conducting your investigation, it will be important to take note of the precision of your chosen instruments and how this affects the accuracy and validity of your results.

Modifying the methodology

The methodology may need modifying as the investigation is carried out. The following actions will help to determine any issues in the methodology and how to modify them:

- Record everything.
- Be prepared to make changes to the approach.
- Note any difficulties encountered and the ways they were overcome. What were the failures and successes? Every test carried out can contribute to the understanding of the investigation as a whole, no matter how much of a disaster it may first appear.
- Do not panic. Go over the theory again, and talk to the teacher and other students. A different perspective can lead to a solution.

If the expected data is not obtained, don't worry. As long as it can be critically and objectively evaluated, the limitations of the investigation are identified and further investigations proposed, the work is worthwhile.

COMPLYING WITH ETHICAL AND SAFETY GUIDELINES

Ethical considerations

Some investigations require an ethics approval—consult with the teacher. In fact, when deciding on an investigation, identify all possible ethical considerations and evaluate whether those parts of the investigation are necessary or if there are ways you can reduce or mitigate them.

Occupational health and safety

While planning for an investigation, it is important for the safety of yourself and the safety of others that the potential risks are considered.

FIGURE 10.1.4 When planning an investigation, you need to identify, assess and control hazards.

Everything we do has some risk involved. Risk assessments are performed to identify, assess and control hazards. A risk assessment should be performed for any situation, in the laboratory or outside in the field. Always identify the risks and control them to keep everyone safe. For example, carry out voltage–current experiments with low voltages (less than 6.0 V DC or 4 × 1.5 V batteries) coupled to resistors so that the currents in the circuits are of the order of milliamps. *At all times* avoid direct exposure to 240 V AC household voltages (Figure 10.1.4).

To identify risks, think about:

- the activity that will be carried out
- the equipment or chemicals that will be used.

The following hierarchy of risk controls is organised from most effective to least effective:

1. *Elimination:* Eliminate dangerous equipment, procedures or substances.
2. *Substitution:* Find different equipment, procedures or substances to use that will achieve the same result, but have less risk associated with them.
3. *Isolation:* Ensure there is a barrier between the person and the hazard. Examples include physical barriers such as guards in machines, or fume hoods to work with volatile substances.
4. *Engineering controls:* Modify equipment to reduce risks.
5. *Administrative controls:* Provide guidelines, special procedures, warning signs and information about safe behaviours for any participants.
6. *Personal protective equipment (PPE):* Wear safety glasses, lab coats, gloves and respirators etc. where appropriate, and provide these to other participants.

Science outdoors

Sometimes investigations and experiments will be carried out outdoors. Working outdoors has its own set of potential risks and it is equally important to consider ways of eliminating or reducing these risks.

As an example, Table 10.1.2 contains examples of risks associated with field work outdoors.

TABLE 10.1.2 Risks associated with fieldwork outdoors.

Risks	Control measures
sunburn	wear sunscreen, a hat and sunglasses
hot or cold weather	wear clothing to protect against heat or cold
projectile launch	create barriers so that people know not to enter the area
trip hazards	minimise the use of cables (electrical, computer) and cover them up with matting be aware of tree roots, rocks etc.

First aid measures

Minimising the risk of injury reduces the chance of requiring first aid assistance. However, it is still important to have someone with first aid training present during practical investigations. Always tell the teacher or laboratory technician if an injury or accident happens.

Personal protective equipment

Everyone who works in a laboratory wears items that help keep them safe. This is called **personal protective equipment (PPE)** and includes:

- safety glasses
- shoes with covered tops
- disposable gloves when handling chemicals
- a disposable apron or a lab coat if there is risk of damage to clothing
- ear protection if there is risk to hearing.

10.1 Review

SUMMARY

- An aim is a statement describing in detail what will be investigated. For example: The aim of the experiment is to investigate the relationship between force, mass and acceleration.
- A hypothesis is a definite statement of the relationship between the independent and dependent variables based on previous knowledge and evidence or observations that attempts to answer the research question. For example: With the force kept constant, the acceleration decreases with increasing mass as an inverse relationship.
- Once a question has been chosen, stop to evaluate the question before progressing. The question may need further refinement or even further investigation before it is suitable as a basis for an achievable and worthwhile investigation. Make sure that it is possible to complete the activity in the time available and with the resources on hand. It might be a little difficult to create a particularly complicated device with the facilities available in the school laboratory.
- There are three categories of variables:
  - The independent variable is the variable that is controlled by the researcher (the one that is selected and changed).
  - The dependent variable is the variable that may change in response to a change in the independent variable. This is the variable that will be measured or observed.
  - Controlled variables are all the variables that must be kept constant during the investigation so that it is a fair test.
- The methodology of your investigation is a step-by-step procedure. When detailing the methodology, ensure it complies as a valid, reliable and accurate investigation.
- It is also important to determine how many times the independent variable needs to be changed and how many trials need to be run for each change in the independent variable.
- Data analysis is part of the method. Consider how the data will be presented and analysed. A wide range of analysis tools could be used. For example, tables organise data so that patterns can be established and graphs can show relationships and comparisons.
- In every investigation you need to consider the risks and potentially hazardous situations, and act to minimise those risks.

KEY QUESTIONS

1 In a practical investigation the student changes the voltage by adding or subtracting batteries in series to the circuit.
  a How could the voltage be a discrete value?
  b How could it be continuous?

2 In another experiment the student uses the following range of values to describe the brightness of a light: dazzling, bright, glowing, dim, off
  What type of measurement is the variable 'brightness'?

3 Select the best hypothesis from the three options below. Give reasons for your choice.
  A Hypothesis 1: The greater the mass of the marble you drop, the greater the final velocity of the marble as a linear relationship.
  B Hypothesis 2: The greater the potential difference across a resistor, the greater the current through it.
  C Hypothesis 3: Different metals will have different resistances.

4 Give the correct term that describes an experiment with each of the following conditions.
  a The experiment addresses the hypothesis and aims.
  b The experiment is repeated and consistent results are obtained.
  c Appropriate equipment is chosen for the desired measurements.

5 A student wanted to find out how the tension in an elastic band affects the initial velocity of the band when it is launched from her finger. State:
  a the independent variable
  b the dependent variable
  c three controlled variables.

10.2 Conducting investigations, and recording and presenting data

FIGURE 10.2.1 When carrying out your investigation try to maintain high standards to minimise potential errors.

Once the planning and design of a practical investigation is complete, the next step is to undertake the investigation and record the results. As with the planning stages, there are key steps and skills to keep in mind to maintain high standards and minimise potential error throughout the investigation (Figure 10.2.1).

This section will focus on the best methods of conducting a practical investigation, of systematically generating, recording and processing data, and of presenting it in a concise and clear manner.

CONDUCTING INVESTIGATIONS TO COLLECT AND RECORD DATA

For an investigation to be scientific, it must be objective and systematic. Ensuring familiarity with the methodology and protocols before beginning will help you to achieve this.

When working, keep asking questions. Is the work biased in any way? If changes are made, how will they affect the study? Will the investigation still be valid for the aim and hypothesis?

It is essential that during the investigation the following are recorded in the logbook:

- all quantitative data collected
- the methods used to collect the data
- any incident, feature or unexpected event that may have affected the quality or validity of the data.

The data recorded in the logbook is the **raw data**. Usually this data needs to be processed in some manner before it can be presented. If an error occurs in the processing of the data or you decide to present the data in an alternative format, the recorded raw data will always be available for you to refer back to.

IDENTIFYING ERRORS

Most practical investigations have errors associated with them. Errors can occur for a variety of reasons. Being aware of potential errors helps you to avoid or minimise them. For an investigation to be accurate, it is important to identify and record any errors.

There are two types of errors:

- systematic errors
- random errors.

Systematic errors

A **systematic error** is an error that is consistent and will occur again if the investigation is repeated in the same way.

Systematic errors are usually a result of instruments that are not calibrated correctly or methods that are flawed.

An example of a systematic error would be if a ruler mark indicating 5 cm from 0 cm was actually only 4.9 cm from 0 cm due to a manufacturing error or shrinkage of the wood. Another example would be if the researcher repeatedly used a piece of equipment incorrectly throughout the entire investigation. Figure 10.2.2 shows how traffic police reduce systematic errors in their data collection.

FIGURE 10.2.2 To avoid a systematic error, make sure that you are using measuring equipment correctly. Laser speed guns, for example, need to be placed on a stationary support so the aim point is held on a single target point for the duration of the read.

Random errors

Random errors occur in an unpredictable manner and are generally small. A random error could be, for example, the result of a researcher reading the same result correctly one time and incorrectly another time. Another example would be if an instrument were affected by a power cut, or low battery power.

Techniques for reducing error

Designing the method carefully, including selection and use of equipment, will help reduce errors.

Appropriate equipment

Use the equipment that is best suited to the data that needs to be collected to validate the hypothesis. Determining the units of the data being collected and at what scale will help to select the correct equipment. Using the right unit and scale will ensure that measurements are more accurate and precise (with smaller systematic errors).

Significant figures are the numbers that convey meaning and precision. The number of significant figures used depends on the scale of the instrument. It is important to record data to the number of significant figures available from the equipment or observation. Using either a greater or smaller number of significant figures can be misleading.

Review the following examples to learn more about significant figures:

- 15 has two significant figures
- 3.5 has two significant figures
- 3.50 has three significant figures
- 0.037 has two significant figures
- 1401 has four significant figures.

To calculate gravitational potential energy (E_g), the formula is $E_g = mg\Delta h$.

If $g = 9.80\,\text{m}\,\text{s}^{-2}$, mass = 7.50 kg, height = 0.64 m (64 cm):

$E_g = 9.80 \times 7.50 \times 0.64 = 47.09\,\text{J}$

But only quote the answer to the least number of significant figures in the data; that is, to two significant figures, so $E_g = 47\,\text{J}$.

Although digital scales can measure to many more than two figures and calculators can give 12 figures, be sensible and follow the significant figure rules.

Calibrated equipment

Some equipment, such as some motion sensors, needs to be calibrated before use to account for the temperature at the time. Before carrying out the investigation, make sure the instruments or measuring devices are properly calibrated and are, in general, functioning correctly. For example, measure the temperature and apply a correction to the speed of sound to calibrate a motion sensor if necessary.

Correct use of equipment

Use the equipment properly. Ensure any necessary training has been done to use the equipment and that you have had an opportunity to practice using the equipment before beginning the investigation. Improper use of equipment can result in inaccurate, imprecise data with large errors, and the validity of the data can be compromised.

Incorrect reading of measurements is a common misuse of equipment. Make sure all the equipment needed in the investigation can be used correctly and record the instructions in detail so they can be referred back to if the data doesn't appear correct.

RECORDING AND PRESENTING QUANTITATIVE DATA

Raw data is unlikely to be used directly to validate the hypothesis. However, raw data is essential to the investigation and plans for collecting the raw data should be made carefully. Consider the formulae or graphs that will be used to analyse the data at the end of the investigation. This will help to determine the type of raw data that needs to be collected in order to validate the hypothesis.

For example, to calculate take-off velocity for a vertical jump, three sets of raw data will need to be collected using a force platform: the athlete's standing body weight, the ground reaction force and the time during the vertical jump. The data can then be processed to obtain the take-off impulse.

Once you have determined the data that needs to be collected, prepare a table in which to record the data.

ANALYSING AND PRESENTING DATA

The raw data that has been obtained needs to be presented in a way that is clear, concise and accurate.

There are a number of ways of presenting data, including tables, graphs, flow charts and diagrams. The best way of visualising the data depends on its nature. Try several formats before making a final decision, to create the best possible presentation.

Presenting raw and processed data in tables

Tables organise data into rows and columns, and can vary in complexity according to the nature of the data. Tables can be used to organise raw data and processed data or to summarise results.

The simplest form of a table is a five-column format. In a five-column table, the first column should contain the independent variable (the one being changed) and the second to fourth columns should contain the three trials of the dependent variable (the one that may change in response to a change in the independent variable). The final column should contain the average of the three trials.

Tables should have the following features:

- a descriptive title that contains both the independent and depended variables
- column headings (including the unit and the uncertainty)
- aligned figures (align the decimal points)
- the independent variable placed in the left column
- the three trials of the dependent variable placed in the right columns with the average column on the end.

Processed data with the units and uncertainty if this is required for graphing e.g. if T^2 needs to be plotted instead of T.

Look at the table in Figure 10.2.3, which has been used to organise raw and processed data about the effect of current on voltage.

Effect of potential difference on the current through a resistor ← clear title

Potential difference (V)	Current trial 1 ± 0.01 (A)	Current trial 2 ± 0.01 (A)	Current trial 3 ± 0.01 (A)	Average current (A)
1.50	0.31	0.34	0.32	0.32 ± 0.02
2.00	0.42	0.45	0.41	0.43 ± 0.02
2.50	0.50	0.51	0.52	0.51 ± 0.01
3.00	0.62	0.65	0.58	0.62 ± 0.04
3.50	0.70	0.71	0.70	0.70 ± 0.01

← headings for each column with units and uncertainties

← consistent number of significant figures

↑ independent variable

↑ consistent number of decimal places

↑ averages with uncertainties

FIGURE 10.2.3 A simple table listing the raw data obtained in the second, third, and fourth columns and processed data in the fifth column.

A table of processed data usually presents the average values of trials, the **mean**. However, the mean on its own does not provide an accurate picture of the results.

To report processed data more accurately, the uncertainty should be presented as well.

Uncertainty

When there is a range of measurements of a particular value, the mean must be accompanied by the **uncertainty**, for your results to be presented as a mean in an accurate way. In other words, the mean must be accompanied by a description of the range of data obtained.

Uncertainty = ± (maximum variance from the mean)

For example, the speed, in $km\,h^{-1}$, of cars travelling down a certain road was:

46, 50, 55, 48, 50, 58, 45

The average speed would be:

$(46 + 50 + 55 + 48 + 50 + 58 + 45) \div 7 = 50\,km\,h^{-1}$

The uncertainty would be the maximum variance from the average: 58 is 8 above the average, so the uncertainty is 8.

This data should be presented as:

Average speed is $50 \pm 8\,km\,h^{-1}$.

Other descriptive statistics measures

The mean and the uncertainty are statistical measures that help describe data accurately. Other statistical measures that can be used, depending on the data obtained, are:

- **mode**: the mode is the value that appears most often in a data set. This measure is useful to describe qualitative or discrete data (for example, the mode of the values 0.01, 0.01, 0.02, 0.02, 0.02, 0.03, 0.04 is 0.02).
- **median**: the median is the 'middle' value of an ordered list of values (for example, the median of the values 5, 5, 8, 8, 9, 10, 20 is 8). The median is used when the data range is spread, for example, due to the presence of unusual results, making the mean unreliable.

Graphs

In general, tables provide more detailed data than graphs, but it is easier to observe trends and patterns in data in graphical form than in tabular form.

Graphs are used when two variables are being considered and one variable is dependent on the other. The graph shows the relationship between the variables.

There are several types of graphs that can be used, including line graphs, bar graphs and pie charts. The best one to use will depend on the nature of the data.

General rules to follow when making a graph (Figure 10.2.4) include the following:

- Keep the graph simple and uncluttered.
- Use a descriptive title that contains the independent and dependent variables.
- Represent the independent variable on the x-axis and the dependent variable on the y-axis.
- Make axes proportionate to the data.
- Clearly label axes with both the variable and the unit in which it is measured.
- Include error bars showing the uncertainty of each point. The error bar should extend above and below the plotted point equal to the uncertainty in the dependent variable and to the left and right of the plotted point equal to the uncertainty of the independent variable.

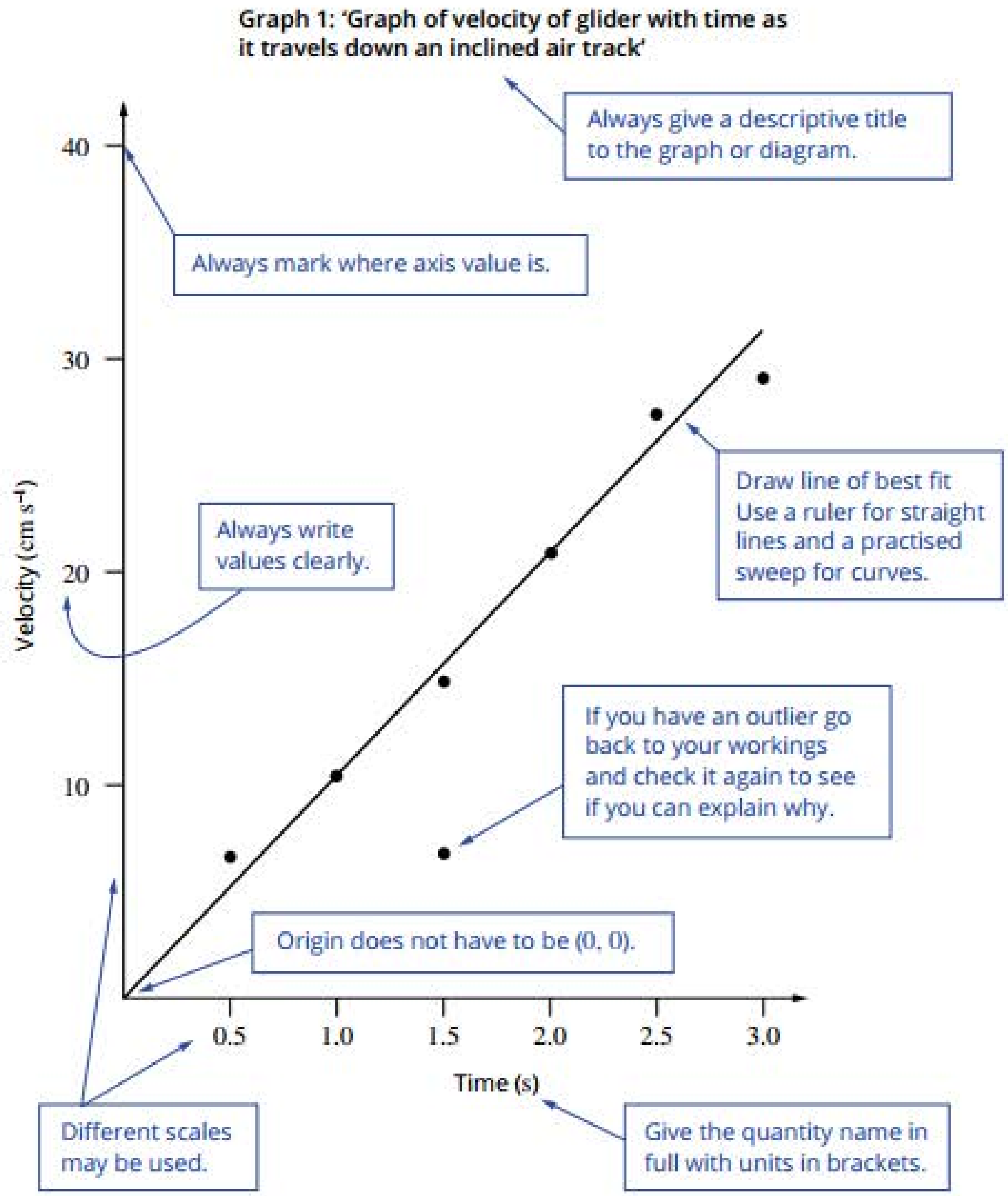


FIGURE 10.2.4 A graph shows the relationship between two variables.

Line graphs

Line graphs are a good way of representing continuous quantitative data. In a line graph, the values are plotted as a series of points with error bars on the graph. There are two ways of joining these points:

- A line can be ruled from each point to the next (Figure 10.2.5(a)). It shows the overall trend; it is not meant to predict the value of the points between the plotted data. These graphs are seldom used in physics.
- The points can be joined with a single smooth straight or curved line (Figure 10.2.5(b)). This creates a trend line, also known as a line of best fit. The line of best fit does not have to pass through every point but should go through as many error bars as possible. It is used when there is an obvious trend between the variables. These graphs are most commonly used in physics.

Outliers

Sometimes when the data is collected, there may be one point that does not fit with the trend and is clearly an error. This is called an **outlier**. An outlier is often caused by a mistake made in measuring or recording data, or from a random error in the measuring equipment. If there is an outlier, include it on the graph, but ignore it when adding a line of best fit (as in Figure 10.2.4, where the point (1.5, 6) is an outlier).

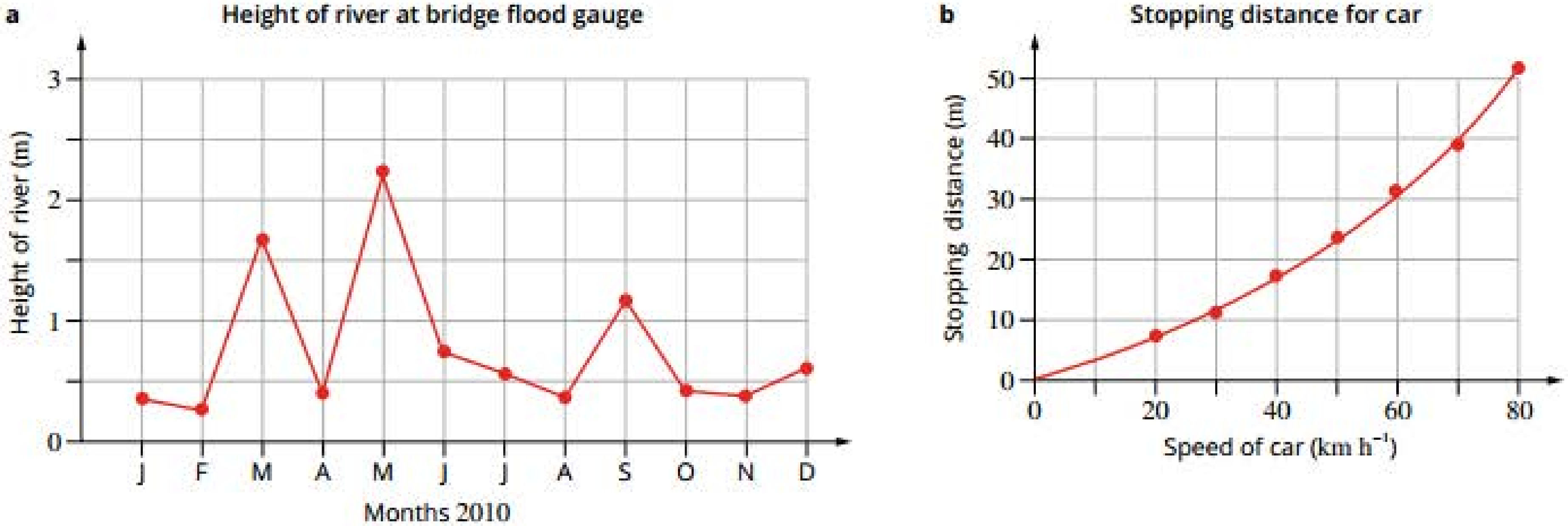


FIGURE 10.2.5 (a) The data in the graph is joined from point to point. (b) The data in the graph is joined with a line of best fit, which shows the general trend.

10.2 Review

SUMMARY

- It is essential that during the investigation, the following are recorded in the logbook:
  - all quantitative data collected
  - the methods used to collect the data
  - any incident, feature or unexpected event that may have affected the quality or validity of the data.
- A systematic error is an error that is consistent and will occur again if the investigation is repeated in the same way. Systematic errors are usually a result of instruments that are not calibrated correctly or methods that are flawed.
- Random errors occur in an unpredictable manner and are generally small. A random error could be, for example, the result of a researcher reading the same result correctly one time and incorrectly another time.
- The number of significant figures used depends on the scale of the instrument used. It is important to record data to the number of significant figures available from the equipment or observation.
- The simplest form of a table is a five-column format in which the first column contains the independent variable (the one being changed), the second to fourth columns contain the trials of the dependent variable (the one that may change in response to a change in the independent variable) and the final column contains the average of the trials.
- When there is a range of measurements of a particular value, the average must be accompanied by the uncertainty.
- General rules to follow when making a graph include the following:
  - Keep the graph simple and uncluttered.
  - Use a descriptive title that contains the independent and dependent variables.
  - Represent the independent variable on the *x*-axis and the dependent variable on the *y*-axis.
  - Make axes proportionate to the data.
  - Clearly label axes with both the variable and the unit in which it is measured.
  - Include error bars showing the uncertainty of each point. The error bar should extend above and below the plotted point equal to the uncertainty in the dependent variable and to the left and right of the plotted point equal to the uncertainty of the independent variable.

KEY QUESTIONS

1 The masses of 1 cm^3 cubes of potato were recorded and the cubes placed in distilled water. After 60 minutes, the cubes were weighed again and the difference in mass was calculated. What type of error is involved:

a if the electronic scales only measured in 1 g increments?

b if the electronic scales were affected briefly by a power surge?

2 If using the quantities mass = 7.50 kg and speed = 1.4 m s^{-1} in a calculation, what would be the appropriate number of significant figures in the answer?

3 For the data set 21, 28, 19, 19, 25, 24, determine:

a the mean

b the mode

c the median

d the uncertainty in the mean.

4 Plot the following data set with error bars, assigning each variable to the appropriate axis on the graph.

Current ± 0.01 (A)	Voltage ± 0.01 (V)
0.06	2.07
0.05	1.56
0.04	1.24
0.03	0.93
0.02	0.63

5 How can the general pattern (trend) of the data set in Question 4 be represented once the points are plotted?

10.3 Discussing investigations and drawing evidence-based conclusions

Now that the chosen topic has been thoroughly researched and the investigation has been conducted and data collected, it is time to draw it all together. The final part of the investigation involves summarising the findings in an objective, clear and concise manner.

FIGURE 10.3.1 To discuss and conclude your investigation, use the raw and processed data.

EXPLAINING RESULTS IN THE DISCUSSION

The discussion is the part of the investigation where the evaluation and explanation of the investigation methods and results takes place. It is the interpretation of what the results mean.

The key sections of the discussion are:

- analysing and evaluating data
- evaluating the investigative method
- explaining the link between the investigation findings and the relevant physics concepts.

When writing the discussion, consider the message you want to convey to the audience. Statements should be clear and concise. At the conclusion of the discussion, the audience must have a clear idea of the context, results and implications of the investigation.

ANALYSING AND EVALUATING DATA

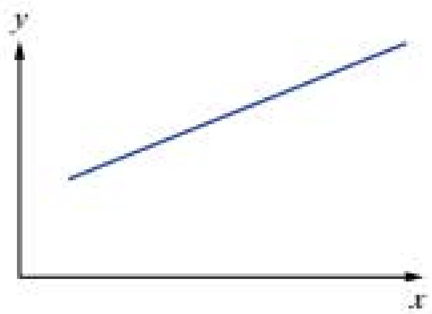
In the discussion, the findings of the investigation need to be analysed and interpreted.

- State whether a pattern, trend or relationship was observed between the independent and dependent variables. Describe what kind of pattern it was and specify under what conditions it was observed.
- Were there discrepancies, deviations or anomalies in the data? If so, these should be acknowledged and explained.
- Identify any limitations in the data you have collected. Perhaps a larger sample or further variations in the independent variable would lead to a stronger conclusion.

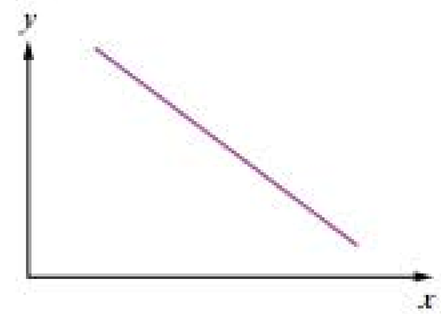
Trends in line graphs

Graphs are drawn to show the relationship, or trend, between two variables, as shown in Figure 10.3.2.

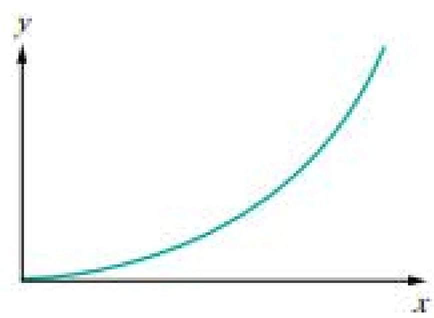
- Variables that change in linear or direct proportion to each other produce a straight, sloping trend line.
- Variables that change exponentially in proportion to each other produce a curved trend line.
- When there is an inverse relationship, one variable increases as the other variable decreases.
- When there is no relationship between two variables, one variable will not change even if the other changes.



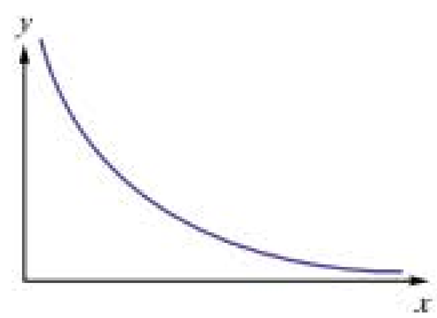
Direct or linear proportional relationship
- Variables change at the same rate (graph line is straight, slope is constant).
- Positive relationship—as x increases, y increases.



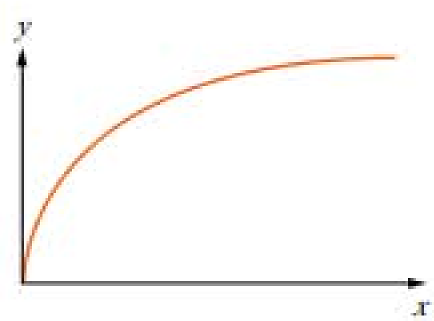
Inverse direct or linear proportional relationship
- Variables change at the same rate (graph line is straight, slope is constant).
- Negative relationship—as x increases, y decreases.



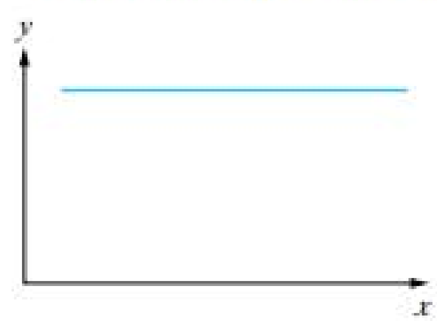
Exponential relationship
- As x increases, y increases slowly, then more rapidly.



Inverse exponential relationship
- As x increases, y decreases rapidly, then more slowly, until a minimum y value is reached.



Exponential rise, then levels off or plateaus (stops rising)
- As x increases, y increases rapidly at first, then slows, then finally does not increase at all—y reaches a maximum value.



No relationship between x and y
- As x increases, y remains the same.

FIGURE 10.3.2 Various relationships can exist between two variables.

Remember that the results may be unexpected. This does not make the investigation a failure. However, the findings must be related to the hypothesis, aims and method.

EVALUATING THE METHOD

It is important to discuss the limitations of the investigation method. Evaluate the method and identify any issues that could have affected the validity, accuracy, precision or reliability of the data. Sources of errors and uncertainty must also be stated in the discussion.

Once any limitations or problems in the methodology have been identified, recommend improvements on how the investigation could be conducted if repeated; for example, suggest how bias could be minimised or eliminated.

Bias

Bias may occur in any part of the investigation method, including sampling and measurements.

Bias is a form of systematic error resulting from the researcher's personal preferences or motivations. There are many types of bias, including:

- poor definitions of both concepts and variables (for example, classifying cricket pitch surfaces as slow or fast without defining 'slow' and 'fast')
- incorrect assumptions (for example, that footwear type, model and manufacturer do not affect ground reaction forces, and as a result failing to control this variable during an investigation on slip risk on different indoor and outdoor surfaces)
- errors in the investigation design and methodology (for example, taking a sample of a particular group of athletes that includes one gender more than the other in the group).

Some biases cannot be eliminated, but should at least be addressed in the discussion.

Accuracy and precision

In the discussion, evaluate the degree of accuracy and precision of the measurements for each variable of the hypothesis. Comment on the uncertainties obtained.

When relevant, compare the chosen method with any other methods that might have been selected, evaluating the advantages and disadvantages of the selected method and the effect on the results.

Reliability

When discussing the results, indicate the range of the data obtained from replicates. Explain how the sample size was selected. Larger samples are usually more reliable, but time and resources might have been scarce. Discuss whether the results of the investigation have been limited by the sample size.

The control group is important to the reliability of the investigation. A control group helps determine if a variable that should have been controlled has been overlooked and may explain any unexpected results.

Error

Discuss any source of systematic or random error and suggest ways of improving the investigation.

DISCUSSING RELEVANT PHYSICS CONCEPTS

To make the investigation more meaningful, it should be explained within the right context; i.e. using related physics ideas, concepts, theories, and models. Within this context, explain the basis for the hypothesis.

FIGURE 10.3.3 Honest evaluation and reflection play important roles in analysing methodology.

For example, if studying the impact of temperature on linear strain of a material (e.g. a rubber band), some of the contextual information to include in the discussion could be:

- the definition of linear strain
- the functions of linear strain
- the relationship between linear strain and temperature
- definitions of material behaviour (such as plastic and elastic)
- factors known to affect linear strain
- existing knowledge on the role of temperature on linear strain
- ranges of temperatures investigated and the reason these temperatures were chosen
- materials studied and the reasons for this choice
- methods of measuring the linear strain of a material.

Relating your findings to a physics concept

Once a context is established, you can use this as a framework in which to discuss whether the data supported or refuted the hypothesis. Ask questions such as:

- Was the hypothesis supported?
- Has the research question been fully answered? (If not, give an explanation of why this is so and suggest what could be done to either improve or complement the investigation.)
- Do the results contradict the hypothesis? If so, why? (The explanation must be plausible and must be based on the results and previous evidence.)

Providing a theoretical context also enables comparison of the results with existing relevant research and knowledge. After identifying the major findings of the investigation, ask questions such as:

- How does the data fit with the literature?
- Does the data contradict the literature?
- Do the findings fill a gap in the literature?
- Do the findings lead to further questions?
- Can the findings be extended to another situation?

Be sure to discuss the broader implications of the findings. Implications are the bigger picture. Outlining them for the audience is an important part of the investigation. Ask questions such as:

- Do the findings contribute to or impact on the existing literature and knowledge of the topic?
- Are there any practical applications for the findings?

DRAWING EVIDENCE-BASED CONCLUSIONS

A conclusion is usually a paragraph that links the collected evidence to the hypothesis and provides a justified response to the research question.

Indicate whether the hypothesis was supported or refuted and the evidence on which this is based (that is, the results). Do not provide irrelevant information. Only refer to the specifics of the hypothesis and the research question and do not make generalisations.

Read the examples of conclusions for the following hypothesis and research question.

Hypothesis: An increase in temperature will cause an increase in linear deformation (change in length) before failure.

- Poor response to the hypothesis: Linear deformation has value y_1 at temperature 1 and value y_2 at temperature 2.
- Better response to the hypothesis: An increase in temperature from 1 to 2 produces an increase in linear deformation of z in the rubber band.

Research question: Does temperature affect the maximum linear deformation the material can withstand?

- Poor response to the research question: The results show that temperature does affect the maximum deformation of a material.
- Better response to the research question: Analysis of the results of the effect of an increase in temperature from 1 to 2 in the rubber band supports current knowledge regarding the effect that an increase in temperature has on increasing maximum linear deformation.

REFERENCES AND ACKNOWLEDGEMENTS

All the quotations, documents, publications and ideas used in the investigation need to be acknowledged in the 'references and acknowledgments' section to avoid plagiarism and to ensure authors are credited for their work. References and acknowledgements also give credibility to the study and allow the audience to locate information sources should they wish to study the topic further.

When referencing a book, include, in this order:

- author's surname and initials
- date of publication
- title
- publisher's name
- place of publication.

For example: Moran G. et al. (2017), *Pearson Physics 11*, Pearson Education, Melbourne, Victoria.

When referencing a website, include, in this order:

- author's surname and initials, or name of organisation, or title
- year website was written or last revised
- title of webpage
- date website was accessed
- website address.

For example: Wheeling Jesuit University/Center for Educational Technologies (2013), *NASA Physics Online Course: Forces and Motion*, accessed 16 June 2015, from http://nasaphysics.cet.edu/forces-and-motion.html.

10.3 Review

SUMMARY

- The discussion is the part of the investigation where the evaluation and explanation of the investigation methods and results takes place. It is the interpretation of what the results mean.
- In the discussion, the findings of the investigation need to be analysed and interpreted.
  - State whether a pattern, trend or relationship was observed between the independent and dependent variables. Describe what kind of pattern it was and specify under what conditions it was observed.
  - Were there discrepancies, deviations or anomalies in the data? If so, these should be acknowledged and explained.
  - Identify any limitations in the data collected. Perhaps a larger sample or further variations in the independent variable would lead to a stronger conclusion.
- It is important to discuss the limitations of the investigation method. Evaluate the method and identify any issues that could have affected the validity, accuracy, precision or reliability of the data. Sources of errors and uncertainty must also be stated in the discussion, and suggestions could be given as to how to reduce these errors.
- When discussing the results, indicate the range of the data obtained from replicates. Explain how the sample size was selected. Larger samples are usually more reliable, but time and resources are likely to have been scarce. Discuss whether the results of the investigation have been limited by the sample size.
- To make the investigation more meaningful, it should be explained within the right context, meaning the related physics ideas, concepts, theories and models. Within this context, explain the basis for the hypothesis.
- Indicate whether the hypothesis was supported or refuted and on what evidence this is based (that is, the results). Do not provide irrelevant information or make generalisations.

KEY QUESTIONS

1. What relationship between the variables is indicated by a sloping linear graph?
2. What relationship exists if one variable decreases as the other increases?
3. What relationship exists if both variables increase or both decrease at the same rate?
4. What might cause a sample size to be limited in an investigation?
5. Consider this investigation hypothesis: An increase in the current passing through a single resistor in an electric circuit will cause an increase in the voltage drop across the resistor.
   Improve this response to the hypothesis:
   When the current was 0.03 A, the voltage was 0.93 V and when the current was 0.05 A, the voltage was 1.81 V.

Chapter review

10

KEY TERMS

controlled variable
dependent variable
independent variable
mean
median
mode
outlier
personal protective equipment (PPE)
qualitative variable
quantitative variable
random error
raw data
reliability
significant figures
systematic error
uncertainty
validity
variable

1 What is a hypothesis and what form does it take?

2 Consider the hypothesis provided below. What are the dependent, independent and controlled variables?
Hypothesis: Releasing an arrow in archery at an angle greater or smaller than 45 degrees will result in a shorter flight displacement (range).

3 What is the dependent variable in each of the following hypotheses?
   a If you push an object with a fixed mass (e.g. shot put) with a larger force, then the acceleration of that object will be greater.
   b The vertical acceleration of a falling object is constant.
   c A springboard diver rotates faster when in a tucked position than when in a stretched (layout) position.

4 List these types of hazard controls from the most effective to the least effective.
substitution, personal protective equipment, engineering controls, administrative controls, elimination, isolation

5 The speed of a toy car rolling down an inclined plane was measured 6 times. The measurements obtained (in cm s^{-1}) were 7.0, 6.5, 6.8, 7.2, 6.5, 6.5.
What is the uncertainty of the average of these values?

6 Which of the statistical measurements of mean, mode and median is most affected by an outlier?

7 What relationship between variables is indicated by a curved trend line?

8 If you hypothesise that impact force is directly proportional to drop height, what would you expect a graph of the data to look like?

9 What is meant by the 'limitations' of the investigation method?

10 What is 'bias' in an investigation?

REVIEW QUESTIONS

Section 1: Short response

1 Describe what an electromagnetic wave is, and briefly summarise the relationship between energy, wavelength and frequency. Name one way in which electromagnetic waves differ from mechanical waves.

2 Physicists often use the expression 'wave–particle duality' to express how light can sometimes be observed to behave like a particle, and how electrons can sometimes be observed to behave like a wave.

 a Describe briefly one experiment in which light is shown to behave as a particle. Explain how this evidence supports a particle model of light.

 b Describe briefly one behaviour of electrons in which they are shown to have wave-like properties. Explain how this evidence supports a wave model of matter.

3 A physicist hoping to determine the speed of light detects two different quanta of light using different instruments. She measures one packet to have a wavelength of 220 nm and knows the other packet to be twice as energetic, with energy of 11.24 eV. Using this information, what is the speed of light?

4 Describe three experimental results associated with the photoelectric effect that cannot be explained by the wave model of light.

5 a State Einstein's two postulates of special relativity.

 b Two space ships are travelling in the same direction at 0.800c and 0.950c respectively, according to an external observer. How fast (in terms of c) is the faster ship moving, according to an observer on the slower ship?

6 A person moving parallel to a 1.50 m ruler at a speed of $1.35 \times 10^8\,\mathrm{m\,s^{-1}}$ will see a difference in its length compared to a person at rest relative to the ruler. Will the moving person see the ruler as being longer or shorter? What length will they observe it to be?

7 A neutron in an evacuated container decays to produce a proton and two other particles. The proton is then attracted to an electron and becomes the nucleus of a hydrogen atom. The atom then slowly drifts to the base of the container. Order the forces that have been involved in the sequence of events described from first to last. Any that were not involved should go last.

 gravity, strong nuclear, weak nuclear, electromagnetic

8 In the early 20th century, astronomers were still trying to figure out if galaxies they could see in the sky were indeed part of the Milky Way or separate from us. One spiral galaxy, the Sombrero Galaxy (NGC 4594), is located 9.55 Mpc away from Earth. In 1912, the astronomer Vesto Slipher measured the redshift of the Sombrero Galaxy and determined it was moving away from us at a rate of $1000\,\mathrm{km\,s^{-1}}$, implying that this galaxy was outside of our galaxy.

 a In measuring the light from the galaxy, would Vesto Slipher have observed a longer or shorter wavelength light than its true wavelength? Explain your answer.

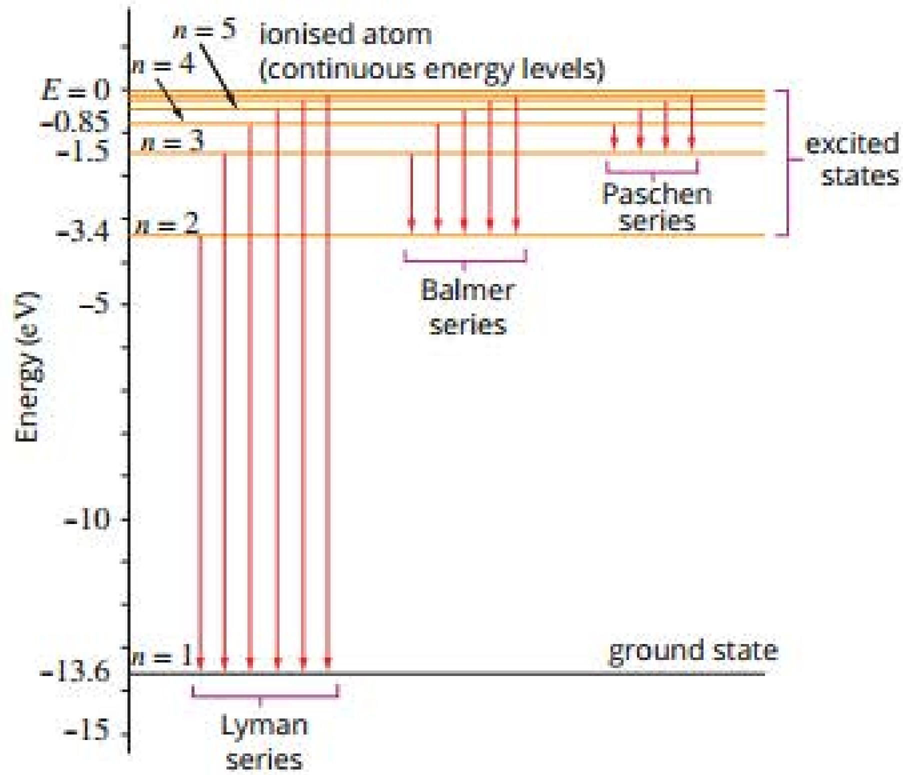
 b Using the above information, calculate the Hubble constant.

Section 2: Problem solving

9 In amateur astrophotography, astronomers must often utilise a number of techniques for overcoming the issue of light pollution. One such technique is the use of narrowband filters to limit the amount of light that hits the sensor of the camera. Typically, these filters select a 3 nm wavelength range in the red, green or blue parts of the optical spectrum and allow only these wavelengths through to the sensor. Since the wavelength of most light pollution is different from that allowed by these filters, this method permits astrophotographers to take deep-space images without the interference of external light sources. One commonly used narrowband filter is that of hydrogen alpha (H-alpha), which corresponds to a particular wavelength of the Balmer series for hydrogen. H-alpha light is important to astronomers because it is emitted by many emission nebulae.

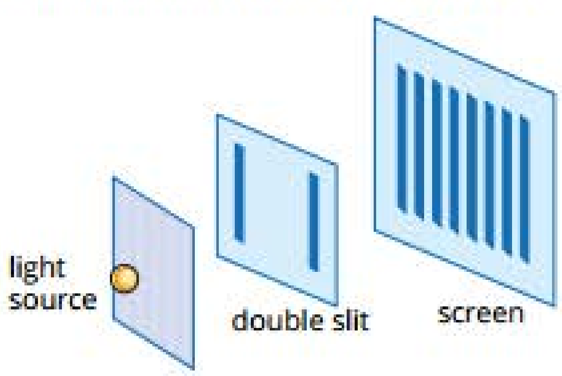
 a Explain the mechanism by which atoms have characteristic emission spectra.

b The energy level diagram for hydrogen is shown below. Given that the wavelength of H-alpha is 656 nm, determine which energy level transition is being made by the electrons.



c Calculate the speed at which an electron would have to be travelling in order to make it through one of these filters.

10 Some students set out to replicate Young's famous double-slit experiment. They set up their apparatus as shown below, and observe a banding pattern on their screen with alternating nodes and antinodes.



a Explain in detail how Young's double-slit experiment supports the wave model of light rather than the particle model.

b The students determined that particular adjacent dark bands on the interference pattern (e.g. the third dark fringe and the fourth dark fringe) differ in their distance from one of the slits by only 500 nm. Determine the wavelength of the monochromatic light being used.

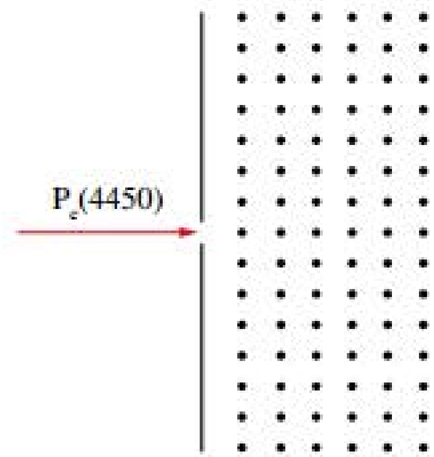
c What is the effect on the observed interference pattern of covering one of the slits?

d What would happen to the spacing of fringes if the slits were moved closer to one another?

11 The elementary particles of the Standard Model are shown on page 4 of the formula and data booklet.

a Two exotic pentaquark states were observed in 2015 by the LHCb collaboration and named $P_c(4450)$ and $P_c(4380)$. Both states are composed of five quarks: two up quarks, a down quark, a charm quark, and an anti-charm quark. Using this information, determine the charge on this pentaquark state.

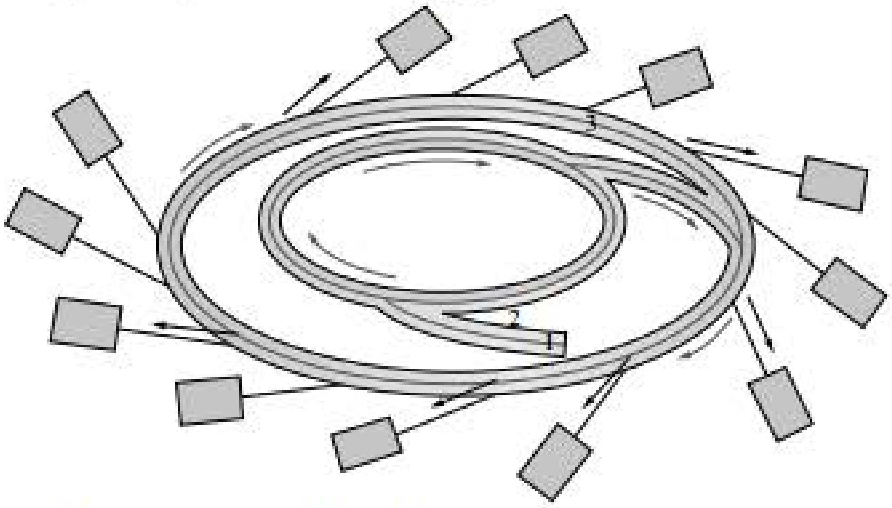
b Copy the diagram below into your notebook and sketch the approximate shape of the particle's trajectory through the magnetic field.



c The Large Hadron Collider uses magnets that are extraordinarily strong—up to 8.33 tesla (approximately 200 000 times stronger than Earth's magnetic field). Assume that immediately after its creation, the pentaquark moves perpendicularly through the field at a velocity of 99.9% of the speed of light. The $P_c(4450)$ has a mass of approximately 7.92×10^{-27} kg. Calculate the radius of the particle's circular path.

d Which of the four fundamental forces would be the most likely to be holding the five quarks together?

12 The Australia Synchrotron is an electron accelerator based in Melbourne, Victoria, capable of bringing electrons extremely close to the speed of light. When travelling in a circle at such high speeds, electrons emit high-intensity radiation across the electromagnetic spectrum, which can then be used in detailed studies at a microscopic level. Its applications are wide reaching, from producing close-up images of concrete and proteins, to medical imaging.



The electrons go through three stages:

1 Electrons are produced by an electron gun and then subject to a high-frequency voltage signal in order to separate them into bunches. A high-voltage potential difference then initially accelerates the electrons as they are shot out.

2 The linear accelerator then provides another potential difference in order to accelerate the electrons even further over the course of about 10 m. After just the first metre of acceleration, the electrons are already travelling at 99.99% of the speed of light.

3 The accelerated electrons are then led through the use of strong magnetic fields into the storage ring. It is here that they spend most of their lifetime, travelling around a ring 216 metres in circumference. The ring typically stores about 200 mA of current for a period of 20 hours.

a The first acceleration of electrons involves subjecting them to a potential difference of 90 keV. Taking into account relativistic effects, determine the speed of the electrons at this stage given that the total energy is given by the equations:

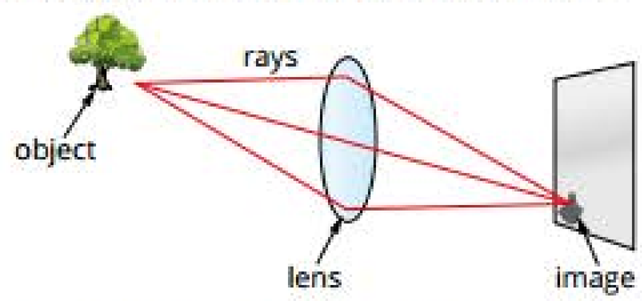
$$E_{tot} = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{and} \quad E_{tot} = E_k + m_0c^2$$

b Similarly, calculate the potential difference in eV to which the electrons are subject over the total 10 m of the linear accelerator (under the simplified assumption that the potential difference increases linearly).

Section 3: Comprehension

13 An example of the culmination of many of the achievements of physics in the 20th century is the development of the modern DSLR camera. In being able to explain the behaviour of both light and electrons, physicists have developed different methods to manipulate these into creating and storing digital images.

The modern DSLR camera is an extraordinarily complicated device. There are two main components key to its function: the lens, and the sensor. Lenses are optical objects that use principles of refraction in order to collect light rays from an object into a single specified point on the sensor. The sensor of a DSLR then detects this optical signal and converts it into a digital output. Although there are many different types of lens systems, we will focus only on a simple single lens as shown. A schematic of how light rays behave through the lens (called a ray-tracing diagram) is shown below.



a i A common problem faced while imaging is the issue of chromatic aberration. Chromatic aberration is the separation of different wavelengths of light from a source object as they focus onto the sensor. Using your knowledge of the refraction of light, explain why this phenomenon would be observed when focusing an image through a lens.

ii In what order would the colours red, violet and green separate, from closest to furthest from the centre. Why?

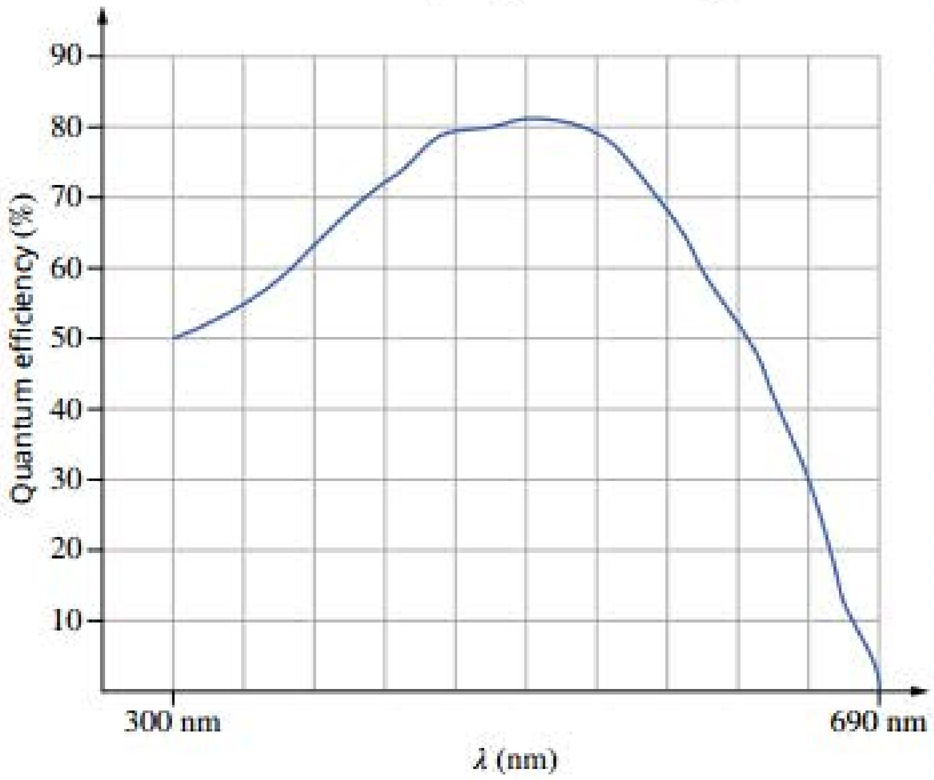
b Given that the refractive index of glass is 1.62, calculate the angle at which light will reflect internally instead of refracting through the lens.

The standard imaging sensor for a DSLR camera is a CMOS (complementary metal-oxide-semiconductor). The sensor is comprised of a large array of small photodiodes called 'pixels'. Photons are incident upon a pixel of the sensor, creating a current that is detected by the camera's computing system. The individual pixel signals are then combined to create the image that you see.

c Suggest and explain a mechanism you have learnt by which the sensor is able to convert the optical image into a digital signal.

d Why would a standard DSLR camera be unaffected by the local radio waves while taking a photo, despite the fact that these are light waves?

e A useful quantity to define for imaging sensors is the quantum efficiency. It is defined as the percentage of photons converted into electrons when incident on an imaging sensor, demonstrating the responsiveness to certain wavelengths of light. Below is a graph of quantum efficiency vs light wavelength for a particular imaging sensor. Using this information, estimate the work function, in electron volts, for the sensor's material. Explain your reasoning.



APPENDIX A SI units

THE STANDARD UNITS OF MEASUREMENT

The accurate and easy measurement of quantities is essential in both everyday life and for scientific investigation. Over the centuries, many different systems of measuring physical quantities have been developed. For example, length can be measured in chains, fathoms, furlongs, yards, feet, rods and microns. Some units were based on parts of the body. The cubit was defined as the distance from the elbow to the fingertip, and so the amount of cloth that you obtained from a tailor depended on the physical size of the person selling it to you.

The metric system was established by the French Academy of Science at the time of the French Revolution (1789–1815) and is now used in most countries. This system includes units such as the metre, litre and kilogram. Countries of the British Empire adopted the British Imperial system of the mile, gallon and pound. These two systems developed independently, and their dual existence created problems in areas such as trade and scientific research. In 1960, an international committee set standard units for fundamental physical quantities. This system was an adaptation of the metric system and is known as the *Système Internationale d'Unités* (International System of Units) or SI system of units.

TABLE A.1 The SI units of the seven fundamental quantities whose basic value is defined to a high degree of accuracy

Fundamental quantity	SI unit	SI unit symbol
mass	kilogram	kg
length	metre	m
time	second	s
electric current	ampere	A
temperature	kelvin	K
luminous intensity	candela	cd
amount of substance	mole	mol

Mass

The kilogram was originally defined as the mass of 1 L of water at 4°C. This is still approximately correct, but a far more precise definition is now used. Since 1897 the measurement standard for the kilogram has been a cylindrical block of platinum–iridium alloy kept at the International Bureau of Weights and Measures in France. Australia has a copy of this standard mass at the CSIRO Division of Applied Physics in Sydney. At times it is returned to France to ensure that the mass remains accurate.

Length

The metre was originally defined in 1792 as one ten-millionth of the distance from the equator to the North Pole (approximately 10 000 km). This definition has changed a number of times since. In 1983, to give a more accurate value, the metre was redefined as the distance that light in a vacuum travels in $\frac{1}{299\,792\,458}$ second. This standard can be reproduced all over the world, as light travels at a constant speed in a vacuum.

PHYSICSFILE

Metric system

The metric system was originally developed in France and is known as the *Système Internationale* (SI). It was adopted in France in 1840 as the official system of units, although it had been developing in that country since 1545. It has remained in use ever since and has gradually been adopted by most other countries. It has been modified a little over the years and now, in Australia, we use SI units that have been standardised by the International Standards Organisation (ISO) since the 1960s. Some countries such as France, Italy and Spain use an earlier form of the metric system that is slightly different. The USA still measures almost everything in the old imperial units such as pounds for mass and feet for distance but, even there, scientists use the SI system of units. There are two major advantages of using the metric system. It is easier to use than other systems in that derived units are straightforward and various sizes of units are created using multiples of ten. The other very big advantage is the international nature of the standards and units. All units are standardised, making comparisons straightforward.

Time

Up to 1967, time had always been based on the apparent motion of the heavens. The second was once defined in terms of the motion of the Sun. Until 1960, one second was defined as $\frac{1}{60}$ of $\frac{1}{60}$ of $\frac{1}{24}$ of an average day in 1900. This reflected the rate of the Earth's rotation on its axis; however, its rotation is not quite uniform. In 1967, a more accurate definition was adopted—one not based on the motion of the Earth. One second is now defined as the time required for a caesium-133 atom to undergo 9 162 631 770 vibrations. These vibrations are stimulated by an electric current and are extremely stable, allowing this standard to be reproduced all over the world.

DERIVED UNITS

As well as the seven fundamental quantities, a wide variety of other physical quantities can be measured. You may have encountered some of these, such as frequency, velocity, energy and density, already. A derived quantity is defined in terms of the fundamental quantities. For example, the SI unit for area is square metres (m^2).

TABLE A.2 Some derived SI quantities and their units

Quantity	SI unit	SI unit symbol	Equivalent unit
velocity	metres per second	$m\,s^{-1}$	—
acceleration	metres per second per second	$m\,s^{-2}$	—
frequency	hertz	Hz	s^{-1}
force	newton	N	$kg\,m\,s^{-2}$
energy/work	joule	J	$kg\,m^2\,s^{-2}$

APPENDIX B Understanding measurement

MEASUREMENT AND UNITS

In every area of physics we have attempted to quantify the phenomena we study. In practical demonstrations and investigations, we generally make measurements and process those measurements in order to come to some conclusions. Scientists have a number of conventional ways of interpreting and analysing data from their investigations. There are also conventional ways of writing numerical measurements and their units.

Correct use of unit symbols

The correct use of unit symbols removes ambiguity, as symbols are recognised internationally. The symbols for units are not abbreviations and should not be followed by a full stop unless they are at the end of a sentence.

Upper-case letters are not used for the names of any physical quantities of units. For example, we write newton for the unit of force, while we write Newton if referring to someone with that name. Upper-case letters are only used for the *symbols* of the units that are named after people. For example, the unit of energy is the joule and the symbol is J. The joule is named after James Joule, who was famous for studies of energy conversions. The exception to this rule is 'L' for litre. We do this because a lower-case 'l' looks like the numeral '1'. The unit of distance is metre and the symbol is m. The metre is not named after a person.

The product of a number of units is shown by separating the symbol for each unit with a dot or a space. Most teachers prefer a space but a dot is perfectly correct. The division or ratio of two or more units can be shown in fraction form, using a slash, or using negative indices. Most teachers prefer negative indices. Prefixes should not be separated by a space.

TABLE B.1 Some examples of the use of symbols for derived units

Preferred	Correct also	Wrong
m s^{-2}	m.s^{-2} m/s^2	ms^{-2}
kW h	kW.h	kWh k Wh
kg m^{-3}	kg.m^{-3} kg/m^3	kgm^{-3}
µm		µ m
N m	N.m	Nm

Units named after people can take the plural form by adding an 's' when used with numbers greater than one. Never do this with the unit symbols. It is acceptable to say 'two newtons' but wrong to write 2 Ns. It is also acceptable to say 'two newton'.

Numbers and symbols should not be mixed with words for units and numbers. For example, twenty metres and 20 m are correct while 20 metres and twenty m are incorrect.

FIGURE B.1 A scientific calculator.

Scientific notation

To overcome confusion or ambiguity, measurements are often written in scientific notation. Quantities are written as a number between one and ten and then multiplied by an appropriate power of ten. Note that 'scientific notation', 'standard notation' and 'standard form' all have the same meaning.

Examples of some measurements written in scientific notation are:

$0.054\,\text{m} = 5.4 \times 10^{-2}\,\text{m}$

$245.7\,\text{J} = 2.457 \times 10^{2}\,\text{J}$

$2080\,\text{N} = 2.080 \times 10^{3}\,\text{N}$ or $2.08 \times 10^{3}\,\text{N}$

You should be routinely using scientific notation to express numbers. This also involves learning to use your calculator intelligently. Scientific and graphics calculators can be put into a mode whereby all numbers are displayed in scientific notation. It is useful when doing calculations to use this mode rather than frequently attempting to convert to scientific notation by counting digits on the calculator display. It is quite acceptable to write all numbers in scientific notation, although most people prefer not to use scientific notation when writing numbers between 0.1 and 1000.

An important reason for using scientific notation is that it removes ambiguity about the precision of some measurements. For example, a measurement recorded as 240 m could be a measurement to the nearest metre; that is, somewhere between 239.5 m and 240.5 m. It could also be a measurement to the nearest ten metres; that is, somewhere between 235 m and 245 m. Writing the measurement as 240 m does not indicate either case. If the measurement was taken to the nearest metre, it would be written in scientific notation as $2.40 \times 10^{2}\,\text{m}$. If it was taken to the nearest ten metres only, it would be written as $2.4 \times 10^{2}\,\text{m}$.

PREFIXES AND CONVERSION FACTORS

Conversion factors should be used carefully. You should be familiar with the prefixes and conversion factors in Table B.2. The most common mistake made with conversion factors is multiplying rather than dividing. Some simple strategies can save you this problem. Note that the table gives all conversions as a multiplying factor.

TABLE B.2 Prefixes and conversion factors

Multiplying factor		Prefix	Symbol
1 000 000 000 000	10^{12}	tera	T
1 000 000 000	10^{9}	giga	G
1 000 000	10^{6}	mega	M
1 000	10^{3}	kilo	k
0.01	10^{-2}	centi	c
0.001	10^{-3}	milli	m
0.000 001	10^{-6}	micro	µ
0.000 000 001	10^{-9}	nano	n
0.000 000 000 001	10^{-12}	pico	p

Do not put spaces between prefixes and unit symbols. It is important to give the symbol the correct case (upper or lower case). There is a big difference between 1 mm and 1 Mm.

There is no space between prefixes and unit symbols. For example, one-thousandth of an ampere is given the symbol mA. Writing it as m A is incorrect. The space would mean that the symbol is for a derived unit—a metre ampere.

Worked example B1

The diameter of a cylindrical piece of copper rod was measured at 24.8 mm with a vernier caliper. Its length was measured at 35 cm with a tape measure.
a Find the area of cross-section in m^2.
b Find the volume of the copper rod in m^3.
Answer
a The area of cross-section is πr^2. The radius is calculated by dividing the diameter by two. Hence the radius is 12.4 mm. To calculate the area in m^2, first halve the diameter and convert it to metres. The radius is $\frac{24.8}{2} = 12.4\,\text{mm} = 12.4 \times 10^{-3}\,\text{m}$. The radius is not written in scientific notation. This is not necessary. All you need to do is multiply by the appropriate factor. The conversion factor for mm to m is 10^{-3}. Just multiply by the conversion factor and don't bother to rewrite the result in scientific notation. This is because it is only going to be used in a calculation and is not a final result. The area of cross-section is $\pi r^2 = \pi(12.4 \times 10^{-3})^2 = 4.8 \times 10^{-4}\,\text{m}^2$.
b The volume is $\pi r^2 h$, where h is the length of the cylinder. The length is $35\,\text{cm} = 35 \times 10^{-2}\,\text{m}$. Hence the volume is $\pi(12.4 \times 10^{-3})^2(35 \times 10^{-2}) = 1.7 \times 10^{-4}\,\text{m}^2$.

Worked example B2

a A car is traveling at 110 km h^{-1}. How fast is this in m s^{-1}?
b Convert 35 miles per hour to metres per second. A mile is approximately 1600 m.
Answer
a 110 km h^{-1} is 110×10^3 metres per 3600 s. $\frac{110 \times 10^3}{3600} = 30.6$ Hence $110\,\text{km h}^{-1} = 30.6\,\text{m s}^{-1}$.
b 35 miles per hour is 35×1600 metres per 3600 s. $\frac{35 \times 1600}{3600} = 15.6$ Hence 35 miles per hour $= 16\,\text{m s}^{-1}$.

DATA

Physicists and physics students collect, analyse and interpret experimental data. In fact, you will do this when you conduct your practical investigation. Working with data requires a good understanding of the meaning and limitations of measurement.

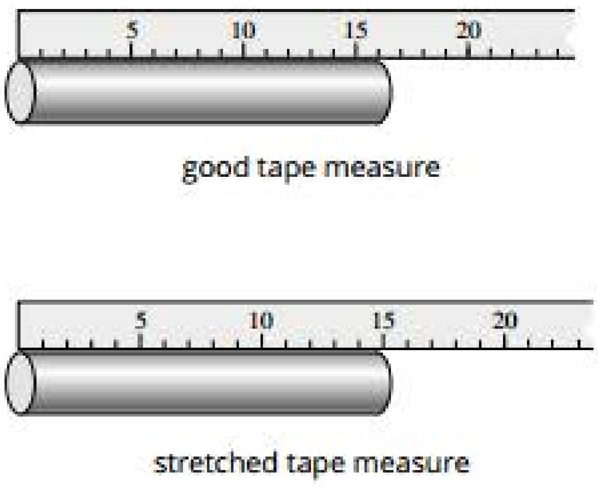


FIGURE B.2 The diagram shows that a correctly manufactured tape measure correctly measures the cylinder to be 16 cm long while the stretched tape measure gives a wrong measurement of 15 cm. The stretched tape measure is inaccurate.

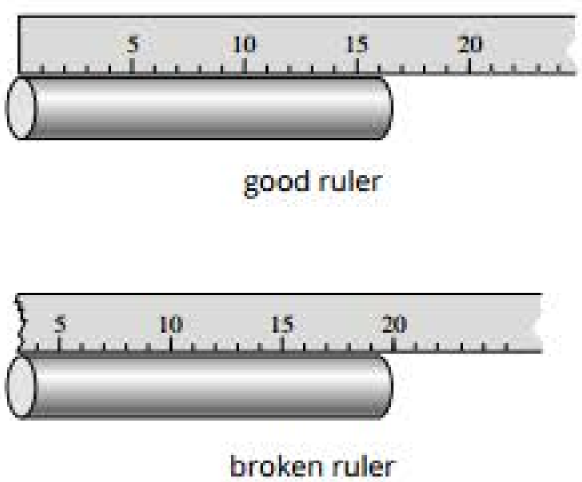


FIGURE B.3 The diagram shows that an undamaged ruler correctly measures the cylinder to be 16 cm long while the broken ruler gives a wrong measurement of 19 cm. The broken ruler is inaccurate but equally as precise as the unbroken ruler.

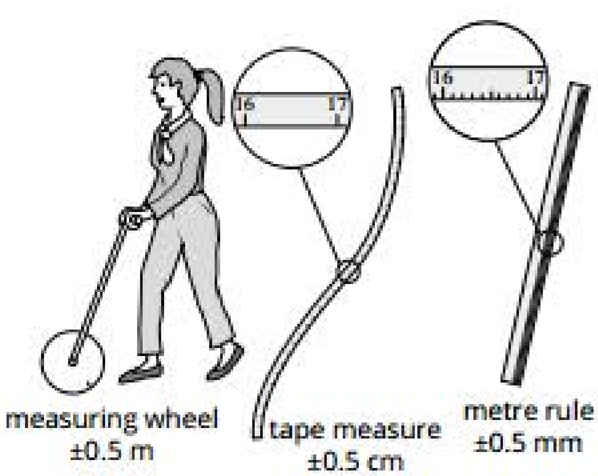


FIGURE B.4 The measuring wheel has low precision and only measures to the nearest metre. It has an uncertainty of 0.5 m. The tape measure has more precision and has an uncertainty of 0.5 cm or 0.005 m. The metre rule has an uncertainty of 0.5 mm or 0.0005 m.

Accuracy and precision

Two very important aspects of any measurement are accuracy and precision. Accuracy and precision are not the same thing. The distinction between the two ideas is only hard to grasp because the two words are defined in a similar way in the dictionary. We often hear the words used together, and in general conversation they tend to be used interchangeably.

Instruments are said to be *accurate* if they truly reflect the quantity being measured. For example, if a tape measure is correctly manufactured it can be used to measure lengths accurately to the nearest centimetre.

Imagine that the tape measure is accidentally stretched during the manufacturing process, as shown in Figure B.2. It would still be used to measure length to the nearest centimetre but all measurements would be wrong. It would be inaccurate.

Suppose an accurate ruler had 3 cm snapped off the end, as shown in Figure B.3. It would now give readings all too large by 3 cm if no allowance were made for the missing piece. This ruler measure would be inaccurate.

In these two examples, the tape measure or ruler is used to measure to the nearest centimetre but is inaccurate. Inaccurate means just plain wrong. Instruments are said to be *precise* if they can differentiate between slightly different quantities. Precision refers to the fineness of the scale being used.

Consider the metre rule, the tape measure and the measuring wheel used to mark out sports fields. All three measure distance. All three can be accurate. The metre rule is more *precise* because it measures to the nearest millimetre, the tape measure has less precision due to measuring only to the nearest centimetre, while the wheel measures only to the nearest metre (Figure B.4).

The tape measure is a more precise instrument than the measuring wheel. Suppose two distances of 2673 mm and 2691 mm are being measured with these two instruments. Each distance would be measured as 3 m, to the nearest metre, by the wheel. They would be measured differently as 2.67 m and 2.69 m, to the nearest centimetre, by the tape measure. The tape measure is more precise because it has a finer scale. You might also say that it has greater resolution. The measuring wheel has such low precision that it can't be used to measure which of the two distances is greater or smaller. Measuring instruments with less precision give measurements that are less certain. The uncertainty in the measurement is due to a coarser scale. The measuring wheel gives less certain measurements than the tape measure even though both instruments may be equally accurate.

All measurements have some amount of *uncertainty*, due to the precision of the instrument that does the measuring. (Note that in Chapter 10 the uncertainty due to collecting a range of data was analysed. This section deals with uncertainty due to precision.) The uncertainty is generally one half of the finest scale division on the measuring instrument. The measuring wheel has an uncertainty of 0.5 m. The metre rule has an uncertainty of 0.5 mm. The tape measure has an uncertainty of 0.5 cm. An electronic balance set to measure grams to two decimal places has an uncertainty of 0.005 g.

Sometimes this uncertainty is referred to as error. It is not error, in that it is not a mistake or something wrong. All measuring instruments have limited precision and, in general, the uncertainty is half of the smallest scale division on the instrument.

The uncertainty is the measure of the precision of an instrument. It is not related to accuracy. A micrometer screw gauge, which measures length to the nearest one-hundredth of a millimetre and hence is very precise, may not be accurate. Usually they are, but if one has been badly manufactured or bent by being over-tightened repeatedly it most likely will be inaccurate. But its precision will still be ±0.000 005 m, or half of one-hundredth of a millimetre.

The uncertainty gives the range in which a measurement falls. If you measured the length of a stick with a metre rule then you would get a measurement 'plus or minus' half a millimetre.

Any stick between 127.5 mm and 128.5 mm long would be measured as 128 mm to the nearest millimetre (refer to Figure B.5). We would record this as 128 ± 0.5 mm.

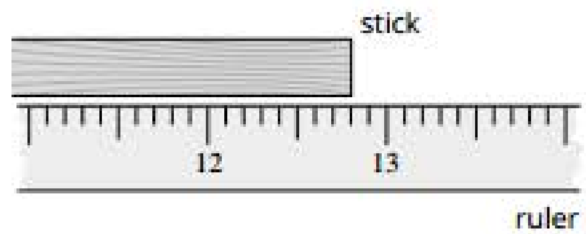


FIGURE B.5 A stick anywhere between 127.5 mm and 128.5 mm would be recorded as having a length of 128 mm if measured by a metre rule with a scale division of 1 mm. Conversely, a measurement recorded as 128 mm could be of an object of length anywhere between 127.5 mm and 128.5 mm.

When using an analogue scale, you might think that you can 'judge by eye' fractions of a scale division and hence get greater precision than half a scale division. You should be able to judge to the nearest half a scale division. You might think you can judge to the nearest tenth of a division. You can't. Research shows that despite the fact that people try to judge the spaces between scale divisions to better than half a division, as soon as this is done, inconsistent measurements are obtained. That is, different people get different measurements of the same thing.

The best judgement you can definitely claim is one half of a scale division. The uncertainty we will still assume, however, is a full half-scale division. Hence, you might measure another stick, one that has a length somewhere between 154 mm and 155 mm, as 154.5 ± 0.5 mm.

Of course, you don't have the option of adding an extra decimal place containing a 0 or a 5 if you are using a digital instrument.

The uncertainty can be recorded as the *absolute* uncertainty as we have done above. The absolute uncertainty is the actual uncertainty in the measurement. In this case it is 0.5 mm. It is often useful to write the uncertainty as a *percentage*: 0.5 mm is 0.32% of 154.5. Hence, the above length would be recorded as 154.5 mm ± 0.32%.

Percentage uncertainty is also called *relative* uncertainty. It is the size of the uncertainty relative to the size of the measured quantity.

PHYSICSFILE

Many people use the term 'error' to refer to uncertainty and many other things. The problem with referring to uncertainty as error is that it is not actually error. Things that are a normal consequence of the limitations of measuring instruments must happen, and are not mistakes. If they are not mistakes or 'something gone wrong' then it makes no sense to call them errors.

Errors are the factors that limit the accuracy of your results. For example, if you perform a calorimetry experiment and do not use a good enough insulator, you will get inaccurate results due to heat losses to the environment. This will contribute to the error in your measurement. Suppose you measured the refraction of light in glass but did not place the protractor in the correct place when measuring angles. This would also cause error.

Many different things can contribute to experimental error. Some are unavoidable. Some are factors in the design of the experiment. Good experimental design seeks to eliminate or at least minimise potential sources of error.

Never quote 'human error' as a source of error. Your data should be examined carefully and mistakes eliminated or at least ignored. So-called human errors, or lack of care, have no place in your experimental work. If you make mistakes then you should repeat the measurements.

Estimating the uncertainty in a result

An experiment or a measurement exercise is not complete until the uncertainties have been analysed. Chapter 10 explained how uncertainties were treated when a range of data had been collected and then averaged out. It is also important to explain how uncertainty due to the precision of instruments affects results.

The following three processes are used for estimating uncertainty in calculations due to the precision of instruments. They are demonstrated in Worked example B3.

- When adding or subtracting data, add the absolute uncertainties.
- When multiplying or dividing data, add the percentage uncertainties.
- When raising data to power n, multiply the percentage uncertainty by n.

In Worked example B3, the analysis of uncertainty reveals the *precision* of an experimental result.

Worked example B3

You might have measured the specific heat capacity of a metal. You could have calculated your result using:

$$c_{metal} = \frac{c_{water} m_{water} \Delta T_{water}}{m_{metal} \Delta T_{metal}}$$

Suppose you had the following data included in your table.

Quantity		Absolute uncertainty	Uncertainty (%)
c_{water}	$4180\,J\,kg^{-1}\,K^{-1}$	$5\,J\,kg^{-1}\,K^{-1}$	0.120
m_{water}	72.5×10^{-3} kg	0.05×10^{-3} kg	0.069
ΔT_{water}	5°C	1°C*	20
m_{metal}	87.3×10^{-3} kg	0.05×10^{-3} kg	0.057
ΔT_{metal}	72°C	1°C*	1.389

*Note that the ΔT values have an absolute uncertainty of 1°C because they are calculated by subtracting one temperature measurement from another.

You would calculate as follows:

$$\begin{aligned} c_{metal} &= 241\,J\,kg^{-1}\,K^{-1} \\ \text{Uncertainty (\%)} &= 0.120 + 0.069 + 20 + 0.057 + 1.389 \\ &= 21.6\% \end{aligned}$$

Hence, you would obtain the following result:

$$\begin{aligned} c_{metal} &= 241\,J\,kg^{-1}\,K^{-1} \pm 21.6\% \\ c_{metal} &= 241 \pm 52\,J\,kg^{-1}\,K^{-1} \end{aligned}$$

Once you have done all of this you can consider the relative success of your measurement exercise.

Your result is:

$$189\,J\,kg^{-1}\,K^{-1} \le c_{metal} \le 293\,J\,kg^{-1}\,K^{-1}$$

If measurements by other people, such as the constants published in data books, fall within this range then you can conclude that your experiment is consistent with established values. That is, within the precision of your technique, there are probably no significant errors although the final measurement is rather imprecise in this case. We might say that it is accurate within the limitations of the equipment.

PHYSICSFILE

In some classes, students are instructed to quote all results to two decimal places or to three significant figures. You should be able to see from Worked example B3 that these rules are not absolutely correct when applied to real data. For ordinary calculations in assignments, tests and examinations, you might just give your answers to three figures.

If a calculation is done in several stages then you should not round off any intermediate results. This will add rounding error to your calculations. Use the memory on your calculator so that there is no rounding until the end of your calculation.

You are also now in a position to refine the experiment by reducing the larger uncertainties. In this case, the largest uncertainty was in the temperature change for the water. Hence, it would not be very helpful to measure the masses to greater precision because the limit to precision in this activity would be the temperature differences. Getting greater precision in the temperature changes would be a useful refinement.

You could consider ways of getting larger temperature changes in the water and hence obtain a smaller percentage uncertainty in the temperature change. Alternatively, you might consider ways of measuring the temperatures to greater precision.

If your measurement range does not include the result you expect, you should think about the origin of the errors. In other words, if you are sure that c_{metal} is less than $189\,J\,kg^{-1}\,K^{-1}$ or more than $293\,J\,kg^{-1}\,K^{-1}$ then there must be some error in your experimental technique or more uncertainty than you realised.

When reviewing an experiment or a measurement exercise, it is a good idea to consider both errors *and* uncertainties.

Significant figures

The number of significant figures in a measurement is simply the number of digits used when the number is written in scientific notation. (Note: Significant figures were explained in Chapter 10.) Your calculator usually has eight or ten digits in the display of the answer for a calculation, but most of them are meaningless. You must round off your answer appropriately.

Consider the result of the experiment described in Worked example B3. It would make no sense to quote the result to two decimal places (or five significant figures) when clearly the precision of the experiment gives less than three significant figures.

Calculated results never have more significant figures than the original data and might have fewer than the original data. If you are not doing a full analysis of the uncertainties, it is customary to give your answers to the same number of significant figures as the least precise piece of data. For example, in Worked example B3, the least precise data is the change in temperature of the water, which has only a single digit. The value for the specific heat might then be quoted simply as $2 \times 10^2\,J\,kg^{-1}\,K^{-1}$, but doing the full calculation of the uncertainty in the result is much more informative.

GRAPHICAL ANALYSIS OF DATA

A major problem with doing a calculation from just one set of measurements is that a single incorrect measurement can significantly affect the result. Scientists like to take a large amount of data and observe the trends in that data. This gives more precise measurements and allows scientists to recognise and eliminate problematic data.

Physicists commonly use graphical techniques to analyse a set of data. In this section, the basic techniques that they use will be outlined and a general method for using a set of data that fits a known mathematical relationship will be developed.

Linear relationships

Some relationships studied in physics are linear, that is they can be represented by a straight line, while others are not. It is possible to manipulate non-linear data so that a linear graph reveals a measurement. Linear relationships and their graphs are fully specified with just two numbers: gradient, m, and vertical axis intercept, c. In general, linear relationships are written:

$$y = mx + c$$

The gradient, m, can be calculated from the coordinates of two points on the line:

$$m = \frac{\text{rise}}{\text{run}}$$

$$= \frac{y_2 - y_1}{x_2 - x_1}$$

where (x_1, y_1) and (x_2, y_2) are any two points on the line. Don't forget that m and c have units. Omitting these is a common error.

PHYSICSFILE

Graphs

When analysing data from a linear relationship, it is first necessary to obtain a graph of the data and an equation for the line that best fits the data. This line of best fit is often called the regression line. The entire process can be done on paper, but most people will use a computer spreadsheet, the capabilities of a scientific or a graphics calculator, or some other computer-based process. In what follows, it is not assumed that you are using any particular technology.

If you are plotting your graph manually on paper then proceed as follows:

1. Plot each data point on clearly labelled, unbroken axes.
2. Identify and label but otherwise ignore any suspect data points.
3. Draw, by eye, the 'line of best fit' for the points. The points should be evenly scattered either side of the line.
4. Locate the vertical axis intercept and record its value as 'c'.
5. Choose two points on the line of best fit to calculate the gradient. Do not use two of the original data points as this will not give you the gradient of the line of best fit.
6. Write $y = mx + c$, replacing x and y with appropriate symbols, and use this equation for any further analysis.

If you are using a computer or a graphics calculator then proceed as follows:

1. Plot each data point on clearly labelled, unbroken axes.
2. Identify suspect data points and create another data table without the suspect data.
3. Plot a new graph without the suspect data. Keep both graphs, as you don't actually discard the suspect data, but do eliminate it from the analysis.
4. Plot the line of best fit—the regression line. The manner in which you do this depends on the model of calculator or the software being used.
5. Compute the equation of the line of best fit that will give you values for m and c.
6. Write $y = mx + c$, replacing x and y with appropriate symbols, and use this equation for any further analysis.

Worked example B4

Some students used a computer with an ultrasonic detector to obtain the speed–time data for a falling tennis ball. They wished to measure the acceleration of the ball as it fell. The students assumed that the acceleration was nearly constant and that the relevant relationship was $v = u + at$, where v is the speed of the ball at any given time, u was the speed when the measurements began, a is the acceleration of the ball and t is the time since the measurement began.

Their computer returned the following data:

Time (s)	Speed (m s^{-1})
0.0	1.25
0.1	2.30
0.2	3.15
0.3	4.10
0.4	5.25
0.5	6.10
0.6	6.95

Find their experimental value for acceleration.

Solution

The data is assumed to be linear, with the relationship $v = u + at$, which can be thought of as being $v = at + u$, which makes it clear that putting v on the vertical axis and t on the horizontal axis gives a linear graph with gradient a and vertical intercept u. A graph of the data is shown in Figure B.6.

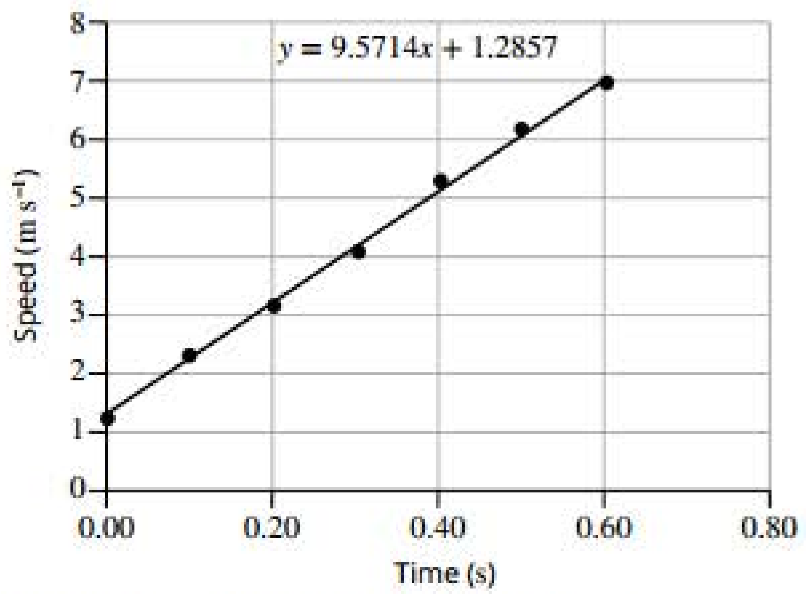


FIGURE B.6 Speed–time profile for a falling tennis ball.

This graph of the data was created on a computer spreadsheet. The line of best fit was created mathematically and plotted. The computer calculated the equation of the line. Graphics calculators can also do this.

A scientific calculator, graphics calculator or spreadsheet gives the regression line as $y = 9.5714x + 1.2857$. If this is rearranged and the constants are suitably rounded, the equation is $v = 1.3 + 9.6t$. This indicates that the ball was moving at $1.3\,m\,s^{-1}$ at the commencement of data collection and accelerating at $9.6\,m\,s^{-2}$.

Manipulating non-linear data

Suppose you were examining the relationship between two quantities B and d and had good reason to believe that the relationship between them is

$$B = \frac{k}{d}$$

where k is some constant value. Clearly, this relationship is non-linear and a graph of B against d will not be a straight line. By thinking about the relationship it can be seen that in 'linear form':

$$\begin{array}{ccccc} B & = & k & \frac{1}{d} & \\ \uparrow & & \uparrow & \uparrow & \\ y & = & m & x & + c \end{array}$$

A graph of B (on the vertical axis) against $\frac{1}{d}$ (on the horizontal axis) will be linear. The gradient of the line will be k and the vertical intercept, c, will be zero. The line of best fit would be expected to go through the origin because, in this case, there is no constant added and so c is zero.

In the above example, a graph of the raw data would just show that B is larger as d is smaller. It would be impossible to determine the mathematical relationship just by looking at a graph of the raw data.

A graph of raw data will not give the mathematical relationship between the variables, but it can give some clues. The shape of the graph of raw data may suggest a possible relationship. Several relationships may be tried and then the best chosen. Once this is done, it is not proof of the relationship but, possibly, strong evidence.

When an experiment involves a non-linear relationship, the following procedure is followed:

1. Plot a graph of the original raw data.
2. Choose a possible relationship based on the shape of the initial graph and your knowledge of various mathematical and graphical forms.
3. Work out how the data must be manipulated to give a linear graph.
4. Create a new data table.

Then follow the steps given in the Physicsfile on page 424. It may be necessary to try several mathematical forms to find one that seems to fit the data.

Worked example B5

Some students were investigating the relationship between current and resistance for a new solid-state electronic device. They obtained the data shown in the table.

According to the theory they had researched, the students believed that the relationship between I and R is $R = dI^3 + g$, where d and g are constants.

By appropriate manipulation and graphical techniques, find their experimental values for d and g. The following steps should be used:

a Plot a graph of the raw data.

b Work out what you would have to graph to get a straight line.

c Make a new table of the manipulated data.

d Plot the graph of manipulated data.

e Find the equation relating I and R.

Current, I (A)	Resistance, R (Ω)
1.5	22
1.7	39
2.2	46
2.6	70
3.1	110
3.4	145
3.9	212
4.2	236

Solution

a Figure B.7 shows the graph obtained using a spreadsheet.

It might be argued that the second piece of data is suspect. The rest of this solution supposes the students chose to ignore this piece of data.

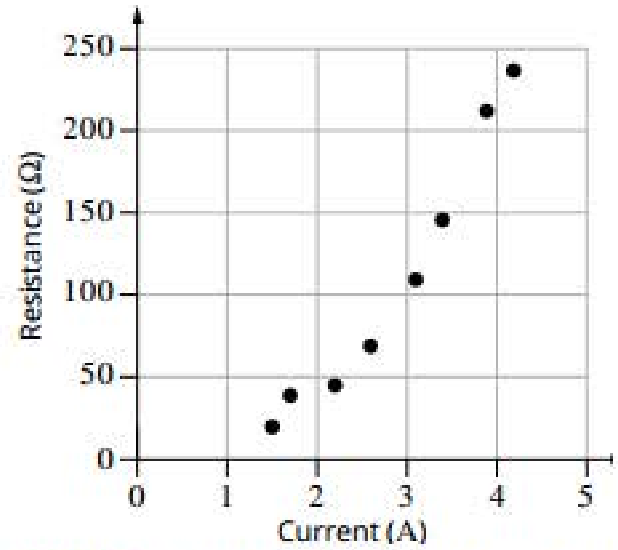


FIGURE B.7 Current–resistance graph of device.

b You can see what to graph if you think of the equation like this:

$$\begin{array}{ccccccc} R & = & d & I^3 & + & g \\ \uparrow & & \uparrow & \uparrow & & \uparrow \\ y & = & m & x & + & c \end{array}$$

A graph of R on the vertical axis and I^3 on the horizontal axis would have a gradient equal to d and a vertical axis intercept equal to g.

c The data is manipulated by finding the cube of each of the values for current.

Current cubed, I^3 (A^3)	Resistance, R (Ω)
3.38	22
10.65	46
17.58	70
29.79	110
39.30	145
59.32	212
74.09	236

d The graph in Figure B.8 was obtained from the spreadsheet.

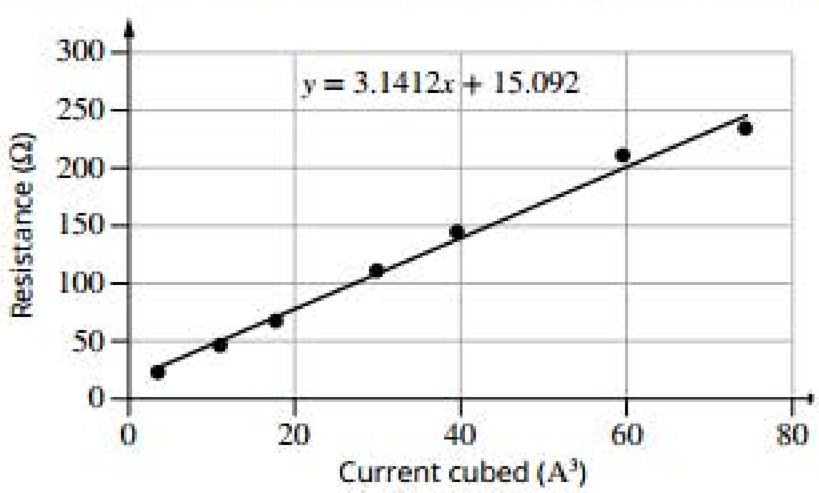


FIGURE B.8 Current–resistance characteristic (manipulated data).

e The regression line has the equation $y = 3.1x + 15.1$, so the equation relating I and R is $R = 3.1I^3 + 15.1$. Hence, the value of d is 3.1 Ω A^{-3} and the value of g is 15.1 Ω.

APPENDIX C Mathematical skills for physics ATAR

1. Transforming decimal notation to scientific notation

Scientists use scientific notation to handle very large and very small numbers.

For example, instead of writing 0.000 000 035, scientists would write 3.5×10^{-8}.

A number in *scientific notation* (also called standard form or power of ten notation) is written as:

$$a \times 10^n$$

where a is a number equal to or greater than 1 and less than 10, that is $1 \leq a < 10$.

n is an integer (a positive or negative whole number).

n is the power that 10 is raised to and is called the index value.

To transform a very large or very small number into scientific notation:

- Write the original number as a decimal number greater than or equal to 1 but less than 10.
- Multiply the decimal number by the appropriate power of 10.
- The index value is determined by counting the number of places the decimal point needs to be moved to form the original number again.
- If the decimal point is moved n places to the right, n will be a positive number. For example:

$$51 = 5.1 \times 10^1$$

- If the decimal point is moved n places to the left, n will be a negative number. For example:

$$0.51 = 5.1 \times 10^{-1}$$

You will notice from these examples that when large numbers are written in scientific notation, the 10 has a positive index value. When very small numbers are written in scientific notation, the 10 has a negative index value.

Practice questions

1 Match each number with its correct scientific notation.

Number	Scientific notation
0.002	2×10^3
2000	1.234×10^{-1}
0.1234	2×10^{-3}
12.34	1.234×10^1
123.4	1.234×10^2

2 Write 7.009×10^{-4} using decimal notation.

2. Identifying significant figures

When giving an answer to a calculation it is important to take note of the number of significant figures that you use.

You should give an answer that is as accurate as possible. However, an answer can't be more accurate than the data or the measuring device used to calculate it. For example, if a set of scales that measures to the nearest gram shows that an object has a mass of 56 g, then the mass should be recorded as 56 g, not 56.0 g. This is because you do not know whether it is 56.0 g, 56.1 g, 56.2 g or 55.8 g.

56 is a number with two significant figures. Recording to three significant figures (e.g. 56.0 g or 55.8 g) would not be scientifically 'honest'.

If this mass of 56 g is used to calculate another value, it would also not be 'honest' to give an answer that has more than two significant figures.

Determining the number of significant figures to give in an answer depends on what kind of calculation you are doing.

If you are multiplying or dividing, use the smallest number of significant figures provided in the initial values.

If you are adding or subtracting, use the smallest number of decimal places provided in the initial values.

Working out the number of significant figures

The following rules should be followed to avoid confusion in determining how many significant figures are in a number.

1 All non-zero digits are always significant. For example, 21.7 has three significant figures.
2 All zeroes between two non-zero digits are significant. For example, 3015 has four significant figures.
3 A zero to the right of a decimal point and following a non-zero digit is significant. For example, 0.5700 has four significant figures.
4 Any other zero is not significant, as it will be used only for locating decimal places. For example, 0.005 has just one significant figure.

Practice questions

1 Which of the following is written to two significant figures?
   A 30.1
   B 0.000 40
   C 0.5
   D 5.12
2 George multiplied 1.22 by 1.364. Which of the options below shows the result of this multiplication with the correct number of significant figures?
   A 1.66
   B 1.664
   C 1.65
   D 1.7
3 How can 41 be written to four significant figures?
   A 00.41
   B 4100
   C 41.00
   D 4.100
4 Alex is getting ready to go for a bike ride. Alex's mass is 65.3 kg. The bicycle has a mass of 12.92 kg.
   a Calculate the combined mass of Alex and the bicycle. Give your answer to the correct number of significant figures.
      A 78
      B 78.2
      C 78.22
      D 78.3
   b Using the combined mass calculated in part (a) above, and the formula $F_{net} = ma$, calculate the force Alex needs to apply to achieve an acceleration of 1.250 m s^{-1}. Give your answer to the correct number of significant figures.

3. Calculating percentages

Scientists use percentages to express a ratio or fraction of a quantity.

To express one quantity as a percentage of another, use the second quantity to represent 100%.

For example, expressing 6 as a percentage of 24 is like saying '6 is to 24 as x is to 100':

$$\frac{6}{24} = \frac{x}{100}$$
$$x = \frac{6}{24} \times 100$$
$$= 25\%$$

To calculate a percentage of a quantity, the percentage is expressed as a decimal then multiplied by the quantity.

For example, to calculate 40% of 20:

$$x = \frac{40}{100} \times 20$$
$$= 0.4 \times 20$$
$$= 8$$

Practice questions

1 What is 9 as a percentage of 12?
A 25%
B 50%
C 75%
D 30%

2 What is 25% of 24?
A 6.6
B 6
C 5
D 0.5

3 Which of the following values expresses 15 as a percentage of 120?
A 8%
B 5%
C 12.5%
D 0.125%

4. Converting between percentages and fractions

To write a percentage as a fraction, divide the percentage by 100.

For example:

$$25\% = \frac{25}{100}$$
$$= \frac{1}{4}$$

$\frac{25}{100}$ is not the simplest form of this fraction. If you divide both the numerator and the denominator by 25 (their highest common factor) then the fraction simplifies to $\frac{1}{4}$.

Whenever you give a fraction as an answer, always try and simplify it by dividing the numerator and denominator by the highest common factor.

To write a fraction as a percentage, multiply the fraction by 100%. In many cases it is easier to convert the fraction to a decimal number first.

For example:

$$\frac{1}{4} = 0.25 \times 100$$
$$= 25\%$$

The value of the fraction or percentage has not changed. It is just being represented in a different way.

Practice questions

1 Choose the option that expresses $\frac{1}{5}$ as a percentage.
A 25%
B 20%
C 30%
D 50%

2 Match each percentage with its corresponding fraction.

Percentage	Fraction
0.2%	$\frac{7}{20}$
2.5%	$\frac{7}{40}$
17.5%	$\frac{1}{40}$
35%	$\frac{111}{250}$
44.4%	$\frac{1}{500}$

5. Changing the subject of an equation

Scientists use equations to represent relationships between variables. In an equation like $A = \pi r^2$, A is called the subject of the equation.

Sometimes the subject of the equation has to be changed in order to express the relationship in a more useful way. For example, if you need to find the radius of a circle, you would want r to be the subject of the equation above.

To change the subject of a simple equation, transpose the equation to leave the new subject on its own. In the example above, the equation needs to read

$r = \dots$

Keep the equation balanced by performing the same operation to both sides of the equation to cancel operations being performed on the desired subject. Inverse operations (the opposite operation; for example, dividing is the inverse of multiplying) will allow cancelling.

For example, make r the subject of the equation $A = \pi r^2$.

1 Divide both sides of the equation by π.

$$A = \pi r^2$$

$$\frac{A}{\pi} = \frac{\pi r^2}{\pi}$$

The π in the numerator and denominator on the right side of the equation cancel out, giving

$$\frac{A}{\pi} = r^2$$

2 To cancel the squaring operation of r, take the square root of both sides of the equation.

$$\frac{A}{\pi} = r^2$$

$$\sqrt{\frac{A}{\pi}} = \sqrt{r^2}$$

The square and square root on the right side of the equation cancel out, giving

$$\sqrt{\frac{A}{\pi}} = r$$

3 r is now the subject of the equation.

$$r = \sqrt{\frac{A}{\pi}}$$

Practice questions

1 Rearrange the formula $A = \frac{2}{3}R$ to make R the subject.

A $R = \frac{2A}{3}$

B $R = \frac{A}{3}$

C $R = \frac{3A}{2}$

D $R = 6A$

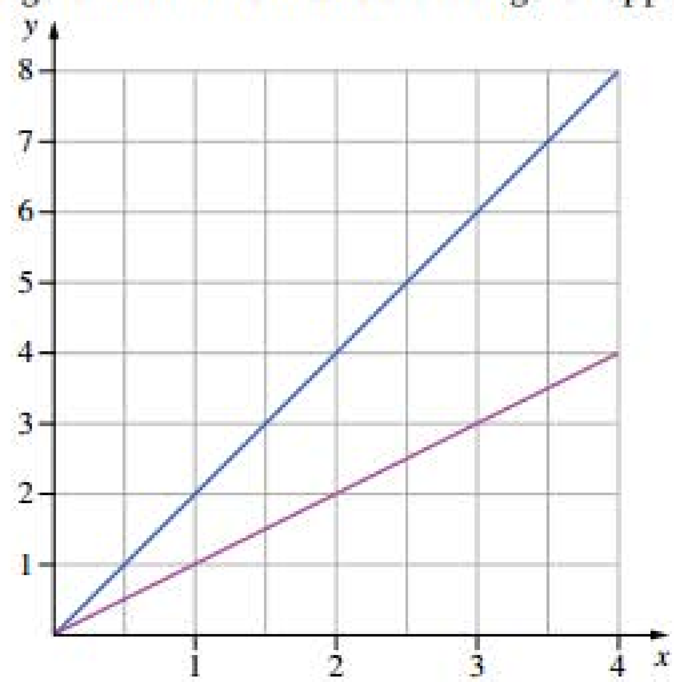
2 Rearrange the formula $y = 3\sqrt{\frac{p}{q}}$ to make p the subject.

6. Interpreting the slope of a linear graph

Scientists often represent a relationship between two variables as a graph. For directly proportional relationships, the variables are connected by a straight line, where the slope (or gradient) of the line represents the constant of proportionality between the two variables.

The slope or gradient of the line is defined as the ratio of change between two points in the vertical axis (ΔY), divided by the change between two points in the horizontal axis (ΔX). In other words, it measures the rate at which one variable (the dependent variable) changes with respect to the other (the independent variable).

The graph below has two straight lines with different slopes. The steeper slope (blue line) indicates that the rate of change is higher. This means the change is happening more quickly. On the other hand, the flatter slope (purple line) indicates that the rate of change is lower. This means the change is happening more slowly.



Practice questions

1 On a graph with two sloped lines, what does the steeper sloped line indicate?
  A a faster rate of change
  B a slower rate of change
  C the same rate of change
  D a much slower rate of change

2 The rate of change of a straight line on a graph is given by the:
  A y-intercept
  B x-intercept
  C gradient
  D area under the graph

7. Understanding mathematical symbols

Part of the language of science is using symbols to represent quantities or to give meanings. For example, the four symbols $<$, $>$, $\leq$ and $\geq$ are known as 'inequalities'.

The following mathematical symbols are commonly used in science.

Symbol	Meaning	Example	Explanation
$<$	less than	$2 < 3$	2 is less than 3
$>$	greater than	$6 > 1$	6 is greater than 1
$\leq$	less than or equal to	$2x \leq 10$	$2x$ is less than or equal to 10
$\geq$	greater than or equal to	$3y \geq 12$	$3y$ is greater than or equal to 12
$\sqrt{\ }$	square root	$\sqrt{4} = 2$	The square root of 4 is 2
Δ	change in (difference between)	Δt	change in t (time)
$\approx$	approximately equal to	$\pi \approx 3.14$	π is approximately equal to 3.14
Σ	summation	$\sum_{i=1}^{4} i$	The sum of consecutive integers from 1 to 4, i.e. $1 + 2 + 3 + 4 = 10$

Practice questions

1 The symbol that means 'less than' is:
  A $<$
  B $>$
  C $\leq$
  D $\geq$

2 Which of these symbols is an inequality?
  A $\approx$
  B Δ
  C $\sqrt{\ }$
  D $\leq$

8. Understanding the difference between discrete and continuous data

Quantitative data forms the backbone of science. Scientists are constantly working with data—measuring, recording, analysing and interpreting it.

Quantitative data consists of numerical values that can either be discrete or continuous.

Discrete data is data that has a set of clearly defined values. For example, the number of students in a class would have a discrete set of possible data.

Continuous data is usually data that is measured in some way and can have an infinite number of values. For example, your height or weight would have a continuous set of possible data.

The easiest way to distinguish between the two types of quantitative data is to ask, 'Is the data measured or counted?' If it is counted, the data set is discrete. If it is measured, the data set is continuous.

Practice questions

1 Which one of these data sets is continuous?
A the number of cars parked in a street
B the temperature of the air over a 24-hour period
C the number of students at a school
D the number of nails used to build a fence

2 Which one of these data sets is discrete?
A the number of cars parked in a street
B the temperature of the air over 24 hours
C the heights of a team of footballers
D the mass of a team of netballers

9. Calculating the mean, median and range of a data set

When handling data, scientists often look for ways to describe patterns in the data. Common terms used when analysing a set of data include the mean, median and range.

Mean: the *average* value in the data set. To calculate the mean, sum all the values in the data set and then divide this total by the number of data values.

Median: the *middle* value in an ordered data set. To calculate the median, arrange the data set in ascending order and then count the number of data values. If the number of values is odd, the median is the middle value. If the number of values is even, calculate the median by adding the two middle values and dividing by 2, i.e. by calculating the average of the two middle numbers.

Range: the *spread* of values in the data set. To calculate the range, take the largest data value and then subtract the smallest data value.

Practice questions

1 The following set of data is recorded:
44, 17, 21, 26, 42, 18
Find the
a mean
b median
c range.

2 The mass in kilograms of each student in a class of 25 students is recorded below. The combined mass of all the students is 1340 kg.
Find the
a mean
b median
c range.
Students' weights: 67, 60, 41, 52, 39, 60, 42, 55, 55, 50, 46, 62, 48, 48, 56, 64, 55, 56, 59, 61, 41, 63, 53, 62, 45

10. Solving simple algebraic equations

To solve an equation means to find the values that make the equation true. Scientists manipulate equations and substitute in known variables in order to solve for the variable required.

For example, you can solve

$$a = \frac{F_{net}}{m}$$

where F_{net} is the net force on the car, which is 2400 N

m is the mass of the car, which is 1200 kg

a is the acceleration of the car in m s^{-2}, which is unknown.

$$a = \frac{2400}{1200}$$

$$= 2\ \text{m s}^{-2}$$

Practice questions

1 Solve the equation $V = IR$ if $I = 3$ and $R = 9$.

A $V = 3$
B $V = 6$
C $V = 12$
D $V = 27$

2 Solve the equation and find the value of Q if $Q = mc\Delta T$, $m = 1.2$, $c = 4200$ and $\Delta T = 30$.

A $Q = 840$
B $Q = 25\ 200$
C $Q = 126\ 000$
D $Q = 6$

11. Completing calculations with more than one operation

Scientists often deal with complex calculations that can involve numerous operations within the one calculation. The order in which these operations are performed can affect the result of the calculation.

For example, the calculation $2 + 3 \times 4$ will give an incorrect answer of 20 if you calculate the $2 + 3$ part first, but it will give the correct answer of 14 if you calculate the 3×4 part first.

Scientists and mathematicians have agreed on a set order in which operations are carried out so that calculations are consistent. You can remember this order using the acronym 'BIDMAS':

- brackets
- indices (powers, square roots, etc.)
- division and multiplication
- addition and subtraction.

The operations present in a calculation are performed in the order shown in the list. If there are multiple instances of division and multiplication, or addition and subtraction, work from left to right.

For example, the 3×4 part in the original example would always be performed first, since multiplication is higher in the list than addition.

When dealing with scientific notation, it is important to keep each individual number complete. Use brackets to help do this, especially when dividing.

For example, when dividing 3.01×10^{21} by 6.02×10^{23}, brackets are used to keep the second number together.

If you used a calculator and entered $\frac{3.01 \times 10^{21}}{6.02 \times 10^{23}}$ (i.e. $3.01 \times 10^{21} \div 6.02 \times 10^{23}$), the answer would come out as 5.00×10^{43}. This is not the correct answer.

The correct answer is only obtained by entering $\frac{3.01 \times 10^{21}}{(6.02 \times 10^{23})}$. This time, the answer comes out correctly as 5.00×10^{-3}.

Using the calculator's *EXP* button or the $x10^x$ button keeps the number and power of 10 together as one number and avoids the problems of using the number multiplied by 10^x. If your calculator does not have an *EXP* or $x10^x$ button, check the user manual. It may be that it is just labelled differently on your calculator. Alternatively, remember to always use brackets to keep the terms in the denominator together, as shown above.

Practice questions

1 What is 3.4×10^{-4} divided by 1.7×10^{-3}?
A 2×10^{-7}
B 2
C 20
D 0.2

2 Substitute $m = 1.4$, $d = 3.9$ and $c = 2.7$ into $W = 6m - 4(d + c)$ and solve for W.
A –4.5
B –18
C 34.8
D 26.7

12. Understanding the relationship between data, graphs and algebraic rules

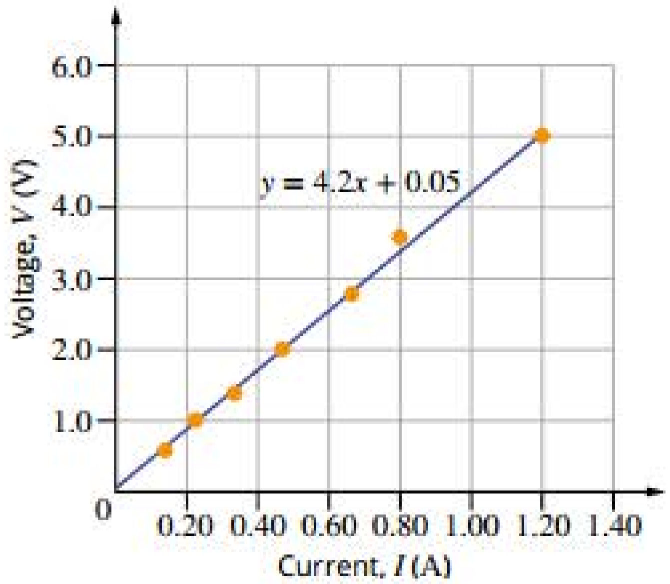
Scientists use graphs to analyse the data they collect from experiments. All graphs tell a story. The shape of the graph shows the relationship between the variables, and this relationship can be written algebraically and numerically. The horizontal axis is known as the x-axis and the vertical axis is known as the y-axis.

Once the algebraic rule is known, the values for one variable can be substituted and the values for the other variable can be calculated. These values can also be determined by reading them from the graph.

For example, when investigating how current and voltage vary across a light bulb, the following data was collected:

Current, *I* (A)	Voltage, *V* (V)
0.14	0.6
0.22	1.0
0.33	1.4
0.47	2.0
0.66	2.8
0.80	3.6
1.20	5.0

Graphing this data produced:



The numerical values from the experiment are listed in the table and plotted on the graph. The algebraic relationship between the variables is given by the equation of the line:

$$y = 4.2x + 0.05$$

The value of the y-intercept is approximately zero, so assuming that the y-intercept is zero, and labelling the x-axis as current and the y-axis as voltage, the relationship can be written as:

$$y = 4.2 \times \text{current}$$

Using the appropriate symbols this can also be written as: $V = 4.2I$

Practice questions

1 If a graph had L on the y-axis and B on the x-axis and the equation of the straight line was $y = 3.7x$, what is the algebraic form of the graph?
A $L = 3.7x$
B $y = 3.7x$
C $y = 3.7B$
D $L = 3.7B$

2 If a relationship was written as $m = 5.9L$, what shape would the graph be and which variable would be plotted on which axis?
A The graph would be non-linear, with m on the y-axis and L on the x-axis.
B The graph would be non-linear, with m on the x-axis and L on the y-axis.
C The graph would be linear, with m on the y-axis and L on the x-axis.
D The graph would be linear, with m on the x-axis and L on the y-axis.

13. Recognising and using ratios

A ratio is the relationship between two numbers of the same kind. It could be the quantities in a recipe, the division of profits from a sale, or the number of different types of the same thing.

Scientists use ratios to compare quantities. This might be the numbers of atoms of different elements in a compound, or the number of primary and secondary windings of a transformer.

You can also use the principle of ratios to solve problems. For example, if 1 reaction produces 2.5 MeV of energy, then how much energy does 34 reactions produce?

The reaction-to-energy ratio of 1:2.5 should remain constant as the number of reactions increases. So you need to find the factor that 1 needs to be multiplied by to give 34:

$$1 \times 34 = 34$$

You then multiply the energy amount by the same factor:

$$2.5 \times 34 = 85\,\text{MeV}$$

You may also see ratios expressed as fractions.

Practice questions

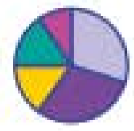
1 If the primary coil of a transformer has 2400 windings and the secondary coil has 600 windings, what is the simplest ratio of primary to secondary windings?
A 2400:600
B 24:6
C 12:3
D 4:1

14. Understanding pie charts, frequency graphs, and histograms

It is essential in science to collect data and arrange it in an orderly way. Tables are often used to organise data, which can then be displayed in a graph.

Pie charts

A pie chart is a circle that is divided into sectors. Each sector represents one item in the data set and is shown as a percentage or fraction of the total data set.



Frequency graphs and histograms

Frequency graphs and histograms are another way of representing data visually.

If data is discrete (i.e. can be counted), each column in a column graph will represent one category, e.g. 'apples' or 'strawberries'. Often these columns have a gap between them.

If the data is continuous (i.e. can be measured), such as the heights of the students in that class, each column will represent a range of possible heights, e.g. 140 to 160 cm, and there will be no gaps between the columns. These are called histograms.

Practice questions

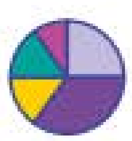
1 Choose the pie chart that correctly shows the data from Emmanuel's poll of soccer fans.

Number of matches watched	10	12	15	18	21
Relative frequency (%)	25	35	15	15	10

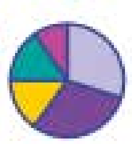
Matches watched:

- ■ = 10
- ■ = 12
- ■ = 15
- ■ = 18
- ■ = 21

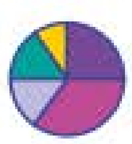
A



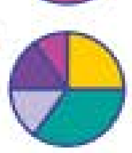
B



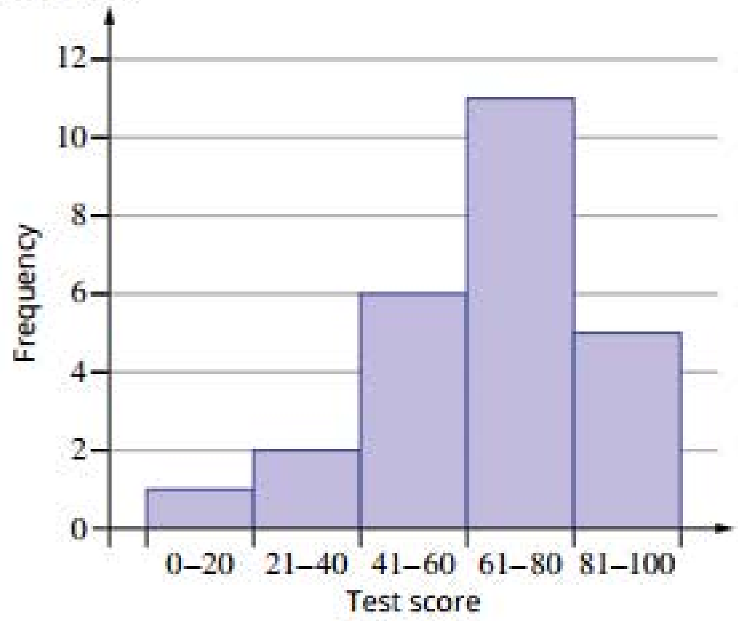
C



D



2 A class of Year 11 Chemistry students had a test. A histogram of their test scores is shown below.



a How many students scored over 80?
b How many students scored below 40?
c How many student are in this class?

15. Understanding the graphical representation of a sine curve

The sine curve is a mathematical curve that describes a smooth repetitive oscillation. It is relevant to the physics topics of sound, AC electricity, simple harmonic motion, waves and many others.

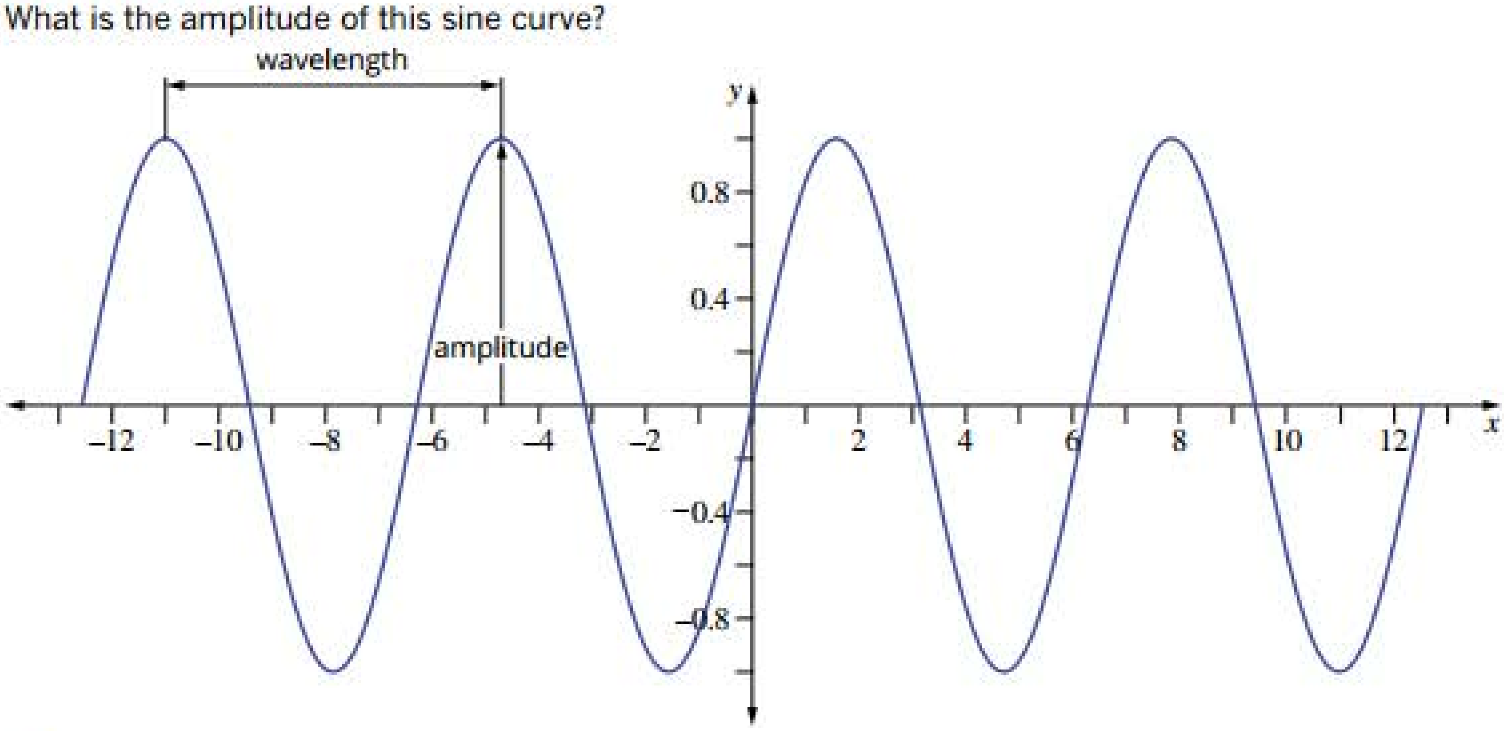
The amplitude of the curve is the distance from the midpoint of the curve to the highest peak or lowest trough, or half the distance from the lowest to the highest point.

The wavelength of the curve is the length of one complete wave.

The period of the curve is the time measurement for one complete cycle of the wave.

Practice questions

1 What is the amplitude of this sine curve?



A 1
B 2
C 4
D 8

2 Sine curves in science can be applied to:
A AC electricity
B simple harmonic motion
C sound waves
D all of the above

16. Understanding sine, cosine and tangent relationships in right-angled triangles

Vector problems in physics (and other applications) often involve using trigonometric relationships to find the unknown side of a right-angled triangle.

You will recall that:

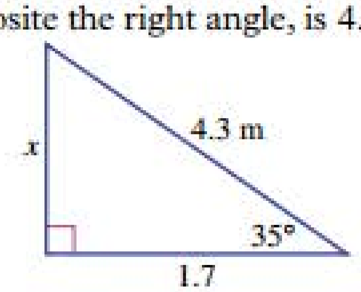
$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}},\ \cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}},\ \tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$

The acronyms for each of these rules are:

Equations	$\sin\theta$	$\cos\theta$	$\tan\theta$
Acronym	SOH	CAH	TOA

To solve a trigonometry problem, the appropriate formula needs to be identified and solved for the unknown side or angle of a triangle.

In the triangle below, side x is opposite the angle shown as 35°. The hypotenuse, which is always the side opposite the right angle, is 4.3 m long.



θ and the length of the hypotenuse are known. To find the length of the opposite side, use the SOH acronym:

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

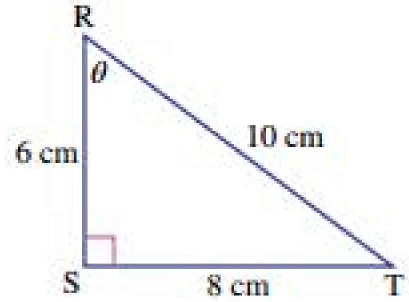
$$\sin 35° = \frac{x}{4.3}$$

$$x = 4.3 \times \sin 35°$$

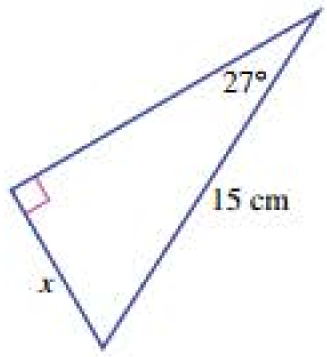
$$= 2.5$$

Practice questions

1 On the following right-angled triangle, label the sides as opposite (O), adjacent (A) and hypotenuse (H) in relation to the angle θ.



2 What is the length of side x in the following triangle? Give your answer to two decimal places.



A 6.81 cm
B 13.37 cm
C 7.64 cm
D 12.25 cm

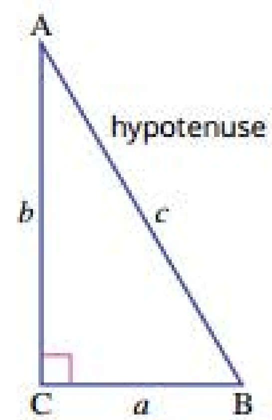
17. Understanding Pythagoras' theorem and similar triangles

Vector problems in physics (and other applications) often involve finding the side lengths of a right-angled triangle and other geometrical unknowns.

Pythagoras' theorem

For any right-angled triangle, Pythagoras' theorem states that the square of the length of the hypotenuse is equal to the sum of the squares of the lengths of the two shorter sides.

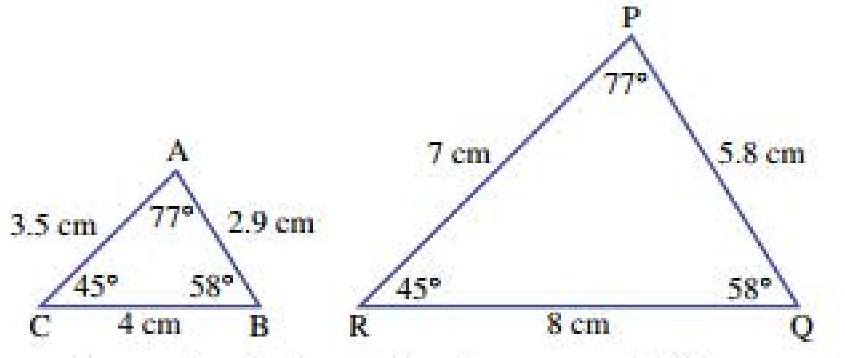
$$a^2 + b^2 = c^2$$



Similar figures

Similar figures are figures that have the same shape, but are not necessarily the same size.

In similar triangles, all pairs of corresponding angles are equal and all pairs of matching sides are in the same ratio. For example, the triangles shown below are similar triangles.



The corresponding angles in these triangles are equal. The corresponding sides in these triangles are in the same ratio. The ratio can be shown as:

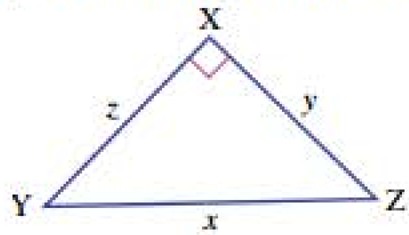
$$\frac{PQ}{AB} = \frac{5.8}{2.9} = 2$$

$$\frac{QR}{BC} = \frac{8}{4} = 2$$

$$\frac{PR}{AC} = \frac{7}{3.5} = 2$$

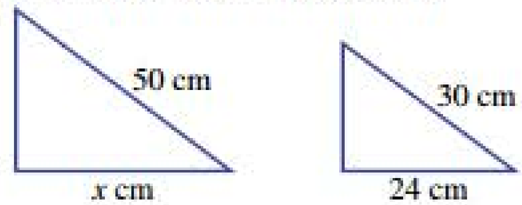
Practice questions

1 For the following triangle, select the correct statement of Pythagoras' theorem.



A $z^2 + x^2 = y^2$
B $z^2 - y^2 = x^2$
C $x^2 + y^2 = z^2$
D $z^2 + y^2 = x^2$

2 These two triangles are similar. Find the value of *x*.



18. Using units in an equation to check for dimensional consistency

Scientists know that each term in an equation represents a quantity. The units used to measure that quantity are not used in the calculations. Units are only indicated on the final line of the solved equation.

For example, this is the equation for the area (A) of a rectangle of length (L) and width (W):

$$A = L \times W$$

If L has a value of 7 m and W has a value of 4 m, it is written:

$$\begin{aligned} A &= L \times W \\ &= 7 \times 4 \\ &= 28\,\text{m}^2 \end{aligned}$$

Note that the units are left out of the actual calculation on the second line and only included at the end, after the numerical answer.

You can use units to check the dimensional consistency of the answer. In the example above, the two quantities of L (length) and W (width) both have to be expressed in consistent units, in this case metres (m), to give an answer that is expressed in square metres ($\text{m} \times \text{m} = \text{m}^2$).

If you had made a mistake, and used the formula $A = L + W$ instead, the answer would be expressed in metres only. This is not the correct unit to express area, so you would know that was wrong.

Practice questions

1 Which formula has the correct dimensions for calculating volume in m^3?
A m × m
B m × s
C m × m × s
D m × m × m

2 Which of these shows the correct substitution into $P = 2L + 2W$ using consistent units for $L = 3.5$ cm and $W = 240$ cm?

A $P = 2 \times 3.5 + 2 \times 240$
B $P = 2 \times 3.5 + 2 \times 24$
C $P = 2 \times 35 + 2 \times 240$
D $P = 2 \times 3.5 + 2 \times 2.4$

3 By using the equation $p = mv$ as a guide, select the correct units in which to measure momentum, p.

A m s
B m s^{-1}
C kg m s^{-1}
D kg m s^{-2}

19. Understanding inverse and inverse square relationships

Some relationships in science involve one quantity in a relationship increasing and the other quantity decreasing proportionally. This is an inverse relationship.

Thicker wires have lower resistance. This means that, as the cross-sectional area of a wire increases, the value of its resistance decreases. This is an inverse relationship and it can be written as:

$$R \propto \frac{1}{A} \text{ or } R \times A$$

(This shows it is a fixed amount that doesn't change even if R and A change.)

An inverse square relationship is similar, but one quantity increases as the square of the other quantity decreases. Wires with a larger radius will have a lower resistance following an inverse square law.

$$R \propto \frac{1}{r^2} \text{ or } R \times r^2$$

Practice questions

1 Which of these rules represents an inverse relationship?

A $B \propto \frac{1}{L}$
B $y \propto d$
C $X \propto \frac{1}{c^2}$
D $K \propto t^2$

2 Which of these rules represents an inverse square relationship?

A $B \propto \frac{1}{L}$
B $y \propto d$
C $X \propto \frac{1}{c^2}$
D $K \propto t^2$

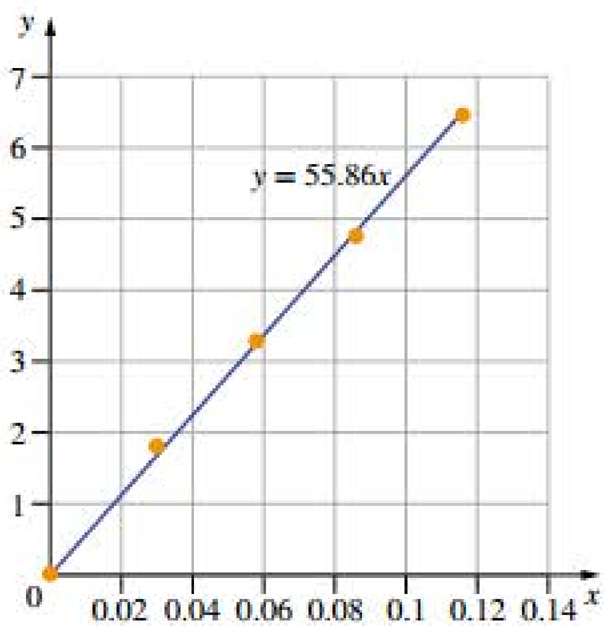
20. Understanding lines of best fit

When scientists observe that the points on a graph seem to form a straight line, then a line of best fit (or trendline) can be drawn through them. Computer programs such as Excel can fit a trendline to a data set; lines of best fit can also be drawn by hand onto a printed or drawn graph.

The line of best fit should pass as close as possible to as many of the points as possible (i.e. it should 'fit' the data closely). It may not pass exactly through any of the points, but once the line of best fit is drawn, the points should be spaced equally on each side, above and below the line. There should be no points very far away from the line, unless they are considered to be unreliable. Unreliable points are called 'outliers' and can be disregarded for the purposes of creating a line of best fit.

The gradient and y-intercept of the line (not the points) can be determined to find the relationship between the variables using the general equation for a straight line: $y = mx + c$.

For example:



Practice questions

1 Are the following statements true or false?
   a All the points on a scatter plot must lie on the line of best fit.
   b The line of best fit may pass through none of the points.
   c The points of the scatter plot should lie close to the line of best fit.

2 Are the following statements true or false?
   a A trendline is the same as a line of best fit.
   b A line of best fit can only be drawn using a computer program such as Excel.
   c Outliers should be included as normal points when considering where to draw a line of best fit.

Mathematical skills for physics ATAR

Answers to practice questions

1. Transforming decimal notation to scientific notation

1

0.002	2×10^{-3}
2000	2×10^{3}
0.1234	1.234×10^{-1}
12.34	1.234×10^{1}
123.4	1.234×10^{2}

0.002: Move the decimal point three places to the right which gives an index of –3, so 0.002 is written as 2×10^{-3}.
2000: Move the decimal point three places to the left which gives an index of +3, so 2000 is written as 2×10^{3}.
0.1234: Move the decimal point one place to the right which gives an index of –1, so 0.1234 is written as 1.234×10^{-1}.
12.34: Move the decimal point one place to the left which gives an index of +1, so 12.34 is written as 1.234×10^{1}.
123.4: Move the decimal point two places to the left which gives an index of +2, so 123.4 is written as 1.234×10^{2}.

2 The –4 index shows the decimal point has been moved four places to the right. Move it back four places to the left to give 0.000 7009.

2. Identifying significant figures

1 B 0.000 40

2 A 1.66
When two numbers are multiplied, use the smallest number of significant figures in the initial values to give your answer. In George's multiplication, the answer is 1.66408, but as 1.22 has three significant figures, the correct answer is 1.66.

3 C. 41.00
The zeroes to the right of the decimal point are significant. When writing an answer to a correct number of significant figures, you may need to use rounding if the initial answer has more figures (digits) than you need. These extra figures are called 'non-significant figures.' If the first non-significant figure is ≥5, you round up. If the first non-significant figure is <5, then do not round up. For example, if the initial answer for a calculation was 2.1259 but you only needed an answer to three significant figures, you would round it to 2.13. If the initial answer was 2.1241, then to three significant figures this becomes 2.12.

4 a B 78.2
When 65.3 is added to 12.92, the answer is 78.22, but only if it is assumed that 65.3 is actually 65.30. When adding or subtracting with significant figures, use the smallest number of significant figures provided in the initial values. As there is no way of knowing the accuracy beyond 65.3, the answer should only be given to three significant figures, which is one decimal place, 78.2.

b When multiplying or dividing, use the smallest number of significant figures in the initial values to give your answer.

$$F_{net} = ma$$
$$= 78.2 \times 1.250$$
$$= 97.8$$

Since the mass value only has three significant figures, the answer should also only have three significant figures. Therefore, even though the calculated answer is 97.75, the answer to three significant figures is 97.8 m s^{-1}.

3. Calculating percentages

1 C 75%

$$\frac{9}{12} = \frac{x}{100}$$
$$x = \frac{9}{12} \times 100$$
$$= 75\%$$

2 B 6

$$x = \frac{25}{100} \times 24$$
$$= 0.25 \times 24$$
$$= 6$$

3 C 12.5%

$$\frac{15}{120} = \frac{x}{100}$$
$$x = \frac{15}{120} \times 100$$
$$= 12.5\%$$

4. Converting between percentages and fractions

1 B 20%
To write a fraction as a percentage, multiply the fraction by 100.

$$\frac{1}{5} = 0.2 \times 100$$
$$= 20\%$$

2

Percentage	Fraction
0.2%	$\frac{1}{500}$
2.5%	$\frac{1}{40}$
17.5%	$\frac{7}{40}$
35%	$\frac{7}{20}$
44.4%	$\frac{111}{250}$

To change a percentage to a fraction, divide by 100.

$$0.2\% = \frac{0.2}{100} = \frac{2}{1000} = \frac{1}{500}$$
$$2.5\% = \frac{2.5}{100} = \frac{25}{1000} = \frac{1}{40}$$
$$17.5\% = \frac{17.5}{100} = \frac{175}{1000} = \frac{35}{200} = \frac{7}{40}$$
$$35\% = \frac{35}{100} = \frac{7}{20}$$
$$44.4\% = \frac{44.4}{100} = \frac{444}{1000} = \frac{222}{500} = \frac{111}{250}$$

5. Changing the subject of an equation

1 C $R = \frac{3A}{2}$

Multiply both sides of the equation by 3

$3A = 3 \times \frac{2}{3}R$

$= 2R$

Divide both sides of the equation by 2

$\frac{3A}{2} = R$

Rewrite the equation so it reads

$R = \frac{3A}{2}$

2 $p = \frac{y^2 q}{9}$

Divide both sides of the equation by 3

$\frac{y}{3} = \sqrt{\frac{p}{q}}$

Square both sides

$\frac{y}{3}^2 = \frac{p}{q}$

Expand the brackets on the left

$\frac{y^2}{9} = \frac{p}{q}$

Multiply both sides by q

$\frac{y^2 q}{9} = p$

Rewrite the equation so it reads

$p = \frac{y^2 q}{9}$

6. Interpreting the slope of a linear graph

1 A. The steepness of the slope indicates the rate of change. A line with a steeper slope indicates a faster rate of change.

2 C. The gradient or slope of a linear graph indicates the rate of change.

7. Understanding mathematical symbols

1 A <. One way is to remember this is that the smaller end of the shape points toward the smaller number. For example, $3 < 6$ means 3 is less than 6.

2 D ≤. The symbols <, >, ≤ and ≥ are all inequalities.

8. Understanding the difference between discrete and continuous data

1 B the temperature of the air over a 24-hour period.
If the data can be counted, the data set is discrete. If the data can be measured, the data set is continuous. The temperature of the air can be measured with a thermometer, so the data set is continuous.

2 A the number of cars parked in a street.
If the data can be counted, the data set is discrete. If the data can be measured, the data set is continuous. The number of cars parked in a street can be counted, so the data set is discrete.

9. Calculating the mean, median and range of a data set

1 **a** $44 + 17 + 21 + 26 + 42 + 18 = 168$

$\frac{168}{6} = 28$

b Place the numbers in ascending order: 17, 18, 21, 26, 42, 44
As there is an even number of values, add the two middle values and divide by 2.

$\frac{21+26}{2} = 23.5$

c $44 - 17 = 27$

2 **a** $\frac{1340}{25} = 53.6$ kg

b 55 kg

c $67 - 39 = 28$kg

10. Solving simple algebraic equations

1 D $V = 27$

Substitute the values and solve the equation.

$V = 3 \times 9$

$= 27$

2 B $Q = 25200$

Substitute the values and solve the equation.

$Q = 0.2 \times 4200 \times 30$

$= 25200$

11. Completing calculations with more than one operation

1 D 0.2

The correct calculation, using brackets, is:

$\frac{3.4 \times 10^{-4}}{(1.7 \times 10^{-3})} = 0.2$

2 B −18

$W = 6 \times 1.4 - 4 \times (3.9 + 2.7)$

$= 6 \times 1.4 - 4 \times 6.6$

$= 8.4 - 26.4$

$= -18$

12. Understanding the relationship between data, graphs and algebraic rules

1 D $L = 3.7B$

Substituting L for y and B for x gives $L = 3.7B$.

2 C. The graph would be linear, with m on the y-axis and L on the x-axis.
The graph would be linear as the equation is written in the form $y = mx + 0$ with m on the y-axis and L on the x-axis.

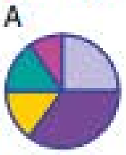
13. Recognising and using ratios

1 D 4:1

The primary to secondary ratio is 2400:600. Dividing both sides of the ratio by 600 simplifies it to 4:1.

14. Understanding pie charts, frequency graphs, and histograms

1 A



The largest sector of the pie chart is purple, representing the percentage of people who watched 12 matches (35%). The next in size is blue (10 matches, 25%), then green (18 matches, 15%) and yellow (15 matches, 15%). The smallest sector is red (21 matches, 10%).

2 **a** 5

b 3

c 25

The height of the 81–100 column is 5. This shows five students scored over 80.

The heights of the 0–20 and 21–40 columns are 1 and 2. This shows that three students scored below 40.

If you add up the heights of all the columns (1 + 2 + 6 + 11 + 5 = 25), it tells you that there are 25 students in the class.

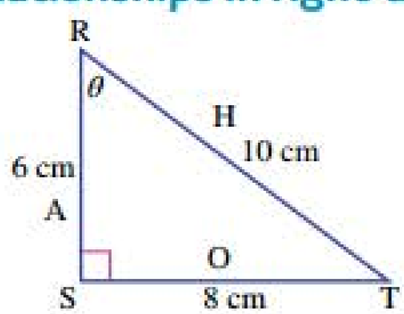
15. Understanding the graphical representation of a sine curve

1 A 1. The amplitude of the curve is the distance from the midpoint (0) to the highest peak (1) or the lowest trough (–1). You could also work out distance from the lowest point to the highest point (1 + 1 = 2) and halve that figure. The amplitude of this sine curve is 1.

2 D all of the above. Sine curves have many applications in physics including those mentioned here.

16. Understanding sine, cosine and tangent relationships in right-angled triangles

1



The hypotenuse is the longest side and the one opposite the right angle, so the 10 cm side is labelled H.

The opposite side is the side opposite the θ, so the 8 cm side is labelled O.

The adjacent side is the side next to the angle θ, so the 6 cm side is labelled A.

2 A 6.81 cm

First, identify the correct trigonometric formula to use. You know the angle 27°, and the hypotenuse is 15 cm long. You want to find the length of x, the side opposite the angle, so use the SOH or sine formula.

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

$$\sin 27° = \frac{x}{15}$$

$\sin 27° \times 15 = x$

$x = 6.81$ cm

The side labelled x is 6.81 cm long.

17. Understanding Pythagoras' theorem and similar triangles

1 D $z^2 + y^2 = x^2$

The hypotenuse, which is the longest side of a right-angled triangle, is opposite the right angle. In this case it is side x.

x^2 is the sum of the squares of the other two sides: $x^2 = z^2 + y^2$

2 $x = 40$ cm

Identify the matching sides. These will be in the same ratio. The hypotenuse of the first triangle is 50 cm. The hypotenuse of the second triangle is 30 cm. The ratio can be written as $\frac{50}{30}$.

The other matching sides will have the same ratio.

$$\frac{x}{24} = \frac{50}{30}$$

$$x = \frac{50 \times 24}{30}$$

$$= 40$$

The value of x is 40.

18. Using units in an equation to check for dimensional consistency

1 D m × m × m

2 D $P = 2 \times 3.5 + 2 \times 2.4$

Both units must be consistent.

Since 100 cm = 1 m the consistent units must be either

$L = 3.5$ m and $W = 2.4$ cm

or

$L = 350$ cm and $W = 240$ cm

The only correct combination is $P = 2 \times 3.5 + 2 \times 2.4$.

3 C kg m s^{-1}

The unit for momentum is taken from the unit for mass (kg) multiplied by the unit for velocity (m s^{-1}). Therefore, momentum is measured in kg m s^{-1}.

19. Understanding inverse and inverse square relationships

1 A $B \propto \frac{1}{L}$

An inverse relationship contains a term in the form $\frac{1}{x}$.

2 C $X \propto \frac{1}{c^2}$

An inverse square relationship contains a term in the form $\frac{1}{x^2}$.

20. Understanding lines of best fit

1 **a** False

b True

c True

2 **a** True

b False

c False

Answers

Chapter 1 The force due to gravity

1.1 Newton's law of universal gravitation

WE 1.1.1 7.1×10^{-9} N **WE 1.1.2** 2.0×10^{20} N
WE 1.1.3 The acceleration of the Earth is 3.3×10^{5} times greater than the acceleration of the Sun.
WE 1.1.4 The equations give the same result to two significant figures.
WE 1.1.5 **a** 589 N **b** 774 N **c** 1.1×10^{3} N

1.1 Review

1 The force of attraction between any two bodies in the universe is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.
2 *r* is the distance between the centres of the two objects.
3 1.8×10^{21} N **4** 2.8×10^{-3} m s^{-2}
5 **a** 3.0×10^{16} N **b** 3.4×10^{22} N
c 0.000088% of the Sun–Earth force
6 The Moon has a smaller mass than the Earth.
7 3.5 m s^{-2} **8** 240 N
9 On Earth, weight is the gravitational force acting on an object near the Earth's surface whereas apparent weight is the contact force between the object and the Earth's surface. In an elevator accelerating upwards, the apparent weight of an object would be greater than its weight.
10 **a** 550 N **b** 490 N
11 299 N **12** D **13** D **14** A **15** B

1.2 Gravitational fields

WE 1.2.1 **a** See Reader+ **b** From strongest to weakest: B, C, A
WE 1.2.2 9.6 N kg^{-1} **WE 1.2.3** 9.75 N kg^{-1} **WE 1.2.4** 3.7 N kg^{-1}

1.2 Review

1 N kg^{-1} **2** 9.3 N kg^{-1} **3** It is one-ninth of the original.
4 **a** 5.67 N kg^{-1} **b** 1.48 N kg^{-1} **c** 0.56 N kg^{-1}
d 0.22 N kg^{-1}
5 0.0008 N kg^{-1} or 8×10^{-4} N kg^{-1}
6 Mercury: 3.7 N kg^{-1}
Saturn: 10.4 N kg^{-1}
Jupiter: 24.8 N kg^{-1}
7 2×10^{12} N kg^{-1}
8 The gravitational field strength at the poles is 1.4 times that at the equator.
9 3.4×10^{8} m **10** 10 Earth radii

1.3 Work in a gravitational field

WE 1.3.1 3.1×10^{3} MJ **WE 1.3.2** 5.2 m s^{-1} **WE 1.3.3** 5.2×10^{8} J
WE 1.3.4 5.4×10^{9} J

1.3 Review

1 C **2** *g* increases. **3** It will accelerate at an increasing rate.
4 A, B, C **5** 2.0×10^{12} J **6** 292 m s^{-1}
7 **a** *F* is between 9 N and 9.2 N. **b** 2.6×10^{6} m or 2600 km
8 **a** 8×10^{6} J **b** 1.9×10^{7} J **c** 2.7×10^{7} J
d 7348 m s^{-1} or 7.3 km s^{-1}
9 1.7×10^{9} J **10** 2.6×10^{11} J

Chapter 1 review

1 730 N **2** 3.78×10^{8} m **3** 2.1×10^{-7} m s^{-2}
4 **a** The forces are equal.
b The acceleration of Jupiter caused by the Sun is greater than the acceleration of the Sun caused by Jupiter.
5 3.7 m s^{-2} **6** **a** 460 N **b** 490 N
7 **a** 2.48×10^{4} N **b** 2.48×10^{4} N **c** 24.8 m s^{-2}
d 1.31×10^{-23} m s^{-2}
8 D **9** **a** D **b** B **c** C **d** A **e** D
10 The direction of the arrowhead indicates the *direction* of the gravitational force and the space between the arrows indicates the *magnitude* of the field. The field lines always point towards the source of the field.
11 9.76 N kg^{-1} **12** **a** 9.79 N kg^{-1} **b** 100.61%
13 **a** 11.1 N kg^{-1} **b** C **14** 16
15 **a** 3×10^{7} J **b** 4×10^{7} J **c** 2000 m s^{-1} or 2 km s^{-1}
d 3.5 N kg^{-1}
16 9 N kg^{-1} **17** D **18** C **19** 3.5×10^{9} J
20 No. Air resistance will play a major part as the satellite re-enters the Earth's atmosphere.

Chapter 2 Motion in a gravitational field

2.1 Inclined planes

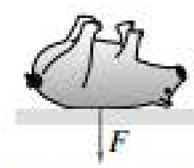
WE 2.1.1 **a** 285 N **b** 783 N **c** 3.4 m s^{-2}

2.1 Review

1 **a**

Force exerted on racquet by ball

b



Force exerted on ground by pig

c

Force exerted on ground by wardrobe

d

Gravitational force of attraction that seal exerts on Earth

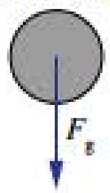
2 **a** A **b** C **c** 490 N up the hill. **d** 4.90 m s^{-2}
e Acceleration is not affected by mass if there is no friction.
3 A, B, D **4** A
5 **a** 849 N **b** 490 N **c** 4.90 m s^{-2}
6 **a** Ball 1: 0.888 N Ball 2: 1.78 N
b Ball 1: 4.14 m s^{-2} Ball 2: 4.14 m s^{-2}
c Ball 1: 0.335 N Ball 2: 0.670 N
d Ball 1: 9.21 m s^{-2} Ball 2: 9.21 m s^{-2}
e For (a) the normal force of ball 2 is double that of ball 1. This is because the normal force is directly proportional to mass. For (b) the two balls have the same acceleration, indicative of the fact that acceleration down an inclined plane depends on the angle of the plane, not the mass. Comparing (a) to (c) it can be seen that by increasing the angle of the inclined plane, the normal force acting on the two objects decreases as a function of the angle. You can think of this as the steeper the inclined plane, the closer the object is to free-fall and therefore apparent weightlessness ($F_N = 0$). Finally, comparing (b) to (d) it can be seen that increasing the angle of the inclined plane increases the acceleration of the objects; however, the acceleration of the two objects remains equal.

2.2 Projectiles launched horizontally

WE 2.2.1 **a** 2.47 s **b** 49.4 m
c 31.4 m s^{-1} at 50.4° below the horizontal
WE 2.2.2 9.74 m s^{-1}

2.2 Review

1 A, D **2** **a** 1.5 m **b** 7.35 m s^{-1} **c** 7.6 m s^{-1}
3 **a** 9.80 m s^{-2} down **b** 6.29 m s^{-1}
4 **a** 1.0 s **b** 20 m **c** 9.80 m s^{-2} down **d** 21.5 m s^{-1}
e 22.3 m s^{-1}
5 **a** 22.7 m s^{-1} **b** 45.6 m s^{-1} **6** B, C
7 The hockey ball travels further. A polystyrene ball is much lighter and is therefore more strongly affected by air resistance than the hockey ball.
8 **a** 0.64 s **b** 0.64 s **c** 3.2 m **9** **a** 54 m s^{-1}
b $\theta = 22°$
10 **a** 10 m s^{-1} forwards **b** 4.4 m s^{-1} down
c 10.9 m s^{-1} at 24° below the horizontal
d 0.45 s **e** 4.5 m **f**



2.3 Projectiles launched obliquely

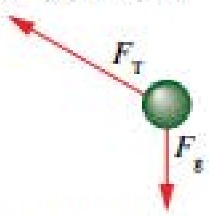
WE 2.3.1 **a** 6.11 m s^{-1} horizontally to the right **b** 0.25 m **c** 0.45 s
WE 2.3.2 **a** 2.25 m s^{-1} to the right **b** 0.366 m
c 3.50 m s^{-1} at 50.0° below the horizontal

2.3 Review

1 B **2** 45° **3** **a** 8.83 m s^{-1} **b** 0.989 m
4 **a** 13.6 m s^{-1} **b** 6.34 m s^{-1} **c** 9.80 m s^{-2} down **d** 13.6 m s^{-1}
5 **a** 4.0 m s^{-1} **b** 6.9 m s^{-1} **c** 0.70 s **d** 3.9 m **e** 4.0 m s^{-1}
6 **a** 24.2 m s^{-1} **b** 24.2 m s^{-1} **c** 24.2 m s^{-1}
7 **a** 14.0 m s^{-1} **b** 4.20 m s^{-1} **c** 5.60 m s^{-1} down
8 24.8 m s^{-1} **9** 69.2 m **10** C **11** 16.2 m s^{-1}

2.4 Circular motion in a horizontal plane

WE 2.4.1 7.5 km h^{-1}
WE 2.4.2 **a** 521 m s^{-2} **b** 3.6×10^3 N
WE 2.4.3 **a** 1.53 m **b**



c 2.34 N towards the left **d** 3.05 N

2.4 Review

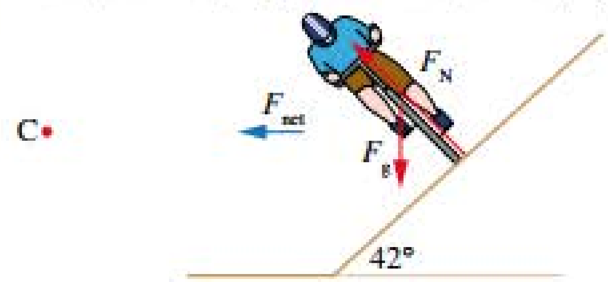
1 B **2** 0.2 s **3** A, D
4 **a** 8.0 m s^{-1} **b** 8.0 m s^{-1} south **c** 7.0 m s^{-2} west
5 8.4×10^3 N west **6** **a** 8.0 m s^{-1} north **b** east
7 The force needed to give the car a larger centripetal acceleration will eventually exceed the maximum frictional force that could act between the tyres and the road surface. At this time, the car would skid out of its circular path.
8 **a** 2.67 m s^{-2}
b The skater has an acceleration, so forces are unbalanced.
c 135 N
9 **a** 28 s **b** 5.0 N **10** **a** 0.5 s **b** 10 m s^{-1}
c 125 m s^{-2} **d** 310 N
11 **a** 1.2 m
b The forces are her weight acting vertically and the tension in the rope acting along the rope towards the top of the maypole.
c Acceleration is directed towards point B, the centre of her circular path.
d 170 N towards B **e** 2.6 m s^{-1}

2.5 Circular motion on banked tracks

WE 2.5.1 **a** 590 N towards the centre of the circle **b** 17 m s^{-1}

2.5 Review

1 towards the centre of the circle
2 The architect could make the banking angle larger or increase the radius of the track.
3 It will travel higher up the banked track because the greater speed means that a greater radius is required in the circular path.
4 friction, normal, weight, balanced, normal, weight
5



6 48 km h^{-1}
7 **a** 640 N
b On a horizontal track, F_N is equal and opposite to the weight force, so $F_N = mg$ = 539 N. This is less than the normal force on the banked track (643 N).
8 47° **9** **a** 4.9 kN **b** 22°
10 A greater radius will make the car travel higher up the banked track. The driver would have to turn the front wheels slightly towards the bottom of the bank.

2.6 Circular motion in a vertical plane

WE 2.6.1 **a** 4.85 m s^{-1} **b** 15.6 N up **c** 3.70 m s^{-1}
d 6.73 N down

2.6 Review

1 **a** It is constant in magnitude.
b at the bottom of its path
c at the top of its path
d at the bottom of its circular path
2 3.8 m s^{-1}
3 **a** weight force from gravity and the normal force from the road
b 1280 N (or 1.3×10^3 N)
c Yes. When the car is moving over a hump, the normal force on the driver is less than her weight mg. Her apparent weight is given by the normal force that is acting and so the driver feels lighter at this point.
d 36 km h^{-1}
4 **a** 31.4 m s^{-1} **b** 19.9 m s^{-1} **c** 8300 N down
5 12.1 m s^{-1} **6** 196 N down **7** 31.3 m s^{-1}
8 188 m s^{-1} **9** **a** 18 m s^{-2} up **b** 1530 N up
10 **a** 9.80 m s^{-2} down **b** 2.2 m s^{-1}
11 **a** 3.6×10^3 N down **b** 1.3×10^3 N up **c** 7.0 m s^{-1}

2.7 Satellite motion

WE 2.7.1 3.08×10^3 m s^{-1}
WE 2.7.2 **a** 6.70×10^5 km **b** 1.90×10^{27} kg **c** 8.20 km s^{-1}

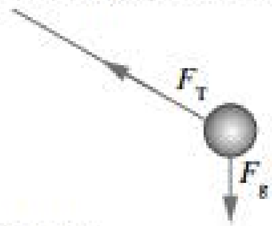
2.7 Review

1 C **2** D
3 **a** 0.22 m s^{-2} **b** 506 N (or 510 N to 2 significant figures)
4 15.6 days

Chapter 2 review

1 B **2** **a** 4.9 m s^{-2} **b** $0.87F_g$
3 **a** 4.9 m s^{-2} **b** 4.9 m s^{-1}
4 **a** 236 N
b 8.88 m s^{-2} down the ramp **c** 506 N down the ramp
d 9.4 m s^{-1} **e** 506 N up the ramp

5 **a** 508 N **b** 137 N
c The reaction to the force of the slide on the teenager due to friction is the force of the teenager on the slide.
d The reaction of the weight force on the teenager is the force of gravitational attraction from the teenager on the Earth.
6 **a** 2.5 m **b** 9.80 $m s^{-2}$ downwards
7 **a** 10 $m s^{-1}$ **b** 4.4 $m s^{-1}$ **c** 11 $m s^{-1}$ (speed only so no direction required)
8 **a** 10.3 $m s^{-1}$ **b** 12.3 $m s^{-1}$ **c** 8.9 m
9 Tait is correct.
10 **a** 1.8 J **b** 1.96 J **c** 8.7 $m s^{-1}$
11 392 J **12 a** 33.0 J **b** 33.0 J **c** 21.4 m
13 **a** 3.70 $m s^{-1}$ **b** 17.1 $m s^{-2}$ towards the centre of the circle
c 0.430 N (size only needed)
14 **a**



b 0.49 N
15 **a** 2.5 $m s^{-2}$ towards the centre of the circle
b The centripetal force is created by the friction between the tyres and the ground.
16 **a** 1.02×10^3 $m s^{-1}$ **b** 1.99×10^{20} N
17 3.40×10^{-2} $m s^{-2}$
18 **a** 10 $m s^{-1}$ south **b** 10 $m s^{-1}$
c 13 s **d** 5.0 $m s^{-2}$ west **e** 7.5×10^3 N west
19 A **20 a** 15.7 $m s^{-1}$
b A frictional force will be acting up the plane.
c 740 N
d No. Once Joe's speed is greater than the design speed, there will be friction contributing to his centripetal acceleration. He would need to speed up and overcome this friction before he slid off the track.
21 **a i** 365 N up **ii** 615 N up **b** D
22 The forces are the force of gravity (weight) and the normal force from the base of the bucket on the water. Both act downwards. The acceleration of the water is towards the centre of the circle, i.e. downwards, and is greater than the acceleration due to gravity.
23 D **24** D **25** A
26 11.2T
27 **a** 0.0540 $m s^{-2}$ **b** 4.38×10^3 $m s^{-1}$ **c** 5.89 days
28 **a** 0.315 $N kg^{-1}$ **b** 344 $m s^{-1}$

Chapter 3 Equilibrium of forces

3.1 Torque

WE 3.1.1 39.5 N m anticlockwise
WE 3.1.2 Yes. The minimum perpendicular distance required is 62.1 cm.
WE 3.1.3 17.1 N m clockwise **WE 3.1.4** 17.1 N m clockwise

3.1 Review

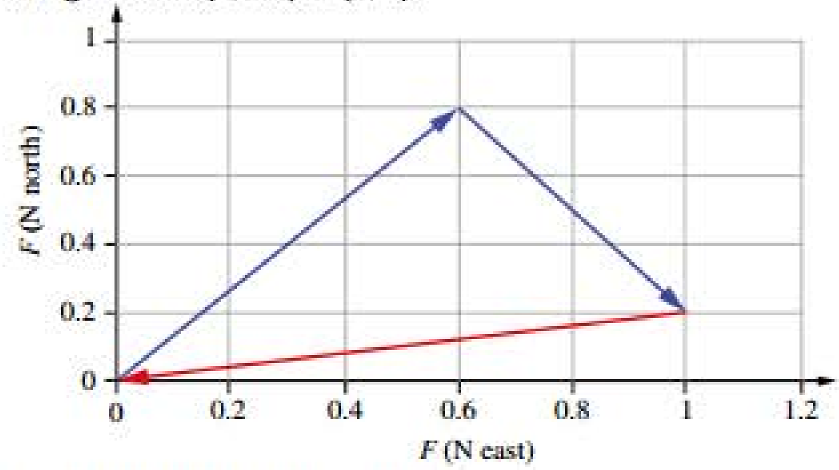
1 A **2** B
3 5.00 N m in the direction in which the spanner is being turned.
4 0.289 m **5** D **6** 1.5×10^4 N m **7** 1.34 m
8 38.6 N m **9** 0.595 N m **10** 56.5 N m
11 0.750 m **12** 9.38 N m clockwise

3.2 Equilibrium of forces

WE 3.2.1 2.9×10^4 N up **WE 3.2.2** 8490 N

3.2 Review

1 D **2** D **3** 3.48 N upwards **4** 0.765 kg
5 29.4 N **6** 911 N
7 The force that causes the trolley to be in equilibrium is the force that goes from (1, 0.2) to (0, 0).



8 Cable A: 1130 N or 1.13×10^3 N
Cable B: 565 N
9 5.25 kg **10** 2120 N

3.3 Static equilibrium

WE 3.3.1 a 1.96 N up **b** 0.300 m or 30.0 cm
WE 3.3.2 Around reference point Y (the position of the boy), the clockwise torque due to the girl's position on the plank is equal to the anticlockwise torque due to the pivot's action on the plank. So the plank is in rotational equilibrium.
WE 3.3.3 555 N
WE 3.3.4 36.8 N downwards on the beam
WE 3.3.5 52.0 N

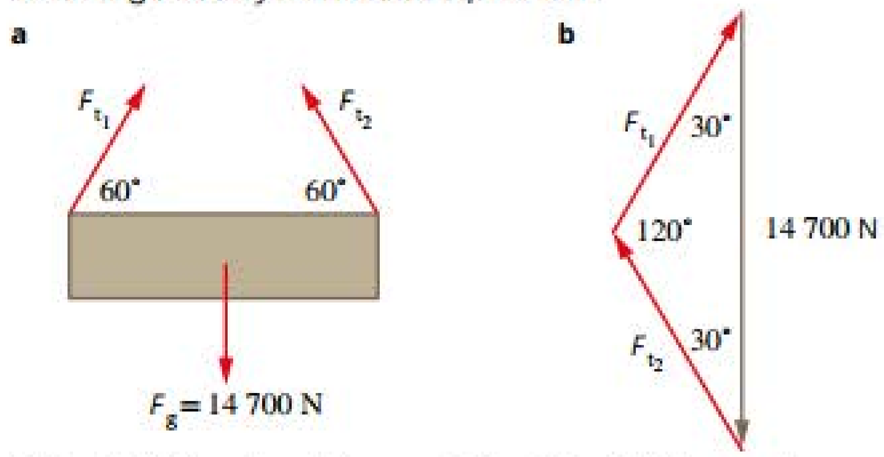
3.3 Review

1 C **2** 0.750 m **3** B **4** 0.889 m
5 340 N **6** 2990 N **7** 9.8×10^3 N
8 downwards **9** 980 N

Chapter 3 review

1 D **2** B **3** C, D **4** 3.60 N m
5 86.5 N m
6 Stability refers to the ability of an object to restore its original static equilibrium after being slightly displaced by an outside force.
An object is in unstable equilibrium if the smallest displacement is sufficient to produce a force or torque that continue to make it move away from equilibrium.
An object is in stable equilibrium if the object returns to its original position once the outside force is removed.
An object is in neutral equilibrium if small outside forces don't create any unbalanced forces or torques to move the object further. The object remains in its new equilibrium position.
7 The 75.0 cm spanner; the longer the lever arm the less force is needed.
8 33.8 N m
9 **a** the spindle of the tap; about 3 cm
b front wheel axle of the wheelbarrow; 1 m (handle plus barrow through to the axle)
c the end of the tweezers, usually a few centimetres
d the point of contact between screwdriver and the rim of the tin, 15–30 cm being the length of the screwdriver
10 4.9×10^5 N m
11 The crane must have a counterweight providing a torque in the opposite direction.
12 A, B, D

13 Diagram (a) shows the forces acting on the beam. Diagram (b) shows these forces added as a vector diagram (head to tail) illustrating that they are in static equilibrium.



14 8.49×10^3 N in each cable 15 2.9×10^4 N upwards
16 109 N upwards 17 848 N 18 F_L = 141 N, F_R = 123 N
19 15.5 cm from the model of the Sun

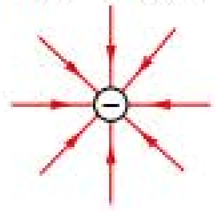
Chapter 4 Electric fields

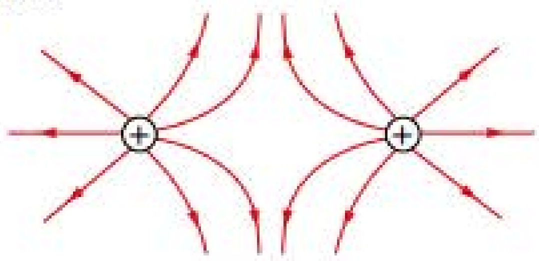
4.1 Electric fields

WE 4.1.1 5.62×10^{-4} N C^{-1}

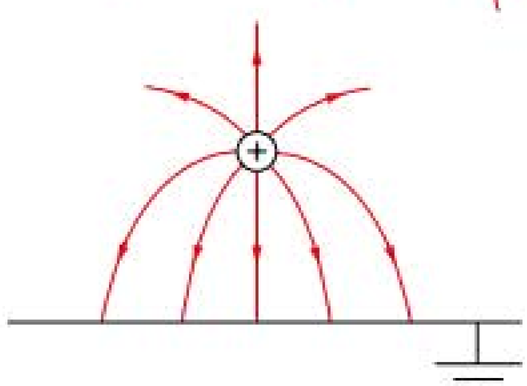
4.1 Review

1 C 2 B
3 **a** True **b** False **c** False **d** True
e True **f** False **g** False
4 **a** **b**





c



5 1.25×10^{-2} N 6 1.39 mC 7 5.72×10^{11} m s^{-2}

4.2 Coulomb's law

WE 4.2.1 6.32×10^{-4} N repulsion
WE 4.2.2 $q_1 = +6.35 \times 10^{-10}$ C, $q_2 = -6.35 \times 10^{-10}$ C
WE 4.2.3 8.0×10^5 N C^{-1} away from the charge or to the right

4.2 Review

1

Force	q_1 charge	q_2 charge	Action
a positive	positive	positive	repulsion
b negative	negative	positive	attraction
c positive	negative	negative	repulsion
d negative	positive	negative	attraction

2 D 3 -8.22×10^{-8} N 4 1.1×10^7 N C^{-1}
5 9000 N 6 1.435 m
7 **a** double, repel **b** quadruple, repel **c** double, attract
d quadruple, repel
8 37 N 9 1.97×10^{13} electrons

4.3 Work done in an electric field

WE 4.3.1 2.16×10^{-18} J on the field

4.3 Review

1 1200 V 2 20 electrons
3 **a** Work is done by the field. **b** No work is done.
c Work is done on the field. **d** No work is done.
e Work is done on the field. **f** Work is done by the field.
4 **a** 1.09×10^{-19} J **b** Work is done on the field.
5 2.23×10^{-6} C

Chapter 4 review

1 0.0225 N 2 D
3 The electrical potential is defined as the work done per unit charge to move a charge from infinity to a point in the electric field. The electrical potential at infinity is defined as zero. When you have two points in an electric field (*E*) separated by a distance (*d*) that is parallel to the field, the potential difference *V* is then defined as the change in the electrical potential between these two points.
4 25 V 5 C 6 field, charged particle
7 4.17×10^{-18} J 8 2×10^{-14} N
9 **a** quarter , repel **b** quadruple, repel **c** halve, attract
10 5.42×10^4 m s^{-1} 11 0.045 N 12 45.8 m
13 $+1.63 \times 10^{-4}$ C 14 7.9×10^{-6} N
15 **a** 8.80×10^{-14} N upwards **b** 24.8 V **c** 3.96×10^{-15} J
16 **a** 1.80×10^5 N attraction **b** -5.5×10^{-6} C per sphere
c 5.98×10^3 N repulsion
17 **a** 1.28×10^{-15} J **b** 8.00×10^3 V
18 6.80×10^{-6} N

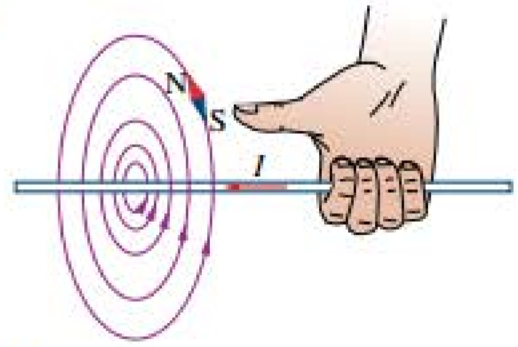
Chapter 5 Magnetic field and force

5.1 The magnetic field

WE 5.1.1 anticlockwise around the wire
WE 5.1.2 2.67×10^{-5} T

5.1 Review

1 B 2 A 3 C 4 C
5



6 A
7 **a** A = east, B = south, C = west, D = north
b A = west, B = north, C = east D = south
8 2.25 A 9 62.5 A 10 1.91 cm

5.2 Forces on charged objects

WE 5.2.1 4.8×10^{-22} N
WE 5.2.2 out of the screen (towards you)

5.2 Review

1 D
2 **a** south (S) **b** path C **c** It remains constant.
d path A **e** particles with no charge, e.g. neutrons
3 D 4 0 N 5 4.8×10^{-24} N south
6 2*F* north
7 Charged particles experience a force from the magnetic field that is perpendicular to the particle's velocity, constantly accelerating the charged particle towards the centre. Thus the magnetic force provides the centripetal force.
8 10.0 m s^{-1} 9 0.0500 T
10 It would need to be halved

5.3 The force on a conductor

WE 5.3.1 2.5×10^{-3} N per metre of power line
WE 5.3.2 vertically downwards
WE 5.3.3 **a** 0 N **b** 1.0×10^{-3} N out from Santa's house
WE 5.3.4 1.6×10^{-7} N m clockwise

5.3 Review

1 No force will apply. **2** 0.4 N upwards
3 **a** 0.18 N downwards **b** Same as (a). **4** 0.1 N
5 Current flows into brush P and around the coil from V to X to Y to W. So force on side VX is down, force on side YW is up, so rotation is anticlockwise.
6 D **7** **a** down the page **b** up the page
8 anticlockwise
9 **a** down the page **b** up the page
c Zero torque acts as the forces are trying to pull the coil apart rather than turn it. The force is parallel to the coil, rather than perpendicular to it.
10 C

Chapter 5 review

1 east—away from the north pole of the left-hand magnet
2 west—away from the north pole of the right-hand magnet
3 The point is equidistant from the two magnets so resulting field would be zero.
4 *B*, into the page **5** 3*B*, into the page
6 zero **7** 1.50 cm **8** 20 μT **9** 75 A
10 D **11** **a** palm **b** fingers **c** thumb
12 equal to, into **13** 2.78 A
14 **a** 5.0×10^{-9} N into the page
b 2.0×10^{-3} N into the page
15 9.6×10^{-15} N **16** The force is downwards.
17 The east–west line because it runs perpendicular to the Earth's magnetic field.
18 2.0×10^{-4} N north **19** A
20 1.0×10^{-2} N into the page **21** 1.0×10^{-2} N out of the page
22 0 N **23** anticlockwise **24** D
25 2.0×10^{-4} N m
26 It reverses the direction of the current in the coil every half turn, to keep the coil rotating in the same direction.

Chapter 6 Magnetic field and emf

6.1 Induced emf in a conductor moving in a magnetic field

WE 6.1.1 No, emf = 0.695 V. This is a very small emf and would not be dangerous.
WE 6.1.2 0.500 m s^{-1}
WE 6.1.3 6.02×10^{-4} V

6.1 Review

1 A **2** 0.0144 V or 1.44×10^{-2} V **3** 0.842 m s^{-1}
4 0.100 m **5** 0 V **6** 0.500 V **7** 3.13×10^{-4} V
8 0.624 V **9** 0.614 V **10** 0.0542 V, 0.668 V

6.2 Induced emf from a changing magnetic flux

WE 6.2.1 8.00×10^{-5} Wb
WE 6.2.2 **a** 5.00×10^{-4} Wb **b** 5.00×10^{-3} V
WE 6.2.3 1000 turns

6.2 Review

1 0 Wb **2** 1.30×10^{-5} Wb **3** 1.20×10^{-6} Wb
4 Zero flux **5** 3.00×10^{-5} V **6** 4.00×10^{-3} V
7 2.00 V **8** 6.00×10^{-3} V **9** 0.0100 m^2
10 0.125 s

6.3 Lenz's law

WE 6.3.1 clockwise when viewed from above
WE 6.3.2 **(i)** through the solenoid from Y to X (through the meter from X to Y)
(ii) no induced emf or current
(iii) through the solenoid from X to Y (through the meter from Y to X)
WE 6.3.3 anticlockwise

6.3 Review

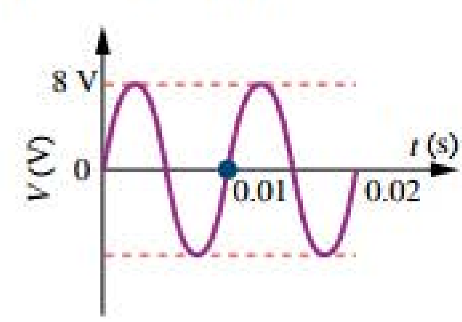
1 C **2** **a** A **b** A
3 **a** anticlockwise
b Any combination of:
1 strength of the magnet
2 speed of the magnet
3 area/diameter of the ring
4 orientation of the ring
5 type of copper making up the ring
6 resistance of the circuit containing the coil.
4 Lenz's law states that an induced current is created, which produces flux to oppose the change in flux. As the copper wire accelerates through the magnetic field, the induced opposing magnetic field produces a repulsive upwards force that reduces the net force acting on the wire. When the upwards magnetic force is equal to the downwards gravitation force, there is no net force and a constant terminal velocity is reached.
5 The compass will oscillate freely above the glass but damping will occur over the aluminium due to induced eddy currents in the aluminium that create a field to oppose the change of flux, according to Lenz's law.
6 top end **7** anticlockwise
8 up at the front of the solenoid
9 into the page
10 One of:
- Induction cookers**:** AC current in coil produced changing flux resulting in eddy currents in metal pans (best with ferromagnetic bases), currents will heat up pan which transfers heat to its content.
- Soft-closing kitchen drawers also use eddy currents created by a magnet on the drawer inducing an eddy current in a metal plate. This causes an induced field that opposes the motion of the magnet on the drawer, causing it to slow down as it closes.
- Melting gold at the Perth Mint is done without the need for fuel and flames by using large alternating voltages that create eddy currents in the gold. The eddy currents heat the gold to beyond its melting point.

6.4 Transforming voltage using changing magnetic field

WE 6.4.1 0.011 m^2 **WE 6.4.2** 2000 W
WE 6.4.3 4000 turns **WE 6.4.4** 0.0125 A
WE 6.4.5 3.00 W **WE 6.4.6** 3.60×10^5 W or 0.360 MW
WE 6.4.7 500.6 kV

6.4 Review

1 B **2** 5.66 V
3



4


5 D 6 40 turns

7 a $P_1 = P_2$ b $\frac{I_2}{I_1} = \frac{N_1}{N_2}$

8 a 80.0 V b 16.0 W c 0.200 A 9 400 W

10 a 5000 A b 100 – 10 = 90 kV

Chapter 6 review

1 3.20×10^{-6} Wb

2 The magnetic flux decreases from 3.20×10^{-6} Wb to 0 after one-quarter of a turn. Then it increases again to 3.20×10^{-6} Wb through the opposite side of the loop after half a turn. Then it decreases to 0 again after three-quarters of a turn. After a full turn it is back to 3.20×10^{-6} Wb again.

3 C

4 The student must induce an emf of 1.00 V in the wire by somehow changing the magnetic flux through the coil at an appropriate rate. A change in flux can be achieved by changing the strength of the magnetic field or by changing the area of the coil. The magnetic field can be changed by changing the position of the magnet relative to the coil. The area can be changed by changing the shape of the coil or by rotating the coil relative to the magnetic field. See Reader+ for fully worked solution.

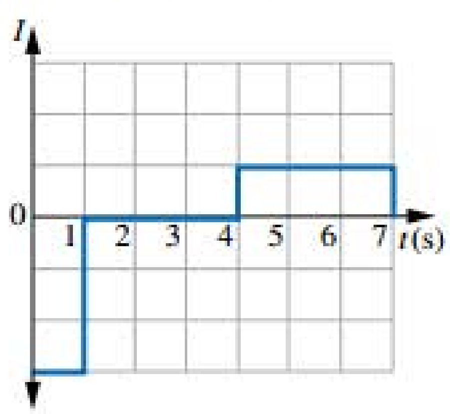
5 B 6 30.0 W 7 3.54 A

8 a 3.20×10^{-3} V b clockwise

9 a 0.0402 V
b from Y to X; clockwise around the coil when viewed from above

10 a 4.00×10^{-3} V or 4.00 mV b from X to Y

11 1.60×10^{-3} V or 1.60 mV

12


Either the graph shown or its inversion is correct.

13 1.00 A 14 10 turns 15 A

16 a 18.0 V b 375 W 17 option C

18 8.91 V

19 Doubling the frequency doubles the rms emf, since the rate of change of flux is doubled.

20 Any two of:
- Using a DC power supply means that the voltage cannot be stepped up or down with transformers.
- There will be significant power loss along the 8 Ω power lines.
- There could be damage to any appliances operated in the shed that are designed to operate on 240 V AC and not on 240 V DC.

21 As the coil area is reduced, the flux into the page will decrease. To oppose this, the induced current will try to increase the flux again in the same direction. Using the right-hand grip rule the direction of the induced current will be clockwise.

22 AB and CD 23 15.0 A 24 9970 V

25 450 W 26 This results in a 30% power loss—a bad idea!

27 anticlockwise 28 B

29 a 40 turns b 0.141 A c 24.0 W d A

30 4.00×10^{7} W or 40.0 MW

31 B. A is incorrect because the *V* in the formula indicates the voltage drop in the transmission lines; it does not refer to the voltage being transmitted.

Unit 3 review

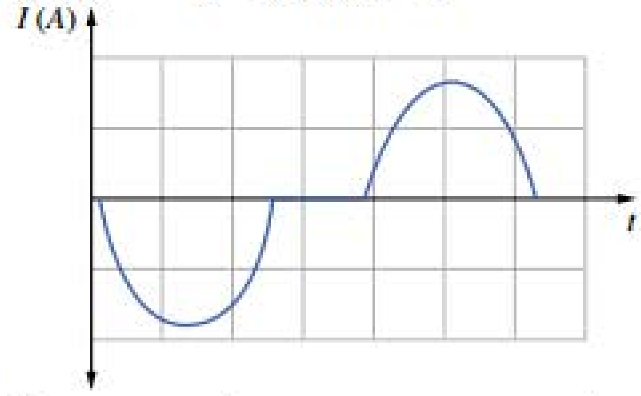
Section 1: Short response

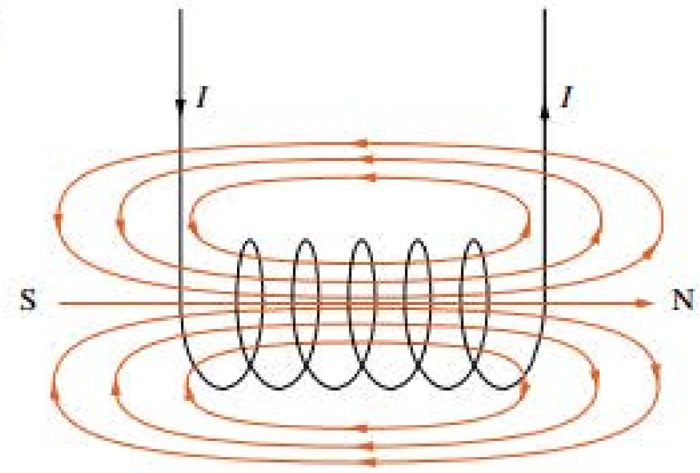
1 3.22×10^{-6} N 2 4340 days

3 a 8.28 m s^{-1} b 4.58×10^{2} J

4 a 35.0 m s^{-1} at 40.0° down from horizontal. b 42.6 m s^{-1}

5 2.57 N 6 9.38×10^{-2} m

7


8 a


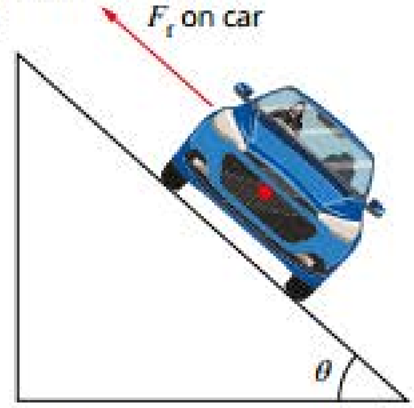
b increasing the current flowing through the coil
increasing the number of turns in the coil
placing a bar of iron inside the coil

c Currents are in opposite directions, so the two magnetic fields will be in opposite directions, so they will attract each other.

Section 2: Problem solving

9 a


b 45.6° c 11.2 N

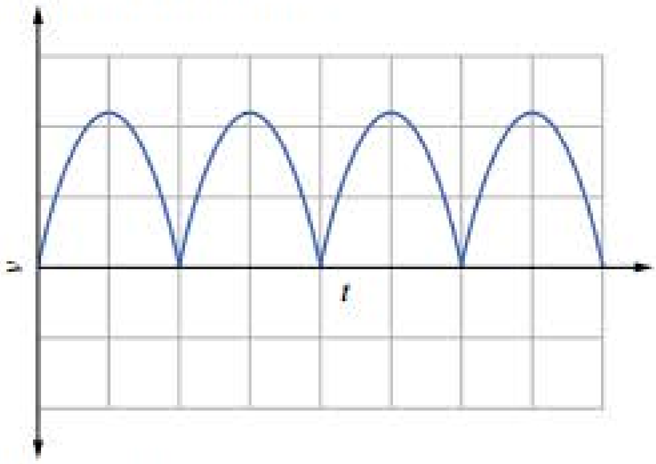
d F_f on car


e 3.83 m s^{-1}

10 a A geostationary orbit is the orbit of a satellite around Earth for which the period of the orbit is exactly one day, and so appears to be stationary in the sky.
b 3.58×10^{7} m c 3.88×10^{3} m s^{-1} d 15 m

11 a 1.80×10^{-2} N upwards

b



c 24.0×10^{-3} V

d Slip rings allow the coil to maintain constant contact with the circuit, and so will have current flowing in a constant direction. Side AB will experience an upwards force, causing the coil to rotate clockwise. Once the coil has rotated by 180°, the side CD will be where AB previously was; however, the current will be flowing in the opposite direction, since the coil has flipped. Side CD will then experience a downwards force. The coil will then oscillate back and forth between each position until it likely reaches an equilibrium position turned halfway.

12 a With more appliances on, the amount of current required would increase, increasing the power loss in the transmission line as $P_{loss} = I^2R$. This power loss would cause a significant drop in potential across the transmission lines according to $\Delta V = IR$.

b $V_{house} = 42.0$ V, $P_{house} = 672$ W

c At the generator end, a step-up transformer is needed, with a turns ratio of 1:20 to convert 250 V up to 5000 V.
At the house end of the line, a step-down transformer is needed, with an approximate turns ratio of 20:1 to convert the approximately 5000 V down to 250 V again.

d 0.800 A **e** 10.4 V **f** 8.32 W **g** 4989.6 V **h** 3991.7 W

i Without transformers: efficiency = 16.8%
With transformers: efficiency = 99.8%.

j With a higher transmission voltage, there is a corresponding decrease in current in the transmission lines. As $P_{loss} = I^2R$, any decrease in current has a significant effect on the power loss.

Section 3: Comprehension

13 a The Doppler spectroscopy method relies on detecting the Doppler shift of the star in order to determine its velocity, so the star must be moving either towards or away from Earth. Thus, a side view will be best because it has all of the star's velocity in the radial direction towards Earth. A top view would give no information about the velocity of the star.

b i In a general stellar system in which the plane of the system is not exactly side-on to Earth, some of the velocity of the star will be perpendicular to the line-of-sight of an astronomer. This velocity is undetectable to observers on Earth. Therefore, the radial velocity detected will only be one component of the total velocity of a star, and its true velocity will be greater than this detected velocity. From the formula $\frac{M_{planet}}{v_{star}} = \frac{M_{star}}{v_{planet}}$, it can be seen that the mass of the planet, $M_{planet} = \frac{v_{star}M_{star}}{v_{planet}}$, is directly proportional to the velocity of the star. Since observations from Earth using this method provide only a lower bound for the star's velocity, they will only provide a lower bound for the planet's mass, with the true quantity determined by the star's true velocity.

ii In order to observe the transit of a planet in front of its local star, the star system must be side-on to observers on Earth. In light of the answers to the previous two questions, one would expect the measurement of the star's velocity to be close to its true velocity, and therefore for these mass calculations to be accurate.

c 2.39×10^{24} kg

Chapter 7 Wave–particle duality and the quantum theory

7.1 Properties of waves in two dimensions

WE 7.1.1 1.52
WE 7.1.2 1.62×10^8 m s^{-1}
WE 7.1.3 28.2°
WE 7.1.4 The light speed is increasing, therefore the angle is refracted way from the normal, as in Figure 7.1.10 on page 236. There will be a significant change in angle and corresponding increase in wavelength.
WE 7.1.5 24.4°

7.1 Review

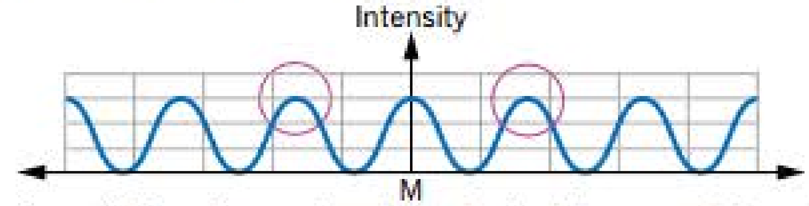
1 **a** wave model **b** wave model **c** particle model
2 C **3** slower than **4** 2.17×10^8 m s^{-1}
5 1.31 **6** 35.3° **7** b and c **8** D
9 Polarisation is a phenomenon in which transverse waves are restricted in their direction of vibration. Polarisation can only occur in transverse waves and cannot occur in longitudinal waves. Since light can be polarised, it must be a transverse wave.

7.2 Interference: Further evidence for the wave model of light

WE 7.2.1 550 nm

7.2 Review

1 D **2** C, D **3** A, D
4



5 Up until Young's experiment, most scientists supported a particle or 'corpuscular' model of light. Young's experiment demonstrated interference patterns, which are characteristic of waves. This led to scientists abandoning the particle theory and supporting a wave model of light.
6 **a** increase **b** decrease **c** increase
7 2610 nm or 2.61×10^{-6} m
8 **a** destructive **b** constructive **c** destructive
9 1400 nm **10** 455 nm

7.3 Electromagnetic waves

WE 7.3.1 5.0×10^{14} Hz

7.3 Review

1 B **2** D **3** D
4 X-rays, visible light, infrared radiation, FM radio waves
5 **a** 4.57×10^{14} Hz **b** 5.09×10^{14} Hz **c** 6.17×10^{14} Hz **d** 7.56×10^{14} Hz
6 0.07% **7** 500 nm **8** 4.3 m
9 1.5×10^{18} Hz **10** 0.122 m

7.4 Light quanta: Blackbody radiation and the photoelectric effect

WE 7.4.1 2.4×10^{-19} J **WE 7.4.2** 1.5 eV
WE 7.4.3 5.0 eV **WE 7.4.4** 2.08 eV

7.4 Review

1 **a** False **b** False **c** True **d** True **e** False
2 1 Molecules vibrate at fixed frequencies or energies.
2 To change energy levels, a molecule needs to absorb or emit a photon of exactly the right energy difference between two levels.
3 **a** 3.03×10^{-19} J **b** 3.38×10^{-19} J **c** 4.09×10^{-19} J **d** 5.01×10^{-19} J
4 In the photoelectric effect, a metal surface may become positively charged if light shining on it causes electrons to be released.

5 **a** True **b** False **c** True
6 **a** 4.1 eV **b** 4.6 eV **c** 6.2 eV
7 D **8** 0.066 eV **9** 0.25 eV **10** C, D
11 **a** True **b** False **c** True **d** True
12 1.68 eV

7.5 Atomic spectra

WE 7.5.1 103 nm, Lyman series
WE 7.5.2 A photon of 6.7 eV corresponds to the energy required to promote an electron from the ground state to the second excited state ($n = 1$ to $n = 3$). The photon may be absorbed.
A photon of 5.0 eV cannot be absorbed.
A photon of 11.0 eV may ionise the mercury atom. The ejected electron will leave the atom with 0.6 eV of kinetic energy.
WE 7.5.3 3.40 eV, 366 nm, ultraviolet

7.5 Review

1 The electrons in a sample become excited when the substance is heated or an electric current flows through it. As the electrons return to their ground state, a photon is emitted.
2 4.0×10^{-19} J **3** 3.0×10^{-6} m
4 **a** light-emitting diode
b light amplification by stimulated emission of radiation
5 675 nm **6** 12.75 eV **7** 9.74×10^{-8} m or 97.4 nm
8 It could not explain high-energy orbits of multi-electron atoms, the continuous emission spectrum of solids and the two close spectral lines in hydrogen that are revealed at high resolution.
9 –0.54 eV
10 The energy difference between $n = 1$ and $n = 3$ is 6.7 eV.
An electron from the electron beam will promote the atomic electron from $n = 1$ to $n = 3$ and lose 6.7 eV. It will then exit the atom with an energy of 0.3 eV.
A photon has exactly 7 eV of energy and needs to give up all of its energy, therefore it cannot promote the electron and will pass straight through.

7.6 The quantum nature of light and matter

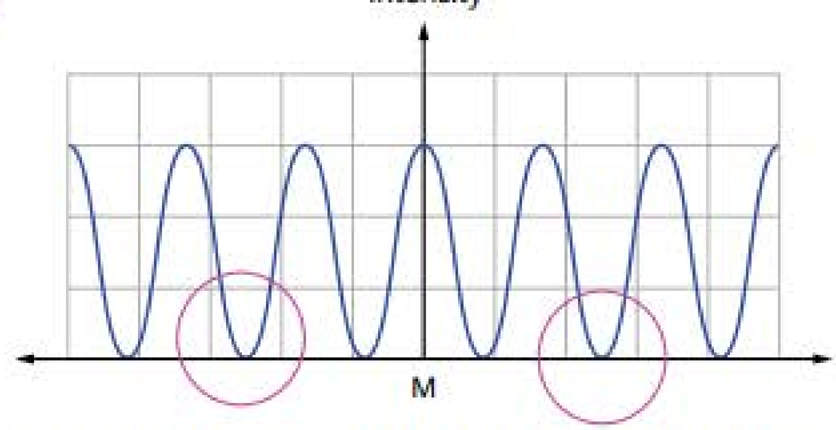
WE 7.6.1 5.7×10^{-13} m **WE 7.6.2** 1.0×10^{-36} m
WE 7.6.3 0.17 nm **WE 7.6.4** 1.47×10^{-27} kg m s^{-1}

7.6 Review

1 7.3×10^{-10} m **2** 1.8×10^{5} m s^{-1} **3** B
4 **a** 3.5×10^{-11} m **b** 2.1×10^{7} m s^{-1}
5 The wavelength of a cricket ball is so small that its wave-like behaviour could not be seen by a cricket player.
6 1.32×10^{5} m s^{-1}
7 $\lambda = \frac{h}{\sqrt{2qVm}}$ **8** $p = \frac{h}{\lambda}$
9 An electron microscope can resolve images in finer detail than an optical microscope because a high-speed electron has a shorter wavelength than a light wave.
10 De Broglie proposed a model in which electrons were viewed as matter waves with wavelengths that formed standing waves within an atomic orbit circumference. A bowed violin string forms standing waves between the bridge of the violin and the violinist's finger.

Chapter 7 review

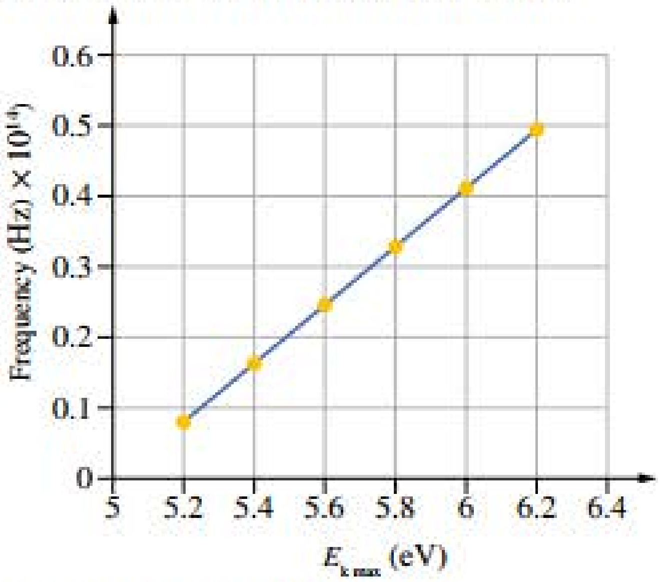
1 A
2 The diffraction pattern for green light would be more spread out.
3 D
4 The light reflected from water and snow is partially polarised. Snowboarders and sailors would benefit from wearing polarising sunglasses as these will absorb the polarised glare from the snow or water respectively.
5 2.25×10^{8} m s^{-1} **6** increases, away from
7 A: incident ray
B: normal
C: reflected ray
D: boundary between media
E: refracted ray
8 2.1×10^{8} m s^{-1} **9** $a = 25.4°$, $b = 25.4°$, $c = 28.9°$
10 **a** 32.0° **b** 53.7° **c** 21.7° **d** 1.97×10^{8} m s^{-1}
11 **a** 19.5° **b** 19.1° **c** 0.4° **d** 1.96×10^{8} m s^{-1}
12 **a** 49.8° **b** 40.5° **c** 27.6°
13 B, D, A, C
14 **a** 581 nm **b** yellow
15



16 radio waves, microwaves, infrared, visible, ultraviolet, X-rays, gamma rays
17 **a** microwaves **b** infrared waves
c X-rays, radio waves (in MRI), infrared radiation, visible light, UV and gamma radiation
18 490 m
19 Young performed his famous experiment in 1803, in which he observed an interference pattern for light. Young shone monochromatic light on a pair of narrow slits. Light passed through the slits and formed a pattern of bright and dark lines/fringes/bands on a screen. He compared this to interference patterns he had observed, and identified that these lines corresponded to regions of constructive and destructive interference. This could only be explained by considering light to be a wave.
20 A microwave oven is tuned to produce electromagnetic waves with a frequency of 2.45 GHz. This is the resonant frequency of water molecules. When food is bombarded with radiation at this frequency, the water molecules within the food start to vibrate. The energy of the water molecules is then transferred to the rest of the food, heating it up.
21 2.5 eV **22** 8.0×10^{-19} J **23** photoelectrons
24 1.2×10^{15} Hz **25** 2.9 eV **26** 1.95 eV

27 Rb = 2.1 eV, Sr = 2.5 eV, Mg = 3.4 eV, W = 4.5 eV

28 a



b 4.1×10^{-15} eV s c 5.0×10^{14} Hz

d No. The frequency of red light (4.41×10^{14} Hz) is below the threshold frequency for rubidium, so no photoelectrons will be emitted.

29 a 4.78 keV

b The electrons have a de Broglie wavelength that is similar to the wavelength of the X-rays. This is evidence for the dual nature of light and matter.

c 2.6×10^{-24} kg m s^{-1}

30 a The detector observed a sequence of maximum and minimum intensities.

b As the electron beam is diffracted, the electrons are exhibiting wave-like behaviour. Electrons are not light but, like light, a beam of electrons can be diffracted.

31 Energy levels in an atom cannot assume a continuous range of values but are restricted to certain discrete values, i.e. the levels are quantised.

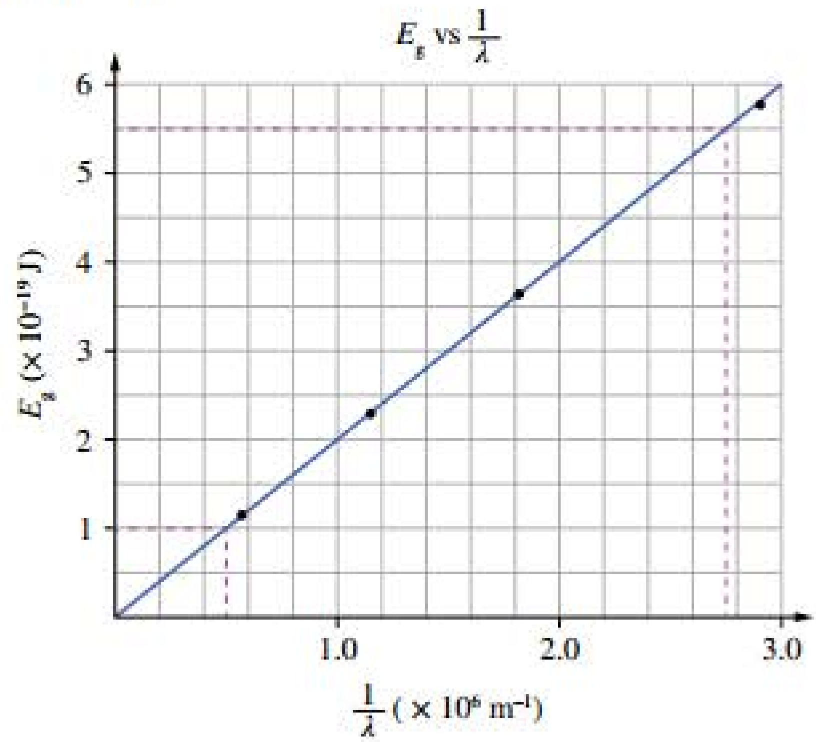
32 2.92×10^{15} Hz

33 Bohr's work on the hydrogen atom convinced many scientists that a particle model was needed to explain the way light behaves in certain situations. It built significantly on the work of Planck and Einstein.

34 The emission spectrum of hydrogen appears as a series of coloured lines. The absorption spectrum of hydrogen appears as a full visible spectrum with a number of dark lines. The colours missing from the absorption spectrum match the colours that are visible in the emission spectrum.

35 As the filament heats up, the free electrons in the tungsten atoms collide, accelerate and emit photons. A wide range of photon wavelengths are emitted due to a wide range of different collisions (some weak, some strong).

36 6.7×10^{-34}



37 B 38 1.7×10^{-35} m

39 No—the wavelength of the bullet is many times smaller than the radius of an atom. Significant diffraction only occurs when wavelength and gap (or object) size are approximately equal, i.e. when $\lambda \geq w$.

40 For the product of the uncertainty in position and the uncertainty in momentum to remain constant then as the uncertainty in position is decreased, the uncertainty in momentum would increase.

41 It is likely that the photon would knock the electron off course and hence the electron's position would be subject to greater uncertainty.

Chapter 8 Special relativity

8.1 Einstein's theory of special relativity

8.1 Review

1 A **2** C **3** D **4** A, D
5 $-20.45\ m s^{-1}$ north or $20.45\ m s^{-1}$ south
6 A hanging pendulum in the spaceship will move from its normal vertical position when the spaceship accelerates.
7 The speed of the ball is greater for Jana than it is for Tom.
The speed of the sound is greater forwards than it is backwards for Jana, while for Tom it is the same forwards and backwards.
The speed of light is the same for Jana and Tom.
8 **a** $370\ m s^{-1}$ **b** $300\ m s^{-1}$ **c** $360\ m s^{-1}$ **d** $340\ m s^{-1}$
9 A
10 **a** $15\ m s^{-1}$ backwards **b** 3 m backwards **c** 0.2 s
11 **a** 0.1 s **b** $50\ m s^{-1}$ in all frames **c** 1 m **d** $50\ m s^{-1}$
e 0.08 s (approx.)
12 Atomic clocks enable extremely short events to be timed to many decimal places. Differences in time for the same event to occur, when measured by observers in different inertial frames of reference, indicate that time is not uniform between the two inertial frames. These measurements support Einstein's special theory of relativity.
13 very short, µs, very similar, should not, do

8.2 Time dilation

WE 8.2.1 520 s

8.2 Review

1 light, oscillation, time, constant
2 'Proper time' is the time measured at rest with respect to the event. Proper times are always shorter than any other times.
3 1.29 s **4** 48.15 s **5** 2.20 s **6** 1.15 s
7 **a** 1 m **b** 3.33×10^{-9} s **c** $d = vt = ct_c$ **d** 7.6×10^{-9} s.
e 2.3
8 **a** 1.74×10^{-5} s or 17.4 µs
b Non-relativistic: 655 m
Relativistic: 5178 m
9 $t_0 = 2.93 \times 10^{-11}$ s
10 The equator clock is moving faster relative to the poles. It is also accelerating and hence will run slower. The effect is well below what we can detect, as the speed of the equator is 'only' about $460\ m s^{-1}$, which is about 1.5 millionths of *c*.

8.3 Length contraction

WE 8.3.1 3.90 m **WE 8.3.2** 20.5 m

8.3 Review

1 The length that a stationary observer measures in their own frame of reference. That is, the object (or distance) that is being measured is at rest with the observer.
2 A **3** 0.812 m
4 3.37 m **5** **a** 0.9c or $2.71 \times 10^8\ m s^{-1}$
b 0.643 m
The fast-moving garage appears even shorter than its proper length to the car driver.
6 Proper time, t_0, because the observer can hold a stopwatch in one location and start it when the front of the carriage is in line with the watch and stop it when the back of the carriage is in line with it.
7 C **8** **a** 0.866*c* or $2.60 \times 10^8\ m s^{-1}$ **b** 0.5
9 23.5 m (At this speed, there is no difference in length.)
10 **a** 1.20 m **b** the proper length, 2.75 m

8.4 Relativistic momentum and energy

WE 8.4.1 **a** $1.56 \times 10^{-21}\ kg m s^{-1}$ **b** 0.998*c*
WE 8.4.2 1.00*c*

8.4 Review

1 $9.53 \times 10^5\ kg m s^{-1}$ **2** $9.65 \times 10^{-18}\ kg m s^{-1}$
3 $1.59 \times 10^{-23}\ kg m s^{-1}$ **4** 5.67×10^{14} J
5 3.11×10^{14} J **6** B **7** −0.435*c*

Chapter 8 review

1 No object can travel at or beyond the speed of light, so the value of $\frac{v^2}{c^2}$ will always be less than 1.
The number under the square root sign will also, therefore, be a positive number less than 1.
The square root of a positive number less than 1 will always be less than 1 as well.
2 1.000 000 014
3 A (postulate 2) and C (postulate 1)
4 At the poles. The Earth has a very small circular acceleration, which is negligible for most purposes; however, at the poles it is even less.
5 C
6 Space and time are interdependent—motion in space reduces motion in time.
7 $3 \times 10^8\ m s^{-1}$ **8** A, B **9** B
10 You could not tell the difference between (i) and (iii), but in (ii) you could see whether an object such as a pendulum hangs straight down.
11 In your frame of reference time proceeds normally. Your heart rate would appear normal. As Mars is moving at a high speed relative to you, people on Mars appear to be in slow motion as time for them, as seen by you, will be dilated.
12 26.8 s **13** **a** 0.992 s **b** 0.992 s **14** C
15 **a** 0.866*c* or $2.598 \times 10^8\ m s^{-1}$
b No, it can't be greater than *c*!
$v = 0.968c$ or $2.90 \times 10^8\ m s^{-1}$
16 **a** 1.67 s **b** Length: 1.80 m
The height is unchanged at 1.0 m.
17 **a** 5.6 years **b** 2.45 years
c Raqu sees the distance as only 2.183 ly.
18 **a** 1.4 mm
b No, the motion is perpendicular to the north–south direction, so this dimension is not affected.
19 **a** $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{(0.995c)^2}{c^2}}} = \frac{1}{\sqrt{1-(0.995)^2}} = 10.01$
b No, they don't experience any difference in their own time frame.
c 25.1 years from our frame of reference
d 2.51 years
e No! They see the distance between Earth and Vega foreshortened because of the high relative speed, so to them the distance is only about 2.5 ly.
20 Earth observer: the observer will not measure the proper time of the muon's life span. Instead they will see that the muon's time is slow according to the equation $t = t_0\gamma$, where t_0 is the rest life span of the muon. The result is that the observer sees the muon live a much longer time, *t*, and therefore make it to the Earth's surface.
Muon: the muon will see the Earth approach at a very high speed (approx. 0.992c) and will see the distance contracted. It will not be 15 km, but instead be much shorter according to the equation $L = \frac{L_0}{\gamma}$. The distance the muon travels is *L*.
21 0.533*c*
22 **a** $1.85 \times 10^{12}\ kg m s^{-1}$ **b** $2.44 \times 10^{12}\ kg m s^{-1}$
c As the velocity of the starship approaches *c* the Lorentz factor becomes significantly larger.
23 **a** 0.949*c*
b Different velocities means they must be in different reference frames regardless of the direction in which each is moving.
24 **a** $9.20 \times 10^8\ kg m s^{-1}$ **b** 3.25×10^{17} J

Chapter 9 The Standard Model

9.1 Particles of the Standard Model

9.1 Review

1
- gauge bosons
- leptons
- hadrons

2 The Standard Model of particle physics describes three of the fundamental forces in the universe but does not incorporate the gravitational force. It predicts that all matter is made of quarks and leptons. The main difference between these two groups of particles is that quarks experience the strong nuclear force but leptons do not. Forces such as the strong nuclear force are mediated by force-carrier particles called gauge bosons that are different for each force. Gluons are the force-carrying particle for the strong nuclear force, photons carry the electromagnetic force and Z, W^- and W^+ particles carry the weak nuclear force.

3 A

4 Particle decay of the neutron: the weak nuclear force
Proton's attraction to the electron: the electromagnetic force
Atom drifts towards the base of the container: the force of gravity
The one force that is not involved is the strong nuclear force.

5 B

6 **a** gluon: gauge boson **b** neutrino: lepton
c neutron: hadron **d** photon: gauge boson
e electron: lepton **f** muon: lepton
g proton: hadron **h** tau: lepton

7 B

8

	Fermions		
Type of fermion	**quark**		**lepton**
Gauge boson	**gluon**		**photon**
Charge	$+\frac{2}{3}$, $-\frac{1}{3}$		**0, −1**
Hadrons formed	**baryon**	**meson**	
Number of particles in the hadron	**3**	**2**	

9 D

10 **a** $\overline{e}^+$ **b** $\overline{\nu}_\tau$ **c** ν_τ **d** $^-$ **e** $\rightarrow$

9.2 Interactions between particles

WE 9.2.1 4.0×10^{-13} J **WE 9.2.2** 0.178 MeV

WE 9.2.3 The total charge before and after reaction is the same, so reaction is allowed according to the law of conservation of charge.

WE 9.2.4 The total baryon number before and after reaction is the same, so reaction is allowed according to the law of conservation of baryon number.

WE 9.2.5 The total electron lepton number before the reaction is different from the total electron lepton number after the reaction. Therefore, this proposed reaction is forbidden according to the law of conservation of electron lepton number.

9.2 Review

1 This interaction results in a force of repulsion between two positrons and is mediated by the exchange of a photon, therefore it involves the electromagnetic force.

2 This interaction is the decay of a muon, which involves the weak nuclear force.

3 0.70×10^3 eV

4 Charge is not conserved, therefore the decay is forbidden according to the law of conservation of charge.
Energy is not conserved, therefore the decay is forbidden according to the law of conservation of energy.

5 Annihilation involves the destruction of a particle and its antiparticle, and the conversion of mass into energy.
Pair production involves the production of a particle and its antiparticle and the conversion of energy into mass.

6 Conservation laws require a quantity or quantum number to be the same before and after an interaction or decay. These laws can be used to predict if a reaction is allowed or forbidden.

7 C

8 Law of conservation of:
a lepton number **b** baryon number **c** charge

9 antielectron neutrino ($\overline{\nu}_e$); law of conservation of electron lepton number

10 antielectron neutrino ($\overline{\nu}_e$)

9.3 Particle accelerators

WE 9.3.1 3.0×10^3 MeV **WE 9.3.2** 1 m

9.3 Review

1 B

2 **a** Bending magnets apply a centripetal force to particles.
b Beam pipe allows an evacuated path for particles to be produced.
c Accelerating cavities increase the velocity of the particles.
d Detectors identify and quantify collision products.

3 1.88×10^{-28} kg

4 10 m (1 sig. fig.)

5 B

6 It was the development of higher-energy accelerators in which particles were able to be collided with sufficient energy to produce heavy bosons.

7 D **8** negative particle, 6.7×10^{-26} kg

9 0.0076 times **10** $2r$

9.4 Expansion of the universe

9.4 Review

1 A

2 Order: galaxy 2, galaxy 1, galaxy 3

3 The Hubble constant represents the rate of expansion of the universe today. For each Mpc a galaxy is away from us, it is receding from us 67.80 km s^{-1} faster.

4 6×10^2 Mpc

5 In a conventional explosion, all matter moves outwards from a single point. However, space expands uniformly in whatever direction you look and from wherever you look. There is no 'centre' of the three-dimensional space of the universe.

6 A, C, D

7 It was emitted when neutral atoms formed in the early universe. The early radiation continues to exist in the universe but has expanded with the universe.

8 Shortly after the Big Bang, the universe was full of very high temperature radiation. This radiation still fills the universe, but the expansion has cooled it to less than 3 K. The Steady State theory has no explanation for the presence of such radiation. Instead, the Steady State theory says the universe has been the same forever.

9 gravity, strong nuclear, weak nuclear, electromagnetic

10 The Steady State theory says that hydrogen atoms are created at a rate of one atom per cubic metre per 300 000 years, while Big Bang theory says that all of the hydrogen and helium was created at the Big Bang out of nothing.

Chapter 9 review

1 All quarks experience the strong nuclear force; leptons do not. Quarks also have non-integer (fractional) charges; leptons have charges of −1 or 0. Leptons are generally not found in the nucleus, while quarks are found in the nucleus. Leptons can exist as free particles but quarks cannot.

2 C **3** **a** photon **b** gluon **c** W^+, W^- and Z

4 The Standard Model is based on the assumption that forces arise through the exchange of particles called gauge bosons (or just bosons). Each of the three forces is mediated by a different particle. The strong nuclear force is mediated by the gluon, electromagnetism by the photon and the weak nuclear force by W^+, W^- and Z bosons.

5 The particles of the Standard Model include gauge bosons, leptons and hadrons. Both gauge bosons and leptons are fundamental particles, which means that they are not made up of other particles, while hadrons are made of two or three smaller particles called quarks. Gauge bosons are force-carrying particles, while leptons and hadrons are not.

6 gravity 7 1.6×10^{-9} J

8 law of conservation of baryon number and the law of conservation of muon lepton number

9 Colliding two particles travelling in opposite directions is possible in a synchrotron, but not in a linear accelerator. In a synchrotron, the sum of the momentums before the collision will be close to zero and so the sum of the momentums carried by particles produced in the collisions is relatively small. This results in most of the energy of the collision being available to create the mass of new particles.

10 0.6T 11 400%

12 rest mass and relativistic kinetic energy

13 Within a linear accelerator, an electric field does work on charged particles by alternating the potential of the drift tubes along the accelerator at a constant rate. This causes particles to gain kinetic energy.

14 It had been found that the Cepheids varied in brightness and that the period of this variation was related to the intrinsic brightness. By comparing the intrinsic and apparent brightness the distance could be found.

15 Baryon number and charge are both conserved, therefore reaction is allowed by both these laws.

16 Annihilation describes the process by which a particle and its corresponding antiparticle collide, and are converted into photons.

17 1.52 MeV 18 1.4×10^{10} years

19 Edwin Hubble's graph of recession velocity vs distance to the galaxy showed a straight-line relationship, indicating that recession velocity is proportional to the distance of a galaxy.
If distant galaxies are all observed to be moving away from us, you might first assume that the galaxies are all moving apart and you are at the centre of the universe. If we were to run time backwards, this would result in all the galaxies having come from the same region of space and therefore the universe must be expanding.
Note: Under the assumption that there is nothing remarkable about our location in the universe, we must conclude that the galaxies are not all flying away from us and that it is in fact the space between the galaxies that is getting bigger.

20 A 21 B

Chapter 10 Practical investigation

10.1 Designing and planning the investigation

10.1 Review

1 **a** If the voltage is measured in units of number of batteries then it is a discrete value.
b If the voltage is measured with a voltmeter then the voltage would be continuous.

2 qualitative

3 A

4 **a** valid **b** reliable **c** accurate

5 **a** the tension in the elastic band
b the initial launch velocity of the elastic band
c the same elastic band, elastic band held in the same way, elastic band launched in the same direction, elastic band placed on the finger in the same way

10.2 Conducting investigations and recording and presenting data

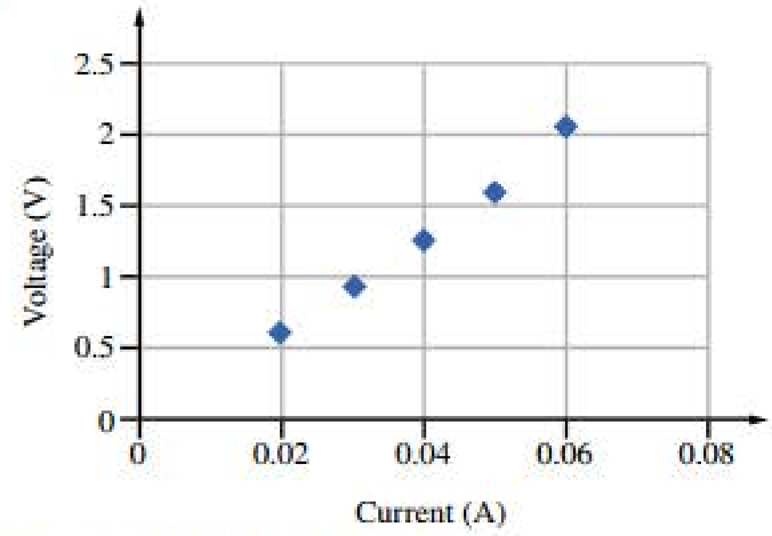
10.2 Review

1 **a** systematic error **b** random error

2 two significant figures

3 **a** mean: 22.7 **b** mode: 19 **c** median: 22.5
d uncertainty in the mean: ±5

4



5 as a line of best fit on the graph

10.3 Discussing investigations and drawing evidence-based conclusions

10.3 Review

1 a proportional relationship between two variables

2 an inversely proportional relationship

3 a directly proportional relationship

4 time restraints and limited resources

5 An increase in current from 0.03 A to 0.05 A produced an increase of 0.88 V across the resistor.

Chapter 10 review

1 A hypothesis is a prediction, based on evidence and prior knowledge, to answer the research question. A hypothesis often takes the form of a proposed relationship between two or more variables.

2 Dependent variable: flight displacement
Independent variable: release angle
Controlled variable: (any of) release velocity, release height, landing height, air resistance (including wind)

3 **a** the acceleration of the object
b the vertical acceleration of the falling object
c the rate of rotation of the springboard diver

4 elimination, substitution, isolation, engineering controls, administrative controls, personal protective equipment

5 $6.8 \pm 0.4\,\text{cm s}^{-1}$

6 the mean

7 an exponential relationship

8 This graph should show a straight line with a positive gradient.

9 Any issues that could have affected the validity, accuracy, precision or reliability of the data plus any sources of error or uncertainty.

10 Bias is a form of systematic error resulting from a researcher's personal preferences or motivations.

Unit 4 review

Section 1: Short response

1 An electromagnetic wave is a propagating wave consisting of oscillations of electric and magnetic fields. The relationship between energy, wavelength and frequency can be summarised in the wave equation $E = hf = \frac{hc}{\lambda}$: the higher the energy, the shorter the wavelength and the higher the frequency.
Electromagnetic waves differ from mechanical waves in that they do not require a medium in which to propagate.

2 a The photoelectric effect demonstrates light exhibiting particle-like behaviour. When a beam of light of high enough energy is incident on a metal, electrons in the metal will be ejected. These electrons can then be collected at the collector electrode and a photocurrent detected. If a negative potential is applied up until the point at which current is reduced to zero, the exact kinetic energy of the electrons can be determined. Several observations were made from this:

- When light intensity increases, the photocurrent increases.
- There is a negative voltage for which no photoelectrons reach the collector. This is known as the stopping voltage. For a particular metal, each frequency of light will give a characteristic stopping voltage. This value is independent of light intensity.

The characteristics of the photoelectric effect could not be explained using a wave model of light. According to the wave model, the frequency of the light should be irrelevant to whether or not photoelectrons are ejected. Since a wave is a form of continuous energy transfer, it would be expected that even low-frequency light should transfer enough energy to emit photoelectrons if left incident on the metal for long enough. Similarly, the wave model predicts that there should be a time delay between the light striking the metal and photoelectrons being emitted, as the energy from the wave builds up in the metal over time.

b When given enough momentum and fired through small slits (or more complex objects, such as a crystalline structure), electrons diffract and produce interference patterns with one another. Furthermore, it can be shown that electrons with the same wavelength as a particular frequency of light will produce the same diffraction pattern. This wavelength is predicted by de Broglie's formula for matter waves: $\lambda = \frac{h}{p}$.

3 $c = 2.99 \times 10^8\,\text{m s}^{-1}$

4
- Only certain frequencies of light will eject photoelectrons.
- There is no time difference between the ejection of photoelectrons by light of different intensities.
- The maximum kinetic energy of the ejected photoelectrons is the same for different light intensities of the same frequency.

5 a 1 The laws of physics are the same in all inertial frames of reference.
2 The speed of light is a constant for all observers.

b 0.625c

6 Moving objects contract, and so the 1.50 m ruler should be shorter in the moving frame.
$L' = 1.34\,\text{m}$

7 Weak nuclear for the decay, electromagnetic attracts the proton and the electron together and gravity accelerates the atom to the base of the container; the strong nuclear is not involved.

8 a The relativistic Doppler effect tells us that light from a source moving away from us is being redshifted, i.e. its wavelength is moved towards the red end of the electromagnetic spectrum. Therefore, the observed wavelength would have been *longer*.

b $105\,\text{km s}^{-1}\,\text{Mpc}^{-1}$

Section 2: Problem solving

9 a Atoms have electrons that are each bound in quantised energy orbitals. When light is incident on the atom, an electron absorbs a photon, which excites the electron to a higher energy state. When the electron returns to any lower energy state, it emits a photon with energy exactly equal to the difference between the energy of the orbitals. Therefore, when observing the light in the emission spectrum of these atoms, the only wavelengths present will be those with energy corresponding to the energy difference of the different atomic orbitals.

b $E = 1.89\,\text{eV}$
This corresponds to the energy transition from $n = 3$ to $n = 2$.

c $1.11 \times 10^3\,\text{m s}^{-1}$

10 a According to the particle theory, light should have passed directly through the slits to produce two bright lines or bands on the screen. Instead, Young observed a series of bright and dark bands or 'fringes'. Young was able to explain this bright and dark pattern by treating light as a wave. He assumed that the monochromatic light was like plane waves and that, as they passed through the narrow slits, these plane waves were diffracted into coherent (in phase) circular waves. The circular waves would interact, causing interference. The interference pattern produced by these two waves would result in lines of constructive (antinodal) and destructive (nodal) interference that would match the bright and dark fringes respectively.

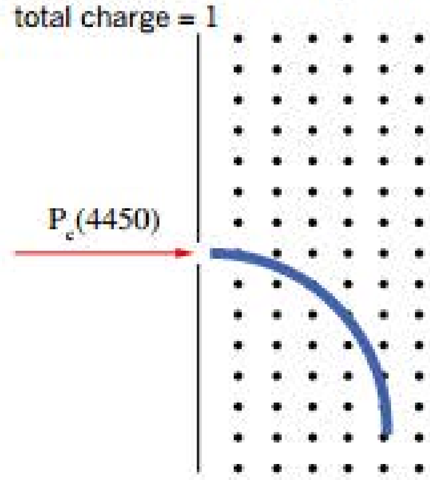
b 500 nm.

c The interference pattern will be no longer uniform. The central band will have the greatest intensity, and each subsequent antinode will decrease in brightness.

d Fringe spacing, Δx, is inversely proportional to the distance between slits, so the fringes will be further apart.

11 a total charge = 1

b



c 39.8 m

d the strong nuclear force

12 a $1.578 \times 10^8\,\text{m s}^{-1}$

b 357 MeV

Section 3: Comprehension

13 a i The amount that light refracts is wavelength dependent. Longer wavelengths, such as those of red light, travel the fastest in the new material so they are refracted the least. Shorter wavelengths, such as those in violet light, are slower so they are refracted the most. This means that when light passes through a lens, the angle by which the wavelengths change upon refracting is different, causing the different colours to spread out.

ii red, green, violet

b $\theta_c = 38.1°$

c For light incident upon a material to create a current, the incident photon must eject an electron (the photoelectric effect). Photons will be refracted by the camera lens onto the sensor, at which point an electron is ejected from the surface of the sensor. This photocurrent can then be detected as a signal upon that particular pixel.

d The photoelectric effect has a minimum threshold at which it will no longer eject electrons. Incident light below this threshold frequency will not be energetic enough to be able to stimulate the electrons into being ejected. Radio waves will be well below this threshold frequency, and therefore unable to be detected by the camera's sensor.

e $\phi = 1.80\,\text{eV}$

Glossary

A

absorption spectrum Spectrum containing dark lines in the positions of the wavelengths that are absorbed by a gas as light passes through it. This is related to the emission spectrum of the gas.

acceleration due to gravity Rate at which a falling object will accelerate in a gravitational field. Equivalent to the gravitational field strength. Measured in $m\,s^{-2}$.

aether An invisible, massless, rigid substance that was proposed as the medium in which light waves propagate. There is no experimental evidence for the existence of the aether.

air resistance A retarding force that acts in the opposite direction to the motion in air of an object or projectile.

alternator An electric generator that produces alternating current (AC).

altitude Height above a planet's surface.

amplitude The absolute value of the maximum displacement from a zero value during one period of an oscillation.

annihilation The process in which matter is completely converted into energy. This is not a chemical process in which matter in one form is converted to matter in another form, as in burning.

antiparticle Particles that have the same mass as their ordinary particle but opposite properties such as charge, spin, baryon number and lepton number, e.g. electron (particle) and positron (antiparticle).

apparent weight The weight felt by a person when their body is stationary or in motion. Sometimes it is higher or lower than their usual weight. Equivalent to the size of the normal reaction force acting on the person.

apparent weightlessness When an object is in free fall and there is no force between it and its surroundings. It appears to be weightless although it is still under the influence of gravity. Occurs when there is no normal reaction force acting between an object and a surface.

armature A revolving structure in an electric motor or generator, wound with coils that carry current. It rotates within a magnetic field to induce an emf.

artificial satellite Body such as Sputnik, the Hubble Space Telescope, or NOAA-19, made by humans and placed in orbit around a planet or the Moon.

axis of rotation An imaginary line through the centre of mass or pivot point of an object, which is perpendicular to the plane of rotation of the object.

B

back emf The opposing emf induced when an electric motor is running, which reduces the current in the circuit and prevents the motor burning out.

banked track A track inclined at some angle to the horizontal, enabling vehicles to travel at higher speeds when cornering compared with around a horizontal curved path.

baryon A composite particle composed of three quarks. Baryons belong to the group of particles called hadrons. The most common examples are the proton and the neutron.

baryon number A quantum number conserved in particle interactions. This means the sum of the baryon numbers before an interaction is equal to the sum of the baryon numbers after the interaction. Baryons (particles containing three quarks) are assigned +1, antibaryons are assigned −1 and all other particles are assigned 0.

base The support for a structure. For example, the base of a car is a rectangle with each of the four tyres at the corners of the rectangle.

Big Bang theory The leading model for how the universe began. It describes the universe starting in a high-density state and then expanding.

blackbody A body that absorbs all incident radiation; it radiates and absorbs energy at the same rate, so the temperature remains constant.

blackbody radiation The radiation emitted by an object due to its temperature. This results in a distribution of different wavelengths or frequencies that is characteristic of the temperature of the object.

brushes Devices that transfer the current in the rotating coil to a stationary external circuit by pressing against the split ring commutator or the slip rings.

C

cantilever A beam that extends out horizontally beyond its supporting structure.

cathode ray tube A vacuum tube in which a hot cathode emits a beam of electrons that pass through a high voltage anode and are focused or deflected before hitting a fluorescent screen.

centre of gravity The position from which the entire weight of the body or system is considered to act; it is the position at which the body will balance.

centre of mass A single point in an object where the mass can be considered to be 'concentrated' for the purposes of analysing motion.

centripetal acceleration Acceleration directed towards the centre of a circle when an object moves with constant speed in a circular path.

centripetal force The force that causes an object to travel in a circular path; can include gravity, tension, normal force and friction.

classical physics The physics of Galileo and Newton, in which the addition of velocities has no limit, and length and time are constant.

coherent Waves that are in phase, i.e. at the same stage at the same time.

commutator A device, usually made from a split ring of copper or another good conductor, that connects the armature coil to the DC electrical power supply. Each half is connected to one end of the coil of wire and so the current is reversed every half a turn to keep the coil rotating in one direction.

components The components of a force are two vectors at right angles to each other that when added together will be equivalent to the original force.

conservation law Describes a condition that something must remain unchanged. An example in particle physics is the conservation of charge. The sum of the charge on the particles before an interaction must be equal to the sum of the charge on the particles after the interaction.

conserved Not created or destroyed, but remaining constant.

constructive interference The process in which two or more waves of the same frequency combine to reinforce each other. The amplitude of the resulting wave is equal to the sum of the amplitudes of the superimposed waves.

cosmic microwave background radiation (CMBR) Electromagnetic radiation that comes from every direction in the universe and is considered to be a remnant of the Big Bang.

critical angle The angle of incidence that produces an angle of refraction of 90°. The largest angle for which refraction will occur; at angles larger than the critical angle, light undergoes total internal reflection.

D

de Broglie wavelength Wavelength associated with a particle due to its mass and velocity.

design speed The speed at which a vehicle experiences no sideways force as it travels around a banked track. It is dependent on angle.

destructive interference The process in which two or more waves of the same frequency combine to cancel each other out. The amplitude of the resulting wave is equal to the difference between the amplitudes of the superimposed waves.

diffraction The bending of waves around obstacles or through gaps in barriers in their path.

diffraction pattern The pattern of dark and light bands that is seen when light passes through a single small gap. Areas of constructive interference appear as bright bands and areas of destructive interference appear as dark bands.

dipole Two electric charges or magnetic poles that have equal magnitudes but opposite signs, usually separated by a small distance.

dispersion The process of splitting light into its component colours to create a spectrum.

E

eddy current A circular electric current induced within a conductor by a changing magnetic field.

electric field A region of space where charged objects experience a force due to the field created by another charged object.

electric field strength A measure of the force per unit charge on a charged object within an electric field, with the units $N\,C^{-1}$. Field strength can also be a measure of the difference in electrical potential per unit distance, with the units $V\,m^{-1}$.

electrical potential The work required per unit charge to move a charged object from infinity to a point in the electric field, with the units $J\,C^{-1}$.

electromagnet A magnet consisting of an iron or steel core wound with a coil of wire, through which a current is passed. The core only becomes magnetised when current is flowing.

electromagnetic force A fundamental force in nature associated with the photon. It is responsible for both electric and magnetic fields exerting forces of attraction or repulsion.

electromagnetic induction The creation of an electric current, or an emf, in a loop of wire as the result of changing the magnetic flux through the loop.

electromagnetic radiation Energy emitted in continuous waves with two transverse, mutually perpendicular components: a varying magnetic field and a varying electric field.

electromagnetic spectrum The range of all possible frequencies of electromagnetic radiation. The visible spectrum is just one small part of the electromagnetic spectrum.

electromotive force (or emf) It is not an actual force but is the energy per coulomb provided by any source of electrical energy. It is measured in volts (V) and is equal to the potential difference across the terminals of the energy source.

electron-volt Amount of work done to move an electron across a potential difference of 1 volt: $1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$. It is an alternative to the joule as a unit in which to measure energy.

electroweak force A force postulated to explain both the electromagnetic force and the weak nuclear force as two aspects of a single force.

electroweak theory The mathematical theory that describes both the electromagnetic and weak forces of nature. This theory is a component of the Standard Model of particle physics.

emf See *electromotive force*

emission spectrum Spectrum of coloured lines in the positions of the wavelengths of light emitted when a gas is heated or has an electric current passed through it. This is related to the absorption spectrum of the gas.

excited state Energy state of an atom above the ground state ($n > 1$).

F

Faraday's law Law stating that the average emf generated in a coil is proportional to the rate of change of magnetic flux and the number of turns in the coil.

fermion A fundamental particle not comprised of other smaller particles. All fermions are either quarks or leptons.

Feynman diagram A graphical representation of the interaction between subatomic particles.

field A region of space around an object with a particular property such as charge, mass, or magnetism, where a force can be felt by other objects with that particular property.

field lines A two-dimensional graphic representation of a field, using arrows to indicate the direction of the field. The closer the field lines, the stronger the field.

force arm The perpendicular distance between the axis of rotation and the line of action of the force.

forward voltage In the photoelectric effect experiment, forward voltage is when the cathode is negative and the anode is positive. It acts to accelerate the emitted photoelectrons.

frame of reference A coordinate system that is usually fixed to a physical system that contains an object and/or an observer. There can be a frame of reference within another frame of reference.

free-fall A motion whereby gravity is the only force acting on a body.

frequency The number of waves passing a given point in one second or number of repeats of a cycle each second. Measured in hertz (Hz).

friction The force that opposes the motion of an object over or through another medium.

G

gauge bosons Exchange particles that mediate the four fundamental forces.

Gedanken German word for 'thought'. Einstein used this term to describe his theoretical 'experiments' on relativity.

generator An electrical device that converts kinetic energy into direct current (DC) electricity. Usually, a coil is rotated, causing it to cut across a magnetic field.

geostationary satellite A satellite that remains in orbit above the same point on the equator on the Earth's surface. It has the same period as the Earth's rotation, i.e. 24 hours. Only occurs at an altitude of 36 000 km above the Earth.

gluons Elementary particles that act as exchange particles for the strong nuclear force between quarks; similar to the exchange of photons in the electromagnetic force between two charged particles. Gluons themselves carry the colour charge of the strong interaction. Gluons can be considered to be the fundamental exchange particle underlying the strong interaction between protons and neutrons in a nucleus.

gravimeter Sensitive instrument used by geologists to detect small variations in gravitational field strength.

gravitational constant, *G* Universal constant of value $6.67 \times 10^{-11}\,\text{N}\,\text{m}^2\,\text{kg}^{-2}$.

gravitational field The region around an object where other objects will experience a gravitational force.

gravitational field strength The strength of gravity, usually measured at the surface of a planet. Equivalent to the acceleration due to gravity, g. Measured in newtons per kilogram ($\text{N}\,\text{kg}^{-1}$).

gravitational force The force of attraction acting between two objects that have mass.

gravitational potential energy The energy that a body possesses due to its position in a gravitational field. A scalar quantity that is measured in joules (J).

ground state Lowest energy state of an atom ($n = 1$).

H

hadron A composite particle that contains quarks held together by the strong nuclear force. Hadrons are subdivided into two families: baryons (e.g. the proton and neutron) and mesons (the pion and kaon).

Heisenberg's uncertainty principle Concept that any measurement of a system creates a disturbance of the system with a resulting uncertainty in the measurement.

Hubble constant The unit of measurement used in Hubble's law, which describes the expansion of the universe. It has a value of around $70\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$.

Hubble's law A law proposed by Edwin Hubble that states that the rate at which astronomical objects in the universe move apart from each other is proportional to their distance from each other.

I

ideal transformer Where the input power and the output power are equal and the transformer is 100% efficient. Real transformers obtain close to this value.

incandescent Emission of light due to very high temperature.

inclined plane Sloping surface or ramp.

induced current Electric current produced by changing a magnetic flux in the region of a conductor or by moving the conductor in a magnetic field.

induction hotplate A stove that heats by inducing eddy currents in metal (iron/steel) pans placed above coils that have a changing magnetic fields.

inertial frame of reference A frame of reference that is either moving with a constant velocity or is stationary. It is not accelerating.

interference The process in which two or more waves of the same frequency combine to reinforce or cancel each other out. The amplitude of the resulting wave is equal to the sum of the amplitudes of the superimposed waves. See also *constructive interference* and *destructive interference*.

intrinsic brightness The actual brightness of a star, regardless of the distance from the observer.

inverse square law Relationship between two variables in which one is proportional to the reciprocal of the square of the other.

K

kinetic energy The energy of a moving body, measured in joules (J). Kinetic energy is a scalar quantity.

L

laser Source of a narrow beam of intense, monochromatic, polarised, coherent radiation.

law of conservation of energy States that energy cannot be created or destroyed, but can only be changed or transformed from one form to another.

law of conservation of momentum In any collision or interaction between two or more objects in an isolated system, the total momentum of the system will remain constant; i.e. the total initial momentum is equal to the total final momentum.

length contraction Length in a moving frame of reference appears shorter when viewed by a stationary observer.

Lenz's law A law stating that the direction of the induced current in a conductor is such that its associated magnetic field opposes the change in flux that caused it.

lepton A fundamental particle that does not experience the strong nuclear force, e.g. electron, muon, neutrino.

lepton number A quantum number conserved in particle interactions. This means the sum of lepton numbers before an interaction is equal to the sum of lepton numbers after the interaction. Leptons are assigned +1, antileptons are assigned −1 and all other particles are assigned 0.

light-emitting diode (LED) Semiconductor diode that uses the excitation of electrons to emit light.

line of action The line along which a force is acting. The line of action extends forwards and backwards from the force vector.

Lorentz factor The factor by which time, length and mass change for an object moving at relativistic speeds (i.e. speeds close to the speed of light): $\gamma = \dfrac{1}{\sqrt{1-\dfrac{v^2}{c^2}}}$

Lorentz force The force experienced by a point charge moving along a wire that is in a magnetic field; the force is at right angles to both the current and the magnetic field. Named for the Dutch physicist who shared a 1902 Nobel Prize for researching the influence of magnetism on radiation.

M

magnetic Of or relating to magnetism or magnets. Having the properties of a magnet. Capable of being magnetised or attracted by a magnet.

magnetic field A magnetic field is a region influenced by a magnet or something with the properties of a magnet.

magnetic flux The strength of a magnetic field in a given area expressed as the product of the area and the component of the field strength at right angles to the area (i.e. $\Phi = B_{\perp}A$).

magnetic flux density Amount of magnetic flux per unit area. In other words, it describes 'the closeness of magnetic field lines'. Same as magnetic field strength.

magnetic pole Magnetic poles are the two limited regions in a magnet at which the field of the magnet is most intense.

mechanical energy The energy that a body possesses due to its position or motion. Kinetic energy, gravitational energy and elastic potential energy are all forms of mechanical energy.

medium A physical substance, such as air or water, through which a mechanical wave is propagated.

metal vapour lamp Lamp that contains a low-pressure gas that becomes excited and emits photons with the colour characteristic of the element in the gas, e.g. sodium vapour lamp.

mnemonic A mnemonic device is any learning technique that aids information retention. Mnemonics aim to translate information into a form that the brain can retain better than its original form. Even the process of learning this conversion might aid in the transfer of information to long-term memory.

momentum The product of the mass and the velocity of an object. Momentum is a vector. It is measured in $kg\,m\,s^{-1}$.

monochromatic Light of a single colour, e.g. red light.

N

natural satellite A body such as the Moon or a planet (not made by humans) that is in orbit around another body.

nebula An interstellar cloud in outer space composed of dust and gasses. The Latin word for cloud is *nebula*.

neutral equilibrium A situation in which an object will remain stationary no matter where it is, e.g. a ball on a horizontal table.

Newton's law of universal gravitation Law that states that the attractive gravitational force between two masses is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

normal reaction force Force with which a surface pushes back on an object, at right angles to the surface. Same as the apparent weight of an object. Symbol F_N or N, measured in newtons.

P

pair production The creation of a particle and its antiparticle. This is commonly the result of two photons interacting or a photon interacting with an atomic nucleus.

parallax movement The apparent movement of the closer stars relative to the background stars, but which is actually due to the motion of Earth around the Sun.

parsec The distance to a star that has a parallax angle of one arcsecond; equal to 206 265 AU.

particle accelerator A machine that can accelerate a charged particle (proton, electron) or an atomic nucleus to very high speeds, including speeds that approach the speed of light.

path difference The difference in the lengths of the paths from each slit to the screen in a double-slit experiment.

Pauli's exclusion principle States that no two particles can occupy the same quantum state at the same time. For example, no two electrons can be in the same shell or orbital around an atom and have the same energy.

period The time interval taken to complete one cycle of a regularly repeating phenomenon, such as a rotating object or in a sound wave. The SI unit for period is seconds (s).

photocurrent Current caused by the flow of photoelectrons during the photoelectric effect.

photoelectric effect Spontaneous emission of electrons from a metal surface when it is illuminated by light of particular frequencies and energies.

photoelectron An electron released from an atom due to the photoelectric effect.

photon Small energy packets of electromagnetic radiation which have no mass, no charge and travel through a vacuum at the speed of light.

pivot point A point about which an object can rotate.

point charge An idealised situation in which all of the charge on an object is considered to be concentrated at a single point. The point size is negligible in relation to the distance between it and another point charge.

polarisation The phenomenon in which transverse waves are restricted in their direction of vibration.

pole The north pole of a freely suspended magnet is attracted to the Earth's magnetic North Pole (a magnetic south). The south pole is attracted to the Earth's magnetic South Pole (a magnetic north). See *magnetic pole*.

postulate A suggestion that is put forward as factual as a basis for further discussion or reasoning.

potential difference The work required per unit charge to move a charged object between two points in the electric field, with the units $J\,C^{-1}$ or volts (V).

principle of moments In order for an object to be in rotational equilibrium, the sum of the moments in a clockwise direction must balance the sum of the moments in an anticlockwise direction.

projectile Object moving freely through the air without an engine or power source driving it.

proper length A measurement of length made from the frame of reference in which the object being measured is stationary.

proper time A measurement of time made with a clock that doesn't move relative to the point at which the start and end of the event occurs.

Q

quantum (plural quanta) According to the quantum model, electromagnetic radiation is emitted from objects as discrete packets called quanta. Each quantum has an energy proportional to its frequency according to the equation $E = hf$.

quantum chromodynamics The theory that describes the interactions of the strong nuclear force, i.e. that between quarks.

quantum mechanics The physics of subatomic particles, including their behaviour and physical interactions.

quantum number The numbers that describe the properties of a particle, e.g. charge, spin.

quarks Six of the fundamental particles in the universe. Quarks cannot be found individually, and only exist bound by the strong nuclear force within hadrons.

R

redshift The change (lowering) of frequency that occurs in any wave phenomenon as the source of the waves moves away from the observer. Likewise, a blueshift occurs if the source is moving towards the observer.

reference point A point about which a rotational equilibrium can be calculated for an object that is in static equilibrium. This point can be anywhere, but is best selected to cancel the torque of an unknown force in a problem.

reflection Occurs when light, other electromagnetic radiation, sound, particles or waves, bounce back after reaching a boundary or surface.

refraction The bending of light, sound, or other type of wave, in passing at an angle to the surface from one medium into another in which its wave speed is different.

refractive index An index or number that is allocated to a medium indicating its refracting properties; ratio of the speed of light in a vacuum, c, to the speed of light in the medium, v, i.e. $n = \frac{c}{v}$.

regenerative braking A system that converts the kinetic energy of a car into electrical energy to recharge the car's battery.

relativistic mass (γm) Used by many to indicate the increase in mass that occurs as an object gains speed. This term was not endorsed by Einstein, who preferred using relativistic momentum. The relativistic mass equation is $m = \gamma m_0$, where γ is the Lorentz factor and m_0 is the rest mass.

reverse voltage In the photoelectric effect experiment, reverse voltage is when the cathode is positive and the anode is negative. It acts to slow the emitted photoelectrons.

right-hand grip rule Used to determine the direction of the magnetic field (curled fingers) around a current (thumb) in a current-carrying wire.

right-hand palm rule Used to indicate the direction of the force (palm) on a current (thumb) in a magnetic field (straight fingers).

root mean square The square root of the arithmetic mean of the squares of the numbers in a given set of numbers. In terms of alternating power, the root mean square value, $P_{rms} = \frac{P_{peak}}{\sqrt{2}}$. Similarly, for voltage and current. Alternatively, it is the effective mean (average) value of an AC supply.

rotational equilibrium A situation in which the sum of all the clockwise torques is equal to the sum of all the anticlockwise torques.

S

satellite Object in a stable orbit around a central body. Could be natural, like a planet, or artificial, like a communications satellite.

simultaneous When two events occur at exactly the same time.

sinusoidal Having a magnitude that varies as a sine curve.

slip rings Components of alternators (AC generators) that allow a constant electrical connection to be made between the rotating armature and the static external circuit through which the generated alternating current flows.

Snell's law Describes the relationship between incident light and refracted light in two media. The ratio of the refractive indices of the two materials is equal to the ratio of the sines of the angle of incidence and the angle of refraction.

solenoid A coil of wire that acts as an electromagnet when electric current is passed through it due to the magnetic field that is set up by the current passing through it. Solenoids are often used to control the motion of metal objects, such as the switch of a relay.

spacetime A term used to describe the situation in which the three-dimensional space coordinate system is linked to the one-dimensional time system.

special relativity Einstein's two postulates of special relativity can be abbreviated as follows: (1) No law of physics can identify a state of absolute rest. (2) The speed of light is the same to all observers.

split-ring commutator A component of DC generators and motors that typically resembles a ring that has been cut into two equal pieces or shells. Each part of the ring has a fixed connection to the ends of the coil, while also making contact with the stationary brushes. This means the connection between the rotating coil and the static circuit is reversed every half turn, which ensures that the direction of current in the circuit is constant (in the case of the generator) or the direction of rotation is constant (in the case of the motor).

stable equilibrium A situation in which an object will return to its equilibrium position, even when it is displaced by a force. For example, a ball placed in a large bowl will always return to the bottom of the bowl.

Standard Model The Standard Model of particle physics is a mathematical description of all known particles and three of the forces acting on them. It is currently the most successful theory for predicting the behaviour and properties of the particles that exist in nature.

standing wave Also called a stationary wave, it is the periodic disturbance in a medium resulting from the combination of two waves of equal frequency and intensity travelling in opposite directions.

static equilibrium A situation in which an object is in both translational equilibrium and rotational equilibrium.

stator A portion of a machine that remains stationary with respect to rotating parts, especially the collection of stationary parts in the magnetic circuits of a motor or generator.

Steady State theory A model of the universe based on an idea called the perfect cosmological principle. This states that the universe on the largest scales looks essentially the same everywhere at all times; therefore, the universe maintains the same average density of matter forever. The Steady State theory is not the accepted model of the universe.

step-down transformer Device that decreases the secondary voltage compared to the primary voltage.

step-up transformer Device that increases the secondary voltage compared to the primary voltage.

stopping voltage The applied voltage required to stop all photoelectrons from reaching the collector electrode. For a particular frequency of incident light on a particular metal, the stopping voltage is a constant.

strong nuclear force A short-range but powerful force of attraction that acts between all the nucleons in the nucleus. The strong nuclear force acts on quarks and binds them together in hadrons. It also acts at larger distances to bind protons and neutrons together within atomic nuclei. In the Standard Model of particle physics, the strong nuclear force is described by quantum chromodynamics and is mediated by an exchange of gluons.

synchrotron Large particle accelerator in a circular shape, which produces a very intense, very narrow beam of electromagnetic radiation called synchrotron light.

T

tangential Describes a direction forming a tangent to a curve.

threshold frequency The minimum frequency of electromagnetic radiation for which the photoelectric effect can occur for a given material.

time dilation When one observer watches events in a frame of reference that is moving (very fast) relative to him/her, time in that frame of reference will appear to go more slowly. People in the moving frame do not experience any difference in the rate at which time passes. This effect is one of the strange consequences of Einstein's theory of special relativity.

torque Any force or system of forces that causes or tends to cause rotation. A turning or twisting force effect.

torsion balance Device used to measure very small twisting forces. Cavendish used this device to measure the force of attraction between lead balls held a small distance apart.

total internal reflection Occurs when the angle of incidence exceeds the critical angle for refraction. Light or waves are then reflected back into the medium; there is no transmission of light.

transformer A device that transfers an alternating current from one circuit to one or more other circuits, usually with an increase (step-up transformer) or decrease (step-down transformer) in voltage. The input goes to a primary coil; the output is taken from a secondary coil or windings linked by induction to the primary coil.

translational equilibrium A situation in which the sum of the forces acting on an object is equal to zero.

U

uniform Constant, unvarying.

unstable equilibrium A situation in which an object will accelerate and will not return to its equilibrium position when it is displaced by a force, e.g. a sphere lying on top of a dome.

V

vector diagram A system of adding vectors in which each vector is drawn head-to-tail, with the resultant vector drawn from the tail of the first vector to the head of the last vector.

voltaic pile An early form of battery consisting of a pile of paired plates of dissimilar metals, such as zinc and copper, each pair being separated from the next by a pad moistened with an electrolyte (mild acid). Also called galvanic pile or Volta's pile.

W

wave–particle duality The theory that, in some experiments, light and matter behave like waves and, in other experiments, they behave like particles.

weak nuclear force The interaction between quarks that is responsible for changing one type of quark to another. This force is described by the electroweak theory and is responsible for radioactive decay and nuclear fission. In the Standard Model of particle physics this force is mediated by the W and Z bosons.

weight Force due to gravity that acts on any mass in a gravitational field. Symbol F_g or W, measured in newtons.

work function The energy required to remove an electron from its state of being bound to an atom; measured in joules or electron-volts.

Index

Note: Page numbers in **bold** refer to the main discussion of key terms (in bold) in the text.

Attributions

The following abbreviations are used in this list: t = top, b = bottom, l = left, r = right, c = centre.

COVER: Shutterstock: KristinaSh.
123RF: 1xpert, p. 24t; Mikhail Abramov, p. 108; claudiodivizia, p. 246c; donatas1205, p. 281b; Pawel Gaca, p. 106b; Patricia Hofmeester, p. 325t; jewhyte, p. 63t; lianem, p. 102b; Felix Lipov, p. 304t; manganganath, p. 310b; nmint, p. 117bl; Andrey Pashkov, p. 106b; Daniel Prudek, p. 28; Roman Sigaev, p. 279b; tristan3d, p. 91; verdateo, p. 99t.
AAP: Adam Davy/PA, p. 75t.
Age Fotostock: David Nunuk, p. 278b.
Alamy Stock Photo: Alchemy, pp. 154tl, 154tr; Stephen Barnes/Sport, p. 47; Oliver Burston, p. 296; dpa picture alliance, p. 60r; epa european pressphoto agency b.v., p. 60l; GL Archive, p. 176bl; Paul Hebditch, p. 75b; H.S. Photos, pp. 271tl, 271tr; imageBROKER, p. 193t; INTERFOTO, p. 294t; Marmaduke St. John, p. 158b; davide palermo, p. 163b; Photos 12, p. 285t; PRISMA ARCHIVO, p. 88bl; Sueddeutsche Zeitung Photo, pp. 262bl, 287; Tetra Images, p. 232; VintageCorner, p. 4c; WENN Ltd, p. 76l; ZUMA Press, Inc, p. 68l.
Bielasko: p. 106b.
Burrows, Keith: p. 152t.
CERN: ATLAS Experiment © 2017 CERN, p. 367.
DK Images: Stephen Oliver, p. 154cr.
EcoSwitch: Courtesy of EcoSwitch, p. 213.
European Space Agency: ESA/AOES Medialab, p. 14; ESA, p. 27t.
Fairfax Photo Sales: Eddie Jim/Fairfax Media, p. 166b.
Fotolia: 06photo, p. 159; Ammentorp, p. 118t; BillionPhotos.com, p. 279t; Joggie Botma, p. 134c; Ezume Images, p. 391; Michael Ireland, p. 134b; Craig Jewell, p. 134t; KAR, p. 392; Siegmar, p. 125t; tadamichi, p. 403.
Fundamental Photographs, NYC: Richard Megna, p. 242lr.
Georgia State University: Courtesy of Rod Nave, HyperPhysics Project, Department of Physics and Astronomy, Georgia State University, p. 262t.
Getty Images: Bellingham Herald/Contributor, p. 22b; Bettmann, p. 260t; Dr Tony Brain/Science Photo Library, pp. 290cc, 290cl, 290cr; Andrew Brookes, National Physical Laboratory, p. 317t; Tim Clayton, p. 50b; Hulton Deutsch, p. 374; Emilio Segre Visual Archives, p. 276; Gary Hincks/Science Photo Library, p. 6b; moodboard, p. 13cl; David Parker/IMI/University of Birmingham High TC Consortium, p. 200; Quinn Rooney, p. 73t; Detlev Van Ravenswaay/Science Photo Library, p. 165br; Royal Astronomical Society, p. 18br; Royal Institution of Great Britain, p. 182t; Science Photo Library, pp. 46, 187t, 220b, 311; Clive Streeter, p. 4b; Stringer, p. 217; Sheila Terry, p. 6c; Trevor Clifford Photography, p. 170; Trevor Clifford Photography/Science Photo Library, p. 155b.
Gold, Aaron: p. 110cl.
ICRAR: Radio image by Natasha Hurley-Walker (ICRAR/Curtin) and the GLEAM Team. MWA tile and landscape by Dr John Goldsmith/Celestial Visions, p. 257tr.
ImageFolk: Minden Pictures, p. 136b.
Leung, Terry: p. 245t.
McBroom, Alice: pp. 156c, 396b.
NASA: Courtesy of nasaimages.org. pp. 18bl, 25b, 87t; Ken Crawford, p. 370t; ESA/Hubble Telescope, p. 15t; European Space Agency/Planch Collatoration, p. 376b; Ed Hengeveld, p. 29; JPL, p. 18t; Dixon Rohr, p. 153b; Simon Swordy (U. Chicago) p. 358t; STSCI, p. 380b.
Newspix: Craig Greenhill, p. 80.
Retrospect Photography: Dale Mann, p. 135.
School Curriculum and Standard Authority, Government of Western Australia: WACE Physics ATAR Course Year 12 Syllabus © School Curriculum and Standard Authority, Government of Western Australia, 2017, pp. 1, 3, 39, 97, 133, 151, 181, 229–31, 303, 341, 386–7.
Science Photo Library: pp. 329, 258cl, 304b; Andrew Lambert Photography, pp. 242cr, 243bl, 243br, 298t; Martin Bond, p. 163t; Atlas Collaboration/CERN, p. 346b; CERN, pp. 361c, 366t; Carlos Clarivan, pp. 271b, 273; Henning Dalhoff, p. 291; DESY, p. 283bl; Emilio Segre Visual Archives/American Institute of Physics, p. 255; ESA/CE/EUROCONTROL, p. 313; Monique Feist/Paul Tafforeau/ESRF/Pascal Goetgheluck, p. 283br; Russell Kightley, p. 249b; Lawrence Berkeley Laboratory, p. 350b; LMSAL/Stanford University/NASA, p. 258cr; Walt Myers, p. 331; David Parker, pp. 285b, 358b, 358c, 366b; Scientifica/Visuals Unlimited, Inc., p. 260c; Alan Sirulnikoff, p. 394; Sputnik, p. 12t; Trevor Clifford Photography, pp. 235bl, 396t; Detlev van Ravenswaay, p. 381; Jim West, p. 169; Mark Williamson, p. 248t.
Science Source - Photo Researchers, Inc: GIPhotoStock, pp. 251b, 252b.
Shutterstock: AGorohov, pp. 99b, 100b, 100t, 101t; Miss Kanithar Aiumla-Or, p. 203tl; Gunnar Assmy, p. 55l; AstroStar, p. 341; Bedrin, p. 412tr; Elnur, p. 257b; Filipe Frazao, p. 5; FWStudio, p. 2–3; Markus Gann, p. 40t;Victoria M Gardner, p. 386–7; Gl0ck, p. 302–3; Arina P Habich, p. 84l; Daniel Hurlimann, p. 38–9; Iculig, p. 388; Iryna1, p. 4t; itsmejust, p. 258bl; Petri Jauhiainen, p. 150–1; Dmitry Kalinovsky, pp. 104b, 106t; Brian Kinney, p. 78r; saied shahin kiya, p. 258t; koya979, p. 277cr; KristinaSh, p. i; krsmanovic, p. 189b; langdu, p. 243cr; Joana Lopes, p. 405; maryo, p. 97; Dominik Michalek, p. 133; MilanB, p. 152c; Mopic, p. 241l; muzsy, p. 63bl; Marius Neacsa, p. 181; oliveromg, pp. 109t, 111; PeterPhoto123, p. 240cr; Pichitchai, p. 203tr; Plati Photography, p. 418; Claire Plumridge, p. 98; Keith Publicover, p. 7; pukach, p. 139t; Fouad A. Saad, p. 189t; Atiketta Sangasaeng, p. 241cr; Carlos E. Santa Maria, p. 272t; Sideways Design, p. 180tl; solarseven, p. 254t; David Steele, p. 64b; James Steidl, p. 55r; Surakit, pp. 102t, 104t, 105b; Targa56, p. 118b; tome213, p. 229, 410–3; Triff, p. 1, 226–8; Tristan3D, p. 25t; Vallka, p. 230–1; Taras Vyshnya, p. 140b; Westend61 Premium, p. 76r; zstock, p. 268.
The Picture Source: Terry Oakley, pp. 288bc, 288bl, p. 288br.
Twostaricebox: p. 199.

Pearson Australia
(a division of Pearson Australia Group Pty Ltd)
707 Collins Street, Melbourne, Victoria 3008
PO Box 23360, Melbourne, Victoria 8012
www.pearson.com.au

First published 2018 by Pearson Australia

eBook credits
Resource authors: Greg Moran
Publisher: Malcolm Parsons
Project Managers: Shelly Wang, Beth Zeme, Rekha Dureja
Production Managers: Anji Bignell, Natalie Lincoln
Development Editors: Zoe Hamilton, Vicky Chadfield
Editor: Marta Veroni, Elaine Cochrane
Rights and Permissions Editor: Amirah Fatin
Illustrator: DiacriTech
eBook Production Coordinator: Louise Grundy
Digital Administrator: Jenny Smirnov

ISBN 978 1 4886 1807 9

Pearson Australia Group Pty Ltd ABN 40 004 245 943