JACARANDA PHYSICS 12

FOR NSW | FOURTH EDITION

KAHNI BURROWS
DAN O'KEEFFE
MICHAEL ANDRIESSEN
MURRAY ANDERSON
GRAEME LOFTS
BRUCE MCKAY
PETER PENTLAND
ROSS PHILLIPS

jacaranda
A Wiley Brand

Fourth edition published 2018 by
John Wiley & Sons Australia, Ltd
42 McDougall Street, Milton, Qld 4064

First edition published 2001
Second edition published 2003
Third edition published 2008

Typeset in 11/14 pt Time LT Std

ISBN: 978-0-7303-5648-6

Front cover image: © sakkmesterke/Shutterstock

Illustrated by various artists, diacriTech and Wiley Composition Services

Typeset in India by diacriTech

Printed in Singapore by
Markono Print Media Pte Ltd

NATIONAL LIBRARY OF AUSTRALIA

A catalogue record for this book is available from the National Library of Australia

10 9 8 7 6 5 4 3 2

All activities have been written with the safety of both teacher and student in mind. Some, however, involve physical activity or the use of equipment or tools. All due care should be taken when performing such activities. Neither the publisher nor the authors can accept responsibility for any injury that may be sustained when completing activities described in this textbook.

CONTENTS

About eBookPLUS

jacaranda plus

This book features eBookPLUS: an electronic version of the entire textbook and supporting digital resources. It is available for you online at the JacarandaPLUS website (**www.jacplus.com.au**).

Using JacarandaPLUS

To access your eBookPLUS resources, simply log on to **www.jacplus.com.au** using your existing JacarandaPLUS login and enter the registration code. If you are new to JacarandaPLUS, follow the three easy steps below.

Step 1. Create a user account

The first time you use the JacarandaPLUS system, you will need to create a user account. Go to the JacarandaPLUS home page (**www.jacplus.com.au**), click on the button to create a new account and follow the instructions on screen. You can then use your nominated email address and password to log in to the JacarandaPLUS system.

Step 2. Enter your registration code

Once you have logged in, enter your unique registration code for this book, which is printed on the inside front cover of your textbook. The title of your textbook will appear in your bookshelf. Click on the link to open your eBookPLUS.

Step 3. Access your eBookPLUS resources

Your eBookPLUS and supporting resources are provided in a chapter-by-chapter format. Simply select the desired chapter from the table of contents. Digital resources are accessed within each chapter via the resources tab.

Once you have created your account, you can use the same email address and password in the future to register any JacarandaPLUS titles you own.

Using eBookPLUS references

eBookPLUS logos are used throughout the printed books to inform you that a digital resource is available to complement the content you are studying.

Searchlight IDs (e.g. **INT-0001**) give you instant access to digital resources. Once you are logged in, simply enter the Searchlight ID for that resource and it will open immediately.

Minimum requirements

JacarandaPLUS requires you to use a supported internet browser and version, otherwise you will not be able to access your resources or view all features and upgrades. The complete list of JacPLUS minimum system requirements can be found at **http://jacplus.desk.com**.

Troubleshooting

- Go to **www.jacplus.com.au** and click on the Help link.
- Visit the JacarandaPLUS Support Centre at **http://jacplus.desk.com** to access a range of step-by-step user guides, ask questions or search for information.
- Contact John Wiley & Sons Australia, Ltd.
  Email: support@jacplus.com.au
  Phone: 1800 JAC PLUS (1800 522 7587)

ACKNOWLEDGEMENTS

The authors and publisher would like to thank the following copyright holders, organisations and individuals for their assistance and for permission to reproduce copyright material in this book.

Images

AAP Newswire: figure **11.4**/ALAN PORRITT/AAP • Alamy Australia Pty Ltd: figure **6.2**/GL Archive/Alamy Stock Photo; figures **4.1**, **8.2**, **8.17**/Science History Images/Alamy Stock Photo; figure **8.15**/sciencephotos/Alamy Stock Photo; figure **8.16**/David A. Barnes/Alamy Stock Photos; figures **11.2**, **13.22**/World History Archive/Alamy Stock Photo; figure **12.7**/Granger Historical Picture Archive/Alamy Stock Photo; figure **12.10**/Photo 12/Alamy Stock Photo; figure **13.1**/Archive Pics/Alamy Stock Photo; figure **14.2**/Pictorial Press Ltd/Alamy Stock Photo; figures **8.21a**, **8.21b**, **8.21c**/Phil Degginger/Alamy Stock Photo • American Institute of Physics: figure **14.3**/AIP Emilio Segrè Visual Archives, W. F. Meggers Collection, Physics Today Collection • Australian Synchrotron: figure **11.27**/Image credit: ANSTO • Brookhaven National Laboratory: figures **16.15a**, **16.15b**/Courtesy of Brookhaven National Laboratory • Bruce McKay: figures **9.6b**, **9.6c**, **16.16**/Bruce McKay • CERN: figure **16.8**/CERN • Digital Stock: figure **1.8**/Digital Stock • European Space Agency Communication Department: figure **12.2**/© ESA and the Planck Collaboration • Fermilab: figures **16.1**, **16.13**/Reidar Hahn/Fermilab National Accelerator Laboratory • Fundamental Photographs: figure **9.6a**/©Ken Kay, Fundamental Photographs, NYC • Getty Images Australia: figure **6.5**/© Hulton-Deutsch Collection/CORBIS/Corbis via Getty Images; figure **6.21**/Marccophoto/Getty Images; figure **9.1**/GIPhotoStock/Getty Images; figures **10.2**, **10.5**, **10.11**/Bettman/Getty Images • Industrial Equipment & Control Pty Ltd: figure **6.41** and accompanying text/Industrial Equipment & Control Pty Ltd • iStockphoto: figure **13.5**/icholakov/iStockphoto; figure **13.20**/stockdevil royalty free - extended use • JAS Photography: figure **1.7**/John Sowden/JAS Photography • John Wiley & Sons Inc.: figure **15.10**/From: Halliday *et al.* 1997, *Fundamentals of Physics*, 5th edn, p. 1079, figure 43.6. Used by permission of John Wiley & Sons, Inc. • Lawrence Berkeley National: figures **16.9**, **16.10**/Adapted from The Particle Adventure, website produced by the Particle Data Group, Lawrence Berkeley National Laboratory • NASA: figures **1.5**, **8.23**, **12.34**, **12.40**, **16.3a**/NASA • Newspix: figure **11.9**/News Ltd/Newspix • Out of Copyright: figure **12.19** • Peter Pentland: figure **7.37**/Peter Pentland • Project CLEA: figure **12.30**/Project CLEA • Public Domain: figures **12.3**, **12.20**, **12.28** • Ragbir Bhathal, Dr: figures **12.31**, **12.32**/Dr. Ragbir Bhathal • Rice University: figure **9.12**/Rice University • Science & Society Pic. Library: figure **13.27**/Science Museum/Science & Society Picture Library • Science Photo Library: figure **7.6**/SCIENCE PHOTO LIBRARY; figure **9.9**/GIPhotoStock/SCIENCE PHOTO LIBRARY • Shelyak Instruments: figure **12.29**/Olivier Thizy/Shelyak Instruments • Shutterstock: figure **1.1**/Ventura/Shutterstock; figure **1.4**/Mauricio Graiki/Shutterstock; figure **1.22**/Diego Barbieri/Shutterstock; figure **2.1**/Racefotos08/Shutterstock; figure **2.2**/muzsy/Shutterstock; figure **2.4**/Evdoha_spb/Shutterstock; figure **2.13a**/Pavel Ilyukhin/Shutterstock; figure **2.14a**/jennyt/Shutterstock; figure **2.19**/Rena Schild/Shutterstock; figure **2.20**/CHEN WS/Shutterstock; figure **2.21**/Shahjehan/Shutterstock; figure **2.28**/Curioso/Shutterstock; figure **3.1**/3Dsculptor/Shutterstock; figure **4.10a**/Fineart1/Shutterstock; figures **5.1a**, **5.1b**, **5.11**/Kim Christensen/Shutterstock; figure **6.3**/Everett Historical/Shutterstock; figure **6.26**/Roger Bruce/Shutterstock; figure **6.27**/New Anawach/Shutterstock; figure **7.1**/wi6995/Shutterstock; figure **7.15**/MARGRIT HIRSCH/Shutterstock; figure **7.32**/Emilia Stasiak/Shutterstock; figure **8.1**/Katarina Christenson/Shutterstock; figure **8.5a**/imagedb. com/Shutterstock; figure **8.5b**/Andrei S/Shutterstock; figure **8.18**/Slaven/Shutterstock; figure **8.19**/Christina Richards/Shutterstock; figure **8.20**/freevideophotoagency/Shutterstock; figure **9.2**/ Mellimage/Shutterstock; figure **9.3**/Roberto Lo Savio/Shutterstock; figure **9.21**/martinnovacek. cz/Shutterstock; figure **9.22**/Marco Crupi/Shutterstock; figure **9.23a**/Vasya Prekrasnaya/Shutterstock; figure **9.23b**/filippo giuliani/Shutterstock; figure **10.1**/ILYA GENKIN/Shutterstock; figure **11.1**/imagIN.gr photography/Shutterstock; figure **11.3**/Ben Jeayes/Shutterstock; figure **11.6**/Georgios Kollidas/Shutterstock; figure **11.20**/dennizn/Shutterstock; figure **11.28**/Triff/Shutterstock; figures **12.1**, **12.23**/NASA images/Shutterstock; figure **12.6**/Sander van Sinttruye/Shutterstock; figure **12.9**/Wolfgang Kloehr/Shutterstock; figure **12.11**/PlanilAstro/Shutterstock; figure **12.12**/Owen65/Shutterstock; figure **12.18**/banderlog/Shutterstock; figure **12.21**/David Fowler/Shutterstock; figure **12.24**/general-fmv/Shutterstock; figure **13.21b**/magnetix/Shutterstock; figure **14.1**/Fotos593/Shutterstock; figure **15.1a**/Rob Crandall/Shutterstock; figure **15.1b**/Efman/Shutterstock; figure **16.2**/IgorGolovniov/Shutterstock • Sloan Digital Sky Survey: figure **12.42**/Sloan Digital Sky Survey • Springer: figure **9.24**/Springer • The National Museum of American History: figure **12.22**/Rowland Solar Spectrum, Division of Medicine & Science, National Museum of American History, Smithsonian Institution • TransGrid: figure **6.25**/Transgrid • Wilson Transformers: figure **6.1**, **6.22**/Wilson Transformer Company Pty Ltd, Australia

Text

CSIRO: **Table 12.4**/CSIRO • NSW Physics Stage 6 Syllabus (2017) © Copyright Board of Studies, Teaching and Educational Standards New South Wales for and on behalf of the Crown in right of the State of New South Wales, 2017.

Every effort has been made to trace the ownership of copyright material. Information that will enable the publisher to rectify any error or omission in subsequent reprints will be welcome. In such cases, please contact the Permissions Section of John Wiley & Sons Australia, Ltd.

TOPIC 1
Projectile motion

1.1 Overview

1.1.1 Module 5: Advanced Mechanics

Projectile motion

Inquiry question: How can models that are used to explain projectile motion be used to analyse and make predictions?

Students:

- analyse the motion of projectiles by resolving the motion into horizontal and vertical components, making the following assumptions:
  - a constant vertical acceleration due to gravity
  - zero air resistance
- apply the modelling of projectile motion to quantitatively derive the relationships between the following variables:
  - initial velocity
  - launch angle
  - maximum height
  - time of flight
  - final velocity
  - launch height
  - horizontal range of the projectile (ACSPH099)
- conduct a practical investigation to collect primary data in order to validate the relationships derived above
- solve problems, create models and make quantitative predictions by applying the equations of motion relationships for uniformly accelerated and constant rectilinear motion.

FIGURE 1.1 This multiple exposure image of a trick jump performed by an FMX rider clearly shows the parabolic shape of the trajectory of rider and bike. This pathway results from the combined effects of ramp angle, initial velocity and the downward acceleration of gravity.

1.2 Falling under gravity

1.2.1 Falling through air

In your earlier studies, you have encountered the concept of **weight**, which describes the attractive force that the Earth exerts on objects within the effect of its gravitational field. The weight of an object is calculated by using the equation

$$\boldsymbol{w} = m\boldsymbol{g}$$

where m is the object's mass and $\boldsymbol{g}$ is the acceleration due to gravity.

Whenever we raise an object from the surface of the Earth and release it, the object falls towards the centre of the Earth. Whether the object is a rock, a feather, a glass of milk or a pencil, every dropped object will move downwards unless something gets between it and the Earth's surface. However, as you may have noticed, not everything falls downwards at the same rate. For example, a feather and a hammer dropped from the same height will hit the ground at different times — the hammer will strike the ground first, then the feather. Making a variety of such observations throughout our lives, we might come to the same conclusion that the ancient Greeks did — that heavy objects fall faster than light objects.

The Italian scientist Galileo Galilei (1564–1642) also noticed that big, heavy objects usually reached the ground faster than small, light ones. However, he had the idea that this behaviour had nothing to do with the masses of the objects themselves but was actually due to the combined effect of their weight (which pulls them downwards) and some sort of resisting force (which pulls them upwards) acting on them as they fell. If the resistant force acting on an object was comparable to its weight, the object fell more slowly. If the resistant force was comparatively small, then it fell more quickly.

This resistant force that acts to oppose an object's motion through the air is known as **air resistance** or **drag** ($\boldsymbol{F}_{drag}$).

The net force acting on an object falling freely under the influence of gravity is equal to the vector sum of its weight and the drag acting upon it: $\boldsymbol{F}_{net} = \boldsymbol{w} + \boldsymbol{F}_{drag}$

As $\boldsymbol{F}_{net} = m\boldsymbol{a}$, we may also write:

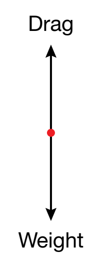
$$m\boldsymbol{a} = \boldsymbol{w} + \boldsymbol{F}_{drag}$$

Therefore,

$$\boldsymbol{a} = \frac{\boldsymbol{w} + \boldsymbol{F}_{drag}}{m}$$

From this, we can see that the rate at which an object accelerates as it falls towards the ground does not depend entirely upon the object's mass or weight.

FIGURE 1.2 The forces acting on a vertically falling object.



1.2.2 Air resistance — what a drag!

Drag is caused by the collision of a falling object with the particles in the air; this has the effect of slowing the object's descent. The greater the number of particles that the object collides with, the greater the size of the drag it experiences.

The drag force exerted on an object is directly proportional to the following:

- the object's cross-section perpendicular to its direction of motion. An object with a large cross-section will collide with more air particles than one with a small cross-section, so it would experience a greater drag.
- the density of the air through which the object moves. Close to the Earth's surface, the air has a density of about 1.2 kg m^{-3}, which is much denser than at, say, an altitude of 10 km, where the air density is around 0.4 kg m^{-3}. This means that an object falling at the 10 km mark will collide with fewer particles and so will experience less drag than an object falling nearer the ground.
- the square of the object's speed. An object travelling at a speed of 10 m s^{-1} will encounter four times the magnitude of drag than if it were to travel at 5 m s^{-1}.

Drag is also affected by the object's shape, the smoothness of its surface and the amount of turbulence that these factors cause in the air as the object falls.

FIGURE 1.3 An object with a larger cross-section collides with more air particles than one with a smaller cross-section.

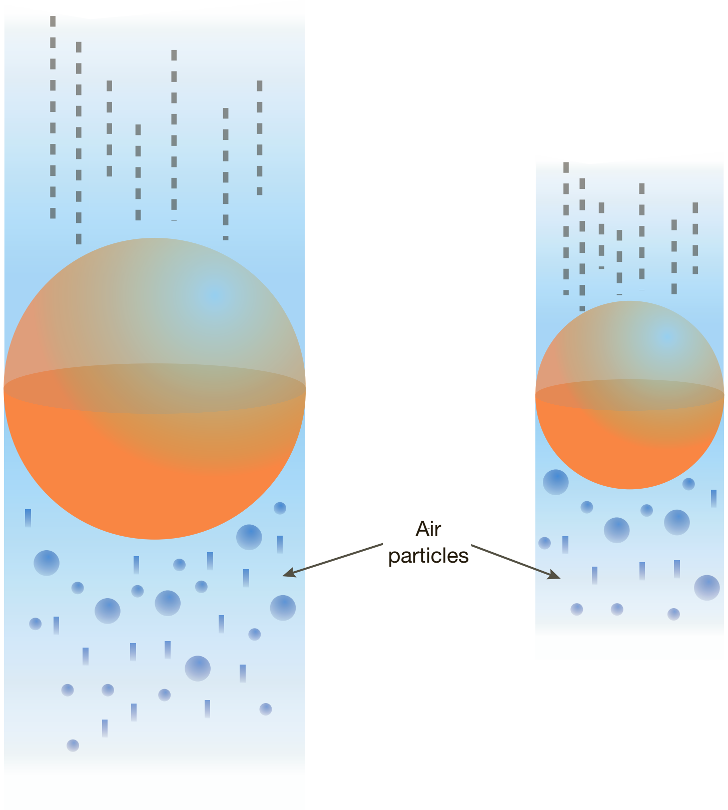


FIGURE 1.4 Formation skydivers spread their limbs out to increase their surface area perpendicular to their direction of fall. This increases their drag, which decreases their descent speed, extending the time that they are in the air.

1.2 SAMPLE PROBLEM 1

At what speed will a 70 kg skydiver be travelling two seconds after stepping out of the plane if we assume a uniform drag force of 20 N acts on him during this time?

SOLUTION:

Given: $\boldsymbol{u} = 0$; $t = 2$ s; $m = 70$ kg; $\boldsymbol{g} = -9.8$ m s^{-2}; $\boldsymbol{F}_{drag} = 20$ N

We need to find $\boldsymbol{v}$.

First, we need to find the net acceleration acting on the skydiver.

Remember, he has two forces acting on him: weight $\boldsymbol{w}$, acting downwards, and $\boldsymbol{F}_{drag}$, acting upwards. This allows us to determine the net force acting on the skydiver:

$$\boldsymbol{F}_{net} = \boldsymbol{w} + \boldsymbol{F}_{drag}$$

Now, as $\boldsymbol{w} = m\boldsymbol{g}$, the weight of the skydiver will be:

$$\begin{aligned}\boldsymbol{w} &= (70\text{ kg})\ (-9.8\text{ m s}^{-2})\\ &= -686\text{ N}\end{aligned}$$

So,

$$\begin{aligned}\boldsymbol{F}_{net} &= (-686\text{ N}) + 20\text{ N}\\ &= -666\text{ N}\end{aligned}$$

As $\boldsymbol{F}_{net} = m\boldsymbol{a}$, we can now find the skydiver's acceleration during the 2 s interval:

$$\boldsymbol{a} = \frac{\boldsymbol{F}_{net}}{m}$$

$$= \frac{-666\ \text{N}}{70\ \text{kg}} = -9.5\ \text{m s}^{-2}$$

We can now find the skydiver's speed at the end of the 2 s interval.

Given: $t = 2$ s; $\boldsymbol{u} = 0\ \text{m s}^{-1}$; $\boldsymbol{a} = -9.5\ \text{m s}^{-2}$; $\boldsymbol{v} = ?$

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{a}\ t$$

$$\begin{aligned}\boldsymbol{v} &= (0\ \text{m s}^{-1}) + (-9.5\ \text{m s}^{-2})(2\ \text{s}) \\ &= -19\ \text{m s}^{-1}\end{aligned}$$

The skydiver will be travelling at $19\ \text{m s}^{-1}$ downwards at the end of the 2 s interval.

WORKING SCIENTIFICALLY 1.1

If you have watched a parachute filling with air as a skydiver deploys it, you would have noticed that the parachute does not form a flat surface as it catches the air; rather, it forms a specific three-dimensional shape according to the way that the parachute has been designed. While many parachutes form dome-like shapes as they catch the air, others form shapes that are more like sausages. As a result of their different geometries, parachutes containing the same surface area of fabric will differ in the volume of air that they 'capture' during descent. Explore the different shapes of parachutes that can be made from a piece of parachute fabric with a surface area of $900\ \text{cm}^2$, determining the effective volume of each when it is dropped with a 10 g payload. Use three of these shapes to determine what relationship (if any) exists between the volume of the parachute and the drag force that it exerts.

1.2.3 Falling in a vacuum

In 1971, Apollo 15 astronaut David Scott dropped a hammer and a feather from the same height while standing on the surface of the Moon (yes, scientists really did pack a hammer and a feather on the spacecraft just so they could test this). The hammer and the feather fell at the same rate and hit the ground at the same time.

FIGURE 1.5 Astronaut David Scott drops a hammer and a feather on the surface of the Moon.

As discussed in the previous section, drag is caused by collisions of air particles with a falling object. The Moon's gravity is too low to hold a thick atmosphere around it like the Earth can, so the hammer and the feather both fell without encountering enough particles to cause any measurable drag. As F_{drag} was equal to zero, the only force acting on the two objects was the attractive force of the Moon's gravity acting towards the centre of the Moon; that is, $\boldsymbol{F}_{net} = \boldsymbol{w}$.

For both the hammer and the feather,

$$m\boldsymbol{a} = m\boldsymbol{g}$$

Cancelling mass m on both sides, we get:

$$\boldsymbol{a} = \boldsymbol{g}$$

In other words, the masses of the two objects had no influence at all on the rate at which the objects fell. They both fell at about $1.6\ \text{m s}^{-2}$ (the acceleration due to gravity on the Moon) and hit the dust (literally) at the same time.

What this little experiment on the Moon proved was that Galileo's theory was correct. In a vacuum, all objects will fall at the same rate, regardless of mass and size.

1.2 SAMPLE PROBLEM 2

(a) How fast will a ball be travelling when it strikes the ground if it is dropped from a height of 100 m? (Ignore air resistance.)

(b) At what speed would the same ball be travelling if it was dropped from the same height on the Moon's surface where gravity is one-sixth that on Earth?

SOLUTION:

(a) $\boldsymbol{u} = 0\,\text{m}\,\text{s}^{-1}$; $\boldsymbol{s} = -100\,\text{m}$; $\boldsymbol{a} = \boldsymbol{g} = -9.8\,\text{m}\,\text{s}^{-2}$; $\boldsymbol{v} = ?$

$$\boldsymbol{v}^2 = \boldsymbol{u}^2 + 2\ \boldsymbol{a}\ \boldsymbol{s}$$

$$\boldsymbol{v}^2 = (0)^2 + 2(-9.8\,\text{m}\,\text{s}^{-2})(-100\ \ \text{m})$$

$$\Rightarrow \boldsymbol{v} = \sqrt{1960} = \pm 44.3\,\text{m}\,\text{s}^{-1}$$

Because the ball is falling, it will be travelling downwards when it strikes the ground, so $\boldsymbol{v}$ must be a negative quantity.

Therefore, if there is no air resistance, the ball is travelling at $-44.3\,\text{m}\,\text{s}^{-1}$ when it strikes the ground.

(b) $\boldsymbol{u} = 0\,\text{m}\,\text{s}^{-1}$; $\boldsymbol{s} = -100\ \ \text{m}$; $\boldsymbol{a} = \dfrac{\boldsymbol{g}}{6} - 1.6\,\text{m}\,\text{s}^{-2}$; $\boldsymbol{v} = ?$

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

$$\boldsymbol{v}^2 = (0)^2 + 2(-1.6\,\text{m}\,\text{s}^{-2})(-100\,\text{m})$$

$$\Rightarrow \boldsymbol{v} = \sqrt{320} = \pm 17.9\,\text{m}\,\text{s}^{-1}$$

The ball would strike the Moon's surface with a velocity of $-17.9\,\text{m}\,\text{s}^{-1}$.

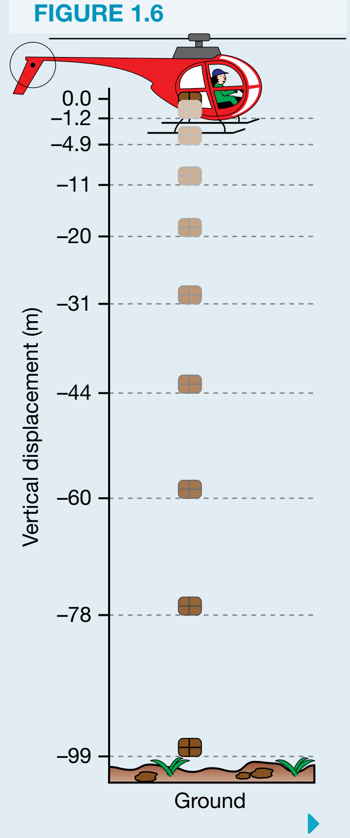
1.2 SAMPLE PROBLEM 3

A helicopter delivering supplies to a flood-stricken farm hovers 100 m above the ground. A package of supplies is dropped from rest, just outside the door of the helicopter. Air resistance can be ignored.

(a) Calculate how long it takes the package to reach the ground.

(b) Calculate how far from its original position the package has fallen after 0.50 s, 1.0 s, 1.5 s, 2.0 s etc. until the package has hit the ground. (You may like to use a spreadsheet here.) Draw a scale diagram of the package's position at half-second intervals.

FIGURE 1.6



SOLUTION:

(a) $\boldsymbol{u} = 0\,\text{m}\,\text{s}^{-1}$; $\boldsymbol{s} = -100\ \ \text{m}$; $\boldsymbol{a} = \boldsymbol{g} = -9.8\,\text{m}\,\text{s}^{-2}$; $t = ?$

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

$$(-100\,\text{m}) = (0\,\text{ms}^{-1})t + \frac{1}{2}(-9.8\,\text{ms}^{-2})t^2$$

$$t = \sqrt{\frac{100}{4.9}} = \pm 4.52\,\text{s},\ \text{rounded to} \pm 4.5\,\text{s}$$

The negative square root can be ignored here as we are interested only in motion that has occurred after the package was released at $t = 0$ (i.e. positive times).

Hence, the package takes 4.5 s to reach the ground.

(b) $t = 0.50\,\text{s}$; $\boldsymbol{u} = 0\,\text{m s}^{-1}$; $\boldsymbol{a} = \boldsymbol{g} = -9.8\,\text{m s}^{-2}$; $\boldsymbol{s} = ?$

$$\boldsymbol{s} = \boldsymbol{u}t + \frac{1}{2}\boldsymbol{a}t^2$$
$$= (0)(0.5\text{s}) + \frac{1}{2}(9.8\,\text{m s}^{-2})(0.5\ \text{s})^2$$
$$= 1.23\ \text{m, rounded to } 1.2\ \text{m}$$

Repeat this for $t = 1$ s, 1.5 s, 2 s etc. to gain the results listed in Table 1.1. The scale diagram is shown in Figure 1.6.

TABLE 1.1 Vertical distance travelled over time

Time (s)	0.50	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
Vertical distance (m)	1.2	4.9	11	20	31	44	60	78	99

1.2.4 Terminal velocity

In the absence of an atmosphere, an object dropped from any height will continue to accelerate until it hits the ground. This is certainly the case with objects dropped on the Moon, but on Earth the effects of air resistance cannot be shrugged off nearly so easily. When an object falls, remember that it is subject to two forces — weight and drag. The weight stays the same throughout the drop, but drag gets larger as the velocity increases. When the object is first dropped and its initial speed is zero, the drag is equal to zero. As the object accelerates under the effect of gravity, it speeds up, so the drag force increases too, reducing the actual rate at which the object accelerates downwards. Eventually, if the object falls for long enough, it will reach a velocity where the downward force of its weight is exactly equal to the upward force of drag. As the net force acting on the object is now zero, the object stops accelerating and will continue to fall at a constant velocity. The velocity at which this occurs is called **terminal velocity**. Once an object has reached its terminal velocity, it will maintain that velocity for the rest of its descent.

Different objects tend to have different terminal velocities. A human falling without a parachute has a terminal velocity of about 200 km h^{-1}, while a cat has a terminal velocity of about 120 km h^{-1}. Very small creatures, such as insects, have very low terminal velocities — about 8 km h^{-1}. This is why a bug will survive a fall of 100 m and walk away, but a human will not.

WORKING SCIENTIFICALLY 1.2

The magnitude of the drag force acting on an object as it falls through a fluid is summarised in the equation

$$\boldsymbol{F}_{\boldsymbol{drag}} = \frac{1}{2}C_D\,\rho\,\boldsymbol{v}^2 A$$

where C_D is the drag coefficient for the object, ρ is the density of the fluid, $\boldsymbol{v}$ is the speed at which the object is falling and $\boldsymbol{A}$ is the cross-sectional area of the object perpendicular to its motion.

The drag coefficient is very difficult to calculate, even for regularly shaped objects, and most values are determined experimentally. Design a method by which the drag force acting on a steel ball bearing falling through a dense fluid such as honey may be measured and its drag coefficient determined. Then use this method to determine whether polished ball bearings of different radii have different drag coefficients.

1.2 Exercise 1

1. A student has two identical pieces of paper, one of which she scrunches up into a little ball. Holding both the scrunched-up piece of paper and the un-scrunched piece of paper at the same height, she lets them go. Explain why, despite having identical compositions and masses, the two pieces of paper land at different times.
2. Explain why an object falls faster at an altitude of 10 km than it does at sea level.
3. When skydivers wish to make a very fast descent, they will jump feet first with their bodies very straight and parallel to their direction of motion and their arms held close to them. Why does this work?
4. Why does a feather fall more slowly than a brick released at the same time from the same height?
5. Calculate the final speed of a hammer dropped from a height of 5 m in the absence of air resistance.
6. From what height does a flowerpot fall if it is travelling at a speed of 6.2 m s^{-1} when it strikes the ground? Ignore air resistance.
7. A ball is thrown downwards at 2 m s^{-1} and it travels 5.0 m before striking the ground. At what speed was it travelling when it hit the ground if we ignore air resistance?
8. A camera is dropped by a tourist from a lookout and falls vertically to the ground. The thud of the camera hitting the hard ground below is heard by the tourist 3.0 seconds later. Air resistance and the time taken for the sound to reach the tourist can be ignored.
   (a) How far did the camera fall?
   (b) What was the velocity of the camera when it hit the ground below?

1.3 Projectile motion

1.3.1 Trajectory

Any object that is launched into the air is a projectile. A basketball thrown towards a goal, a trapeze artist soaring through the air, and a package dropped from a helicopter are all examples of projectiles.

The **trajectory** of a projectile is the path that it follows during its flight. Except for those projectiles whose motion is initially straight up or down, or those that have their own power source (like a guided missile), projectiles generally follow a parabolic path. Deviations from this path can be caused either by air resistance, by spinning of the object or by wind. These effects are often small and can be ignored in many cases. A major exception, however, is the use of spin in many ball sports, but this effect will not be dealt with in this book.

FIGURE 1.7 A stroboscopic photograph of a ball undergoing projectile motion. A **stroboscope** is a light that produces quick flashes at regular (usually small) time periods. If used with a camera, instead of a regular flashgun, a stroboscopic photograph is produced, which shows multiple images of a moving object.

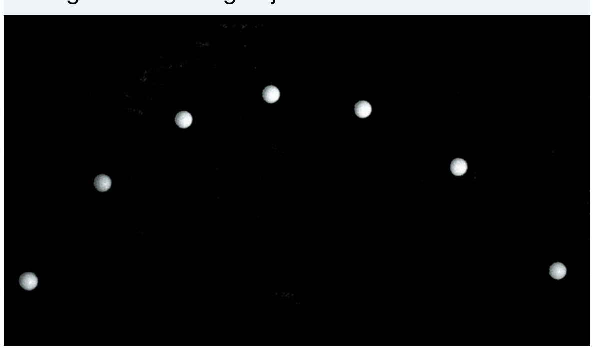
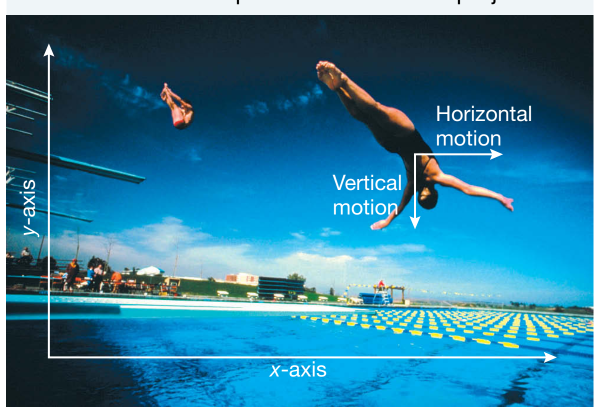


FIGURE 1.8 A frame of reference for the vertical and horizontal component motions of a projectile.



To understand and analyse this motion we must note an observation first made by Galileo: the motion of a projectile can be regarded as two separate and independent motions superimposed upon each other. The first is a vertical motion, which is subject to acceleration due to gravity, and the second is a horizontal motion, which experiences no acceleration. These two motions can be placed within a two-dimensional frame of reference, using the *y*-axis for the vertical motion and the *x*-axis for the horizontal motion.

Because the two motions are perpendicular, and therefore independent, we can treat them separately and analyse them separately.

1.3.2 What goes up must come down

Most projectiles are set in motion with velocity. The simplest case is that of a ball thrown directly upwards. The only force acting on the ball is that of gravity (ignoring air resistance). The ball accelerates downwards while moving upwards. Initially, this results in the ball slowing down. Eventually, it comes to a halt, then begins to move downwards, speeding up as it goes. Notice that, when air resistance is ignored, the motion of the ball is identical whether it is going up or coming down. The ball will return with the same speed with which it was projected. Throughout the motion illustrated in Figure 1.9 (and for which graphs are shown in Figure1.10), the acceleration of the ball is a constant 9.8 m s^{-2} downwards. A common error made by physics students is to suggest that the acceleration of the ball is zero at the top of its flight. If this were true, would the ball ever come down?

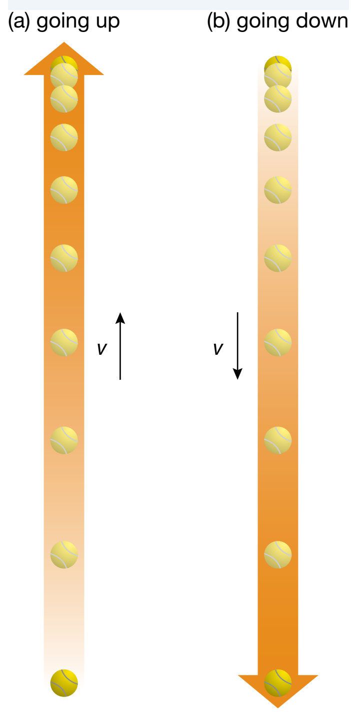


FIGURE 1.9 The motion of a ball projected vertically upwards. (a) going up (b) going down.

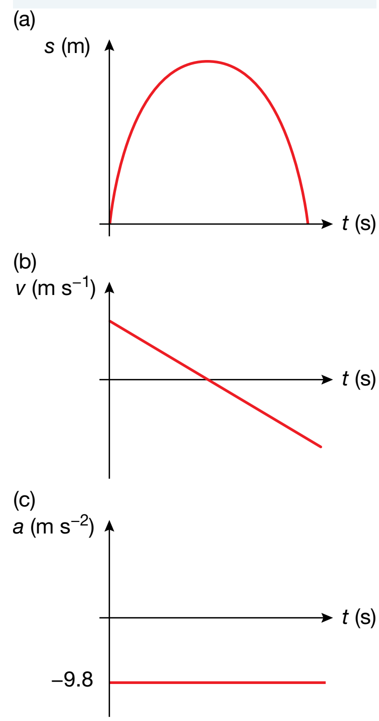


FIGURE 1.10 Graphs of motion for a ball thrown straight upwards.

AS A MATTER OF FACT

The axiom ‘what goes up must come down’ applies equally to bullets as it does to balls. Unfortunately, this means that people sometimes get killed when they shoot guns straight up into the air. If the bullet left the gun at a speed of 60 m s^{-1}, it will return to Earth at roughly the same speed. This speed is well and truly fast enough to kill a person who is hit by the returning bullet.

1.3 SAMPLE PROBLEM 1

A dancer jumps vertically upwards with an initial velocity of $4.0\,\mathrm{m\,s^{-1}}$. Assume the dancer's centre of mass was initially 1.0 m above the ground, and ignore air resistance.

(a) How long did the dancer take to reach her maximum height?

(b) What was the maximum displacement of the dancer's centre of mass?

(c) What is the acceleration of the dancer at the top of her jump?

(d) Calculate the velocity of the dancer's centre of mass when it returns to its original height above the ground.

SOLUTION:

There are several ways of arriving at the same answer. As has been done in this example, it is always good practice to minimise the use of answers from previous parts of a question. This makes your answers more reliable, preventing a mistake made earlier on from distorting the accuracy of your later calculations. For this problem, assign up as positive and down as negative.

(a) $\boldsymbol{u} = 4.0\,\mathrm{m\,s^{-1}}$; $\boldsymbol{a} = \boldsymbol{g} = -9.8\,\mathrm{m\,s^{-2}}$; $\boldsymbol{v} = 0\,\mathrm{m\,s^{-1}}$ (as the dancer comes to a halt at the highest point of the jump); $t = ?$

$$\boldsymbol{v} = \boldsymbol{u} + \mathbf{a}\,t$$

$$(0) = \left(4.0\,\mathrm{m\,s^{-1}}\right) + \left(-9.8\,\mathrm{m\,s^{-2}}\right)\,t$$

$$\Rightarrow t = \frac{4.0}{9.8}$$

$$= 0.41\,\mathrm{s}$$

The dancer takes 0.41 s to reach her highest point.

(b) $\boldsymbol{u} = 4.0\,\mathrm{m\,s^{-1}}$; $\boldsymbol{a} = \boldsymbol{g} = -9.8\,\mathrm{m\,s^{-2}}$; $\boldsymbol{v} = 0\,\mathrm{m\,s^{-1}}$ (as the dancer comes to a halt at the highest point of the jump); $s = ?$

$$\boldsymbol{v}^2 = \boldsymbol{u}^2 + 2\boldsymbol{a}\boldsymbol{s}$$

$$(0)^2 = (4.0\,\mathrm{m\,s^{-1}})^2 + 2\left(-9.8\,\mathrm{m\,s^{-2}}\right)s$$

$$\Rightarrow s = \frac{16}{19.6}$$

$$s = 0.82\,\mathrm{m}$$

The maximum displacement of the dancer's centre of mass is 0.82 m.

(c) At the top of the jump, the only force acting on the dancer is the force of gravity (the same as at all other points of the jump). Therefore, the acceleration of the dancer is acceleration due to gravity: $9.8\,\mathrm{m\,s^{-2}}$ downwards.

(d) For this calculation, only the downwards motion needs to be investigated.
$\boldsymbol{u} = 0\,\mathrm{m\,s^{-1}}$ (as the dancer comes to a halt at the highest point of the jump);
$\boldsymbol{a} = \boldsymbol{g} = -9.8\,\mathrm{m\,s^{-2}}$; $\boldsymbol{s} = -0.82\,\mathrm{m}$ (as the motion is downwards); $v = ?$

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

$$v^2 = (0)^2 + 2\left(-9.8\,\mathrm{m\,s^{-2}}\right)(-0.82\,\mathrm{m})$$

$$v = -4.0\,\mathrm{m\,s^{-1}}$$

(Note: Here, the negative square root is used, as the dancer is moving downwards. Remember, the positive and negative signs show direction only.)

The velocity of the dancer's centre of mass when it returns to its original height is $4.0\,\mathrm{m\,s^{-1}}$ downwards.

1.3.3 Horizontally launched projectiles

If a ball is launched horizontally from the end of a table, it does not simply drop vertically downwards as soon as it clears the table edge. Instead, it follows a semi-parabolic trajectory. The curved flight path of the ball is due to the fact that it is undergoing motion in two dimensions at the same time — horizontal motion and vertical motion. These two motions can be described in terms of two independent vectors: $\boldsymbol{v}_x$, the horizontal component of the ball's velocity, and $\boldsymbol{v}_y$, the vertical component of the ball's velocity.

The only force acting on the ball once it has been released is gravity (ignoring air resistance). As the force of gravity is the same regardless of the motion of the ball, the ball will still accelerate downwards at the same rate as if it were dropped. As the ball falls, the magnitude of $\boldsymbol{v}_y$ increases.

By considering the vertical velocity profile of the ball to be identical to that of an object that is simply dropped vertically, the value of the vertical velocity $\boldsymbol{v}_y$ can be calculated:

$$\boldsymbol{v}_y = \boldsymbol{u}_y + \boldsymbol{g}t$$

where $\boldsymbol{u}_y$ is the initial vertical velocity, $\boldsymbol{g}$ is the downwards acceleration due to gravity and t is equal to the time of flight.

As the initial vertical velocity will be equal to zero,

$$\boldsymbol{v}_y = (0) + \boldsymbol{g}t$$

$$\Rightarrow \boldsymbol{v}_y = \boldsymbol{g}t$$

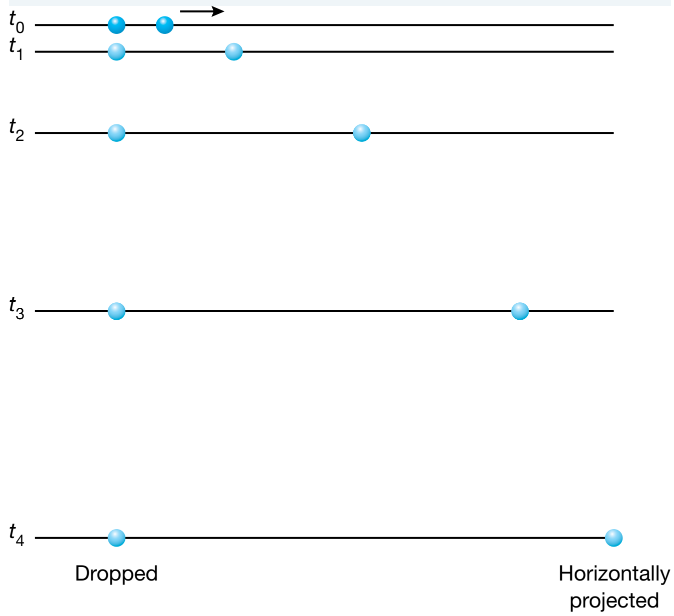
The vertical displacement, $\boldsymbol{y}$, of the ball as it drops during the time t can also be found:

$$\boldsymbol{y} = \boldsymbol{u}_y t + \frac{1}{2}\boldsymbol{g}t^2$$

$$\Rightarrow \boldsymbol{y} = (0)\,t + \frac{1}{2}\boldsymbol{g}t^2$$

$$\Rightarrow \boldsymbol{y} = \frac{1}{2}\boldsymbol{g}t^2$$

FIGURE 1.11 A horizontally projected object will fall at the same rate as it would if it were dropped.

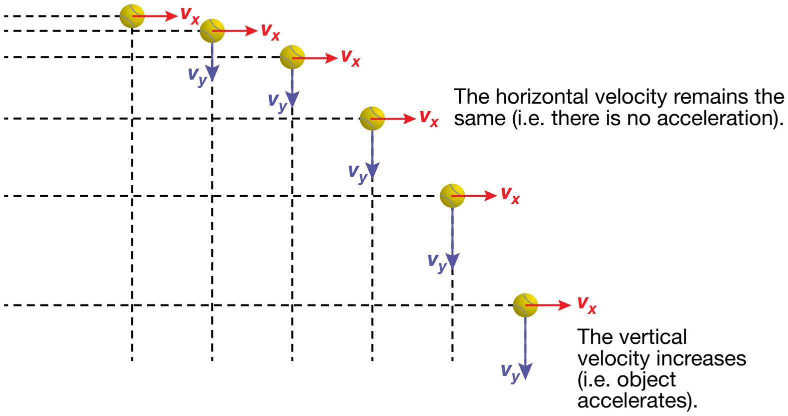


The horizontal component of the ball's motion, $\boldsymbol{v}_x$, remains unchanged for the duration of the ball's flight because there is no accelerating force acting horizontally on the ball. The horizontal displacement of the ball, $\boldsymbol{x}$, from its starting point is then simply:

$$\boldsymbol{x} = \boldsymbol{v}_x t$$

It is the constant horizontal velocity and changing vertical velocity that gives projectiles their characteristic parabolic motion. Notice that the vertical distance travelled by the ball in each time period increases, but that the horizontal distance is constant.

FIGURE 1.12 Position of a horizontally projected ball at constant time intervals.



1.3.4 Keep them separated

In modelling projectile motion, the vertical and horizontal components of the motion are treated separately.

1. The total time taken for the projectile motion is determined by the vertical part of the motion as the projectile cannot continue to move horizontally once it has hit the ground, the target or whatever else it might collide with.
2. This total time can then be used to calculate the horizontal displacement, or **range**, over which the projectile travels.

1.3 SAMPLE PROBLEM 2

Imagine the helicopter described in Section 1.2 Sample Problem 3 is not stationary, but is flying at a slow and steady speed of 20 m s^{-1} and is 100 m above the ground when the package is dropped.

(a) Calculate how long it takes for the package to hit the ground.
(b) What is the range of the package?
(c) Calculate the vertical distance the package has fallen after 0.50 s, 1.0 s, 1.5 s, 2.0 s, etc. until the package has reached the ground. (You may like to use a spreadsheet here.) Then calculate the corresponding horizontal distance, and hence draw a scale diagram of the package's position at half-second intervals. Remember, the horizontal and vertical components of the package's motion must be considered separately.
(d) Determine the velocity at which the package strikes the ground.

SOLUTION:

(a) In this part of the question, the vertical component is important.
Vertical component: $\boldsymbol{u_y} = 0\ \text{m s}^{-1}$; $\boldsymbol{y} = -100\ \text{m}$; $\boldsymbol{a_y} = \boldsymbol{g} = -9.8\ \text{m s}^{-2}$; $t = ?$

$$\boldsymbol{y} = \boldsymbol{u_y} t + \frac{1}{2}\boldsymbol{a_y} t^2$$

$$-100\ \text{m} = (0)\, t + \frac{1}{2}\left(-9.8\ \text{m s}^{-2}\right) t^2$$

$$\Rightarrow \frac{-100}{-4.9} = t^2$$

$t = 4.52$ s, rounded to 4.5 s

(Note: The positive square root is taken as we are concerned only with what happens after $t = 0$.)

(b) The range of the package is the horizontal distance over which it travels until it lands on the ground. It is the horizontal component of velocity that must be considered here.

Horizontal component: $\boldsymbol{u_x} = \boldsymbol{v_x} = 20\,\text{m}\,\text{s}^{-1}$ (The initial velocity of the package is the same as the velocity of the helicopter in which it has been travelling.)

$\boldsymbol{a_x} = 0\,\text{m}\,\text{s}^{-2}$ (No forces act horizontally so there is no horizontal acceleration.)

$t = 4.5\,\text{s}$ (from part (a) of this example); $\boldsymbol{x} = ?$

$$\begin{aligned} x &= \boldsymbol{u_x} t + \frac{1}{2}\boldsymbol{a_x} t^2 \\ &= \left(20\,\text{m}\,\text{s}^{-1}\right)(4.5\,\text{s}) + \frac{1}{2}(0)(4.5\,\text{s})^2 \\ &= 90\,\text{m} \end{aligned}$$

(c) The calculations for t = 0.5 s are shown here in Table 1.2:

TABLE 1.2 Vertical and horizontal components of the package's motion.

Vertical component	Horizontal component
$\boldsymbol{u_y} = 0\,\text{m}\,\text{s}^{-1}$; $t = 0.5\,\text{s}$; $\boldsymbol{a_y} = \boldsymbol{g} = -9.8\,\text{m}\,\text{s}^{-2}$; $\boldsymbol{y} = ?$	$\boldsymbol{u_x} = 20\,\text{m}\,\text{s}^{-1}$; $t = 0.5\,\text{s}$, $\boldsymbol{x} = ?$
$\boldsymbol{y} = \boldsymbol{u_y} t + \frac{1}{2}\boldsymbol{a_y} t^2$ $= 0 + \frac{1}{2}(-9.8\,\text{m}\,\text{s}^{-2})(0.5\,\text{s})^2$ $= -1.225$ (rounded to $-1.2\,\text{m}$)	$\boldsymbol{x} = \boldsymbol{u_x} t$ $= (20\,\text{m}\,\text{s}^{-1})(0.5\,\text{s})$ $= 10\,\text{m}$

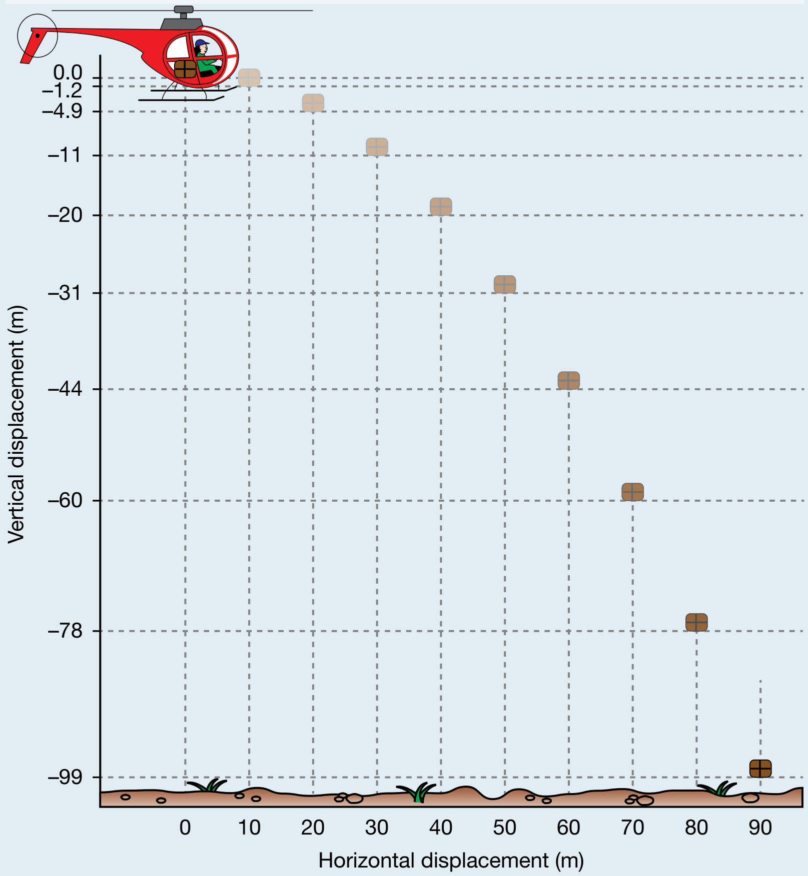
Repeat the calculations shown in Table 1.2 for $t = 1.0\,\text{s},\ 1.5\,\text{s},\ 2.0\,\text{s}$, etc. to gain the results shown in Table 1.3. The scale diagram of the package's position is shown in Figure 1.13.

TABLE 1.3 Vertical and horizontal displacement travelled over time.

Time (s)	Vertical displacement (m)	Horizontal displacement (m)
0.5	−1.2	10
1.0	−4.9	20
1.5	−11	30
2.0	−20	40
2.5	−31	50
3.0	−44	60
3.5	−60	70
4.0	−78	80
4.5	−99	90

(d) The velocity at which the package strikes the ground will be the vector sum of the vertical and horizontal components of the package's velocity (which occurs at $t = 4.5\,\text{s}$ as seen in part (a))

FIGURE 1.13 Scale diagram of the package's position.



Vertical component: $u_y = 0$; $t = 4.5\ \text{s}$; $a_y = g = -9.8\ \text{m s}^{-2}$

$$
\begin{aligned}
v_y &= u_y + a_y t \\
&= (0) + (-9.8\ \text{m s}^{-2})(4.5\ \text{s}) \\
&= -44.1\ \text{m s}^{-1}
\end{aligned}
$$

Horizontal component: $v_x = u_x = 20\ \text{m s}^{-1}$

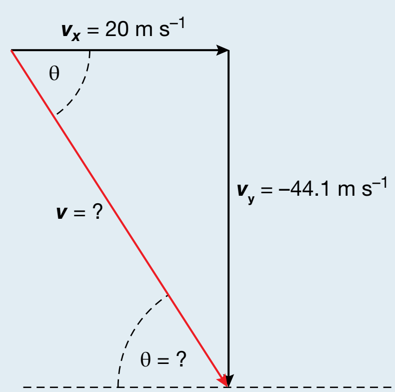
As can be seen in Figure 1.14:

$$
\begin{aligned}
v &= \sqrt{(v_x)^2 + (v_y)^2} \\
&= \sqrt{(20\ \text{m s}^{-1})^2 + (44.1\ \text{m s}^{-1})^2} \\
&= 48.4\ \text{m s}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\text{and } \theta &= \tan^{-1}\left|\frac{44.1}{20}\right| \\
&= 65.6^\circ\ (\text{rounded to } 66^\circ)
\end{aligned}
$$

Therefore, the package strikes the ground with a velocity of $48.4\ \text{m s}^{-1}$ at an angle of 66° to the horizontal surface.

FIGURE 1.14 The vertical and horizontal components of the velocity at which the package strikes the ground.



WORKING SCIENTIFICALLY 1.3

Different types of golf balls have different numbers of dimples on them, which are meant to provide longer flight times and so allow them to travel further. Investigate the mathematical relationship between the number and distribution of dimples on golf balls and their ranges. You will need to design a SAFE method of launching the golf balls so that they consistently have the same initial velocity and launch angle.

1.3 Exercise 1

1 Why don't target shooters aim directly at the bullseye in the middle of the target?

2 A ball is thrown directly upwards with a velocity of 45 km h^{-1}. Ignoring air resistance, determine:
- (a) its peak height
- (b) its time of flight
- (c) its velocity after 0.5 s
- (d) its velocity after 1.5 s.

3 A basketball player jumps directly upwards so that his centre of mass reaches a maximum displacement of 50 cm.
- (a) What is the velocity of the basketballer's centre of mass when it returns to its original height above the ground?
- (b) For how long was the basketballer's centre of mass above its original height?

4 What will be the horizontal and vertical velocity components of a bullet 0.8 s after it is fired horizontally with a speed of 900 m s^{-1}?

5 Determine the horizontal and vertical velocity components of a dart thrown horizontally with a speed of 5 m s^{-1} at a time 0.5 seconds after it leaves the thrower's hand.

6 Calculate the velocity of a horizontally thrown ball after 1.2 s if it has an initial velocity of 7 m s^{-1}.

7 An air gun is fired horizontally at a target 81 m away and the bullet takes just 0.35 s to strike it.
- (a) What was the initial velocity of the bullet?
- (b) How far vertically did the bullet drop from the horizontal by the time it had struck the target?
- (c) What is the range of the air gun if the target is removed and it is fired horizontally at the same velocity as in (a)?

8 How long will a bullet fired horizontally at a velocity of 800 m s^{-1} take to reach a target located 300 m away?

9 A pellet is launched horizontally from a slingshot at a bottle placed on a fence. The pellet strikes the fence 30 cm below the bottle. If the pellet had an initial speed of 50 m s^{-1}, what horizontal distance separated the shooter from the bottle?

10 A shot is thrown horizontally with a velocity of 7 m s^{-1} by a shotput competitor. If the shot left the athlete's hand at a height of 1.5 m above the ground, what was its range?

11 A ball is thrown horizontally at a speed of 40 m s^{-1} from the top of a cliff into the ocean below and takes 4.0 seconds to land in the water. Air resistance can be ignored.
- (a) What is the height of the cliff above sea level if the thrower's hand releases the ball from a height of 2.0 metres above the ground?
- (b) What horizontal distance did the ball cover?
- (c) Calculate the vertical component of the velocity at which the ball hits the water.
- (d) At what angle to the horizontal does the ball strike the water?

1.4 Projection at an angle

1.4.1 Resolving initial velocity components

Generally, projectiles are shot, thrown or driven at some angle to the horizontal.

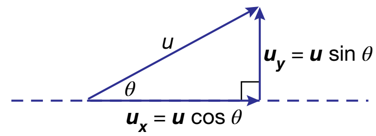
In these cases, the initial velocity may be resolved into its horizontal and vertical components to help simplify the analysis of the motion.

Given a projectile launched with an initial velocity $\boldsymbol{u}$ at an angle θ to the horizontal, the values of the initial horizontal velocity, $\boldsymbol{u}_x$, and the initial vertical velocity, $\boldsymbol{u}_y$, can be determined by trigonometry:

$$\boldsymbol{u}_x = \boldsymbol{u} \cos\theta$$

$$\boldsymbol{u}_y = \boldsymbol{u} \sin\theta$$

FIGURE 1.15 The initial velocity at some angle to the horizontal can be resolved into vertical and horizontal components, $\boldsymbol{u}_y$ and $\boldsymbol{u}_x$.



In this way, a bullet fired at an upward angle of 30° at a speed of 400 m s^{-1} will have an initial horizontal velocity of 346 m s^{-1}($\boldsymbol{u}_x = 400 \cos 30°$) and an initial vertical velocity of 200 m s^{-1}($\boldsymbol{u}_y = 400 \sin 30°$).

As with horizontally launched projectiles, the horizontal velocity component of the projectile launched at an angle will remain the same throughout its journey, that is:

$$\boldsymbol{u}_x = \boldsymbol{v}_x$$

The vertical velocity component, however, will vary in the same manner as any other object that is thrown straight up; it will slow down as it rises, come to rest momentarily at the top of its motion and then fall back towards the Earth again with increasing speed. The position and velocity of any projectile at any moment in time can, as usual, be determined by the equations of motion.

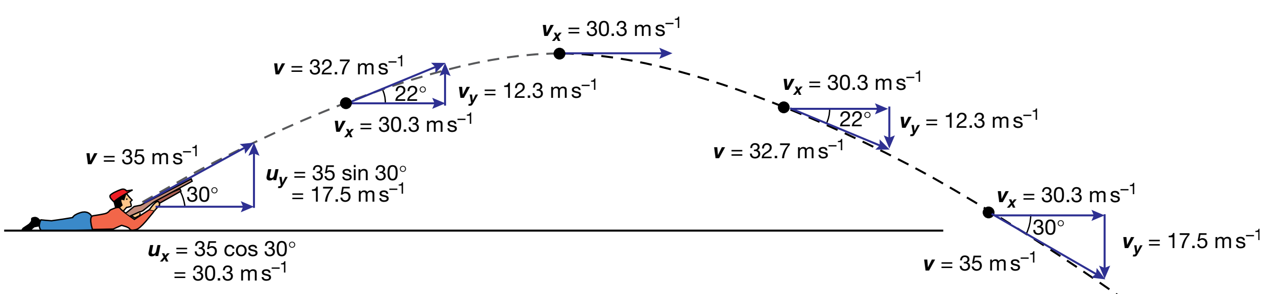
1.4.2 Symmetrical parabolic motion

For a projectile fired from ground level and returning to ground level at the end of its flight, the trajectory of the projectile is considered to be symmetrical, with the axis of symmetry being the vertical line passing through the position of maximum height.

For Figure 1.16, calculations have been performed for several points along the trajectory of a fired bullet to show how the velocity varies throughout the motion. You can see how the velocity reduces to a minimum at the peak, because at this point the vertical velocity is zero although the horizontal velocity remains. As the projectile falls from its peak, its velocity increases again until, at the end of the trajectory, it has the same value as the initial velocity and even the same angle to the horizontal, although now it is directed below the horizontal.

The bullet reaches its maximum height halfway through the total time of flight; also at this point, the bullet has travelled half of its range.

FIGURE 1.16 Velocity determined at many points along the trajectory of a bullet fired from an air gun.



In general, here are some tips for performing projectile motion calculations:

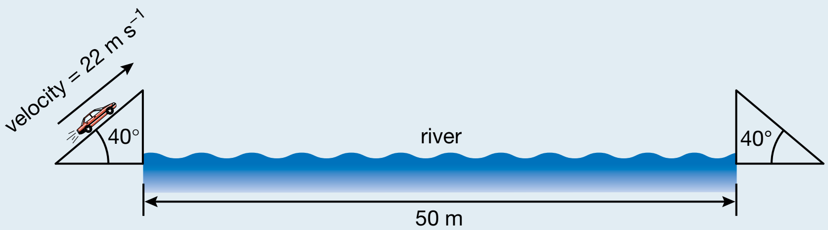
- It helps to draw a diagram.
- Always separate the motion into vertical and horizontal components.
- Remember to resolve the initial velocity into its components if necessary.

- The time in flight is the link between the separate vertical and horizontal components of the motion.
- At the end of any calculation, check to see if the quantities you have calculated are reasonable.

1.4 SAMPLE PROBLEM 1

A stunt driver is trying to drive a car over a small river. The car will travel up a ramp (at an angle of 40°) and leave the ramp travelling at 22 m s^{-1}. The river is 50 m wide. Will the car make it?

FIGURE 1.17



SOLUTION:

Assign up as positive and down as negative.

Before either part of the motion can be examined, it is important to calculate the vertical and horizontal components of the initial velocity.

Therefore, the initial vertical velocity is 14 m s^{-1} and the initial horizontal velocity is 17 m s^{-1}.

In order to calculate the range of the car (how far it will travel horizontally), it is clear that the horizontal part of its motion must be considered. However, the vertical part is also important. The vertical motion is used to calculate the time in the air. Then, the horizontal motion is used to calculate the range.

FIGURE 1.18 The vertical and horizontal components of the stunt car's initial velocity.

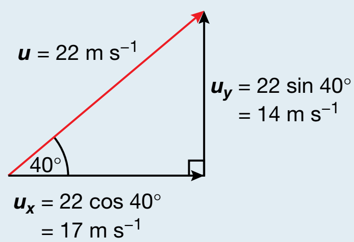


TABLE 1.4 Calculating the horizontal and vertical components

Vertical component	Horizontal component
(Use the first half of the motion — from take-off until the car has reached its highest point. This can be done because the trajectory is symmetrical.) $\boldsymbol{u_y} = 14\text{ m s}^{-1}$; $\boldsymbol{a_y} = \boldsymbol{g} = -9.8\text{ m s}^{-2}$; $\boldsymbol{v_y} = 0$ (as the car comes to a vertical halt at its highest point); t = ? $v_y = u_y + a_yt$ $0 = 14\text{ m s}^{-1} + (-9.8\text{ m s}^{-2})t$ $t = \frac{14}{9.8} = 1.4\text{ s}$ As this is only half the motion, the total time in the air is 2.8 s. (It is possible to double the time in this situation because we have ignored air resistance.The two parts of the motion are symmetrical.)	$\boldsymbol{u_x} = 17\text{ m s}^{-1}$; $t = 2.8\text{ s}$ (being twice the time taken to reach maximum height as calculated for the vertical component), $\boldsymbol{a_x} = 0$, $\boldsymbol{x}$ = ? $x = u_xt + \frac{1}{2}a_xt^2$ $= (17\text{ m s}^{-1})(2.8\text{ s}) + \frac{1}{2}(0)t^2$ $= 48\text{ m}$

Therefore, the unlucky stunt driver will fall short of the second ramp and will land in the river. Maybe the study of physics should be a prerequisite for all stunt drivers!

1.4.3 Asymmetric trajectories

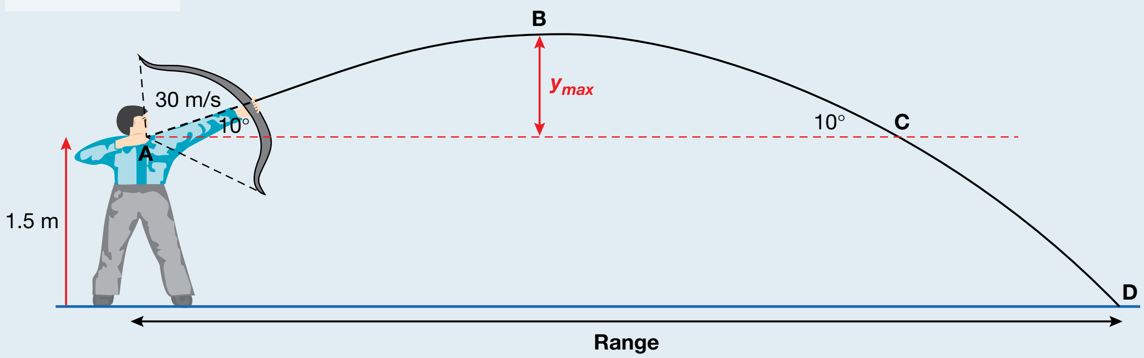
Let's now consider the case of a projectile that is launched at a point above the horizontal plane on which it will land.

1.4 SAMPLE PROBLEM 2

The archer in Figure 1.19 releases his arrow from a point 1.5 m above the ground at an angle of 10°. If the arrow has an initial velocity of $30\,m\,s^{-1}$, calculate:

(a) the maximum height of the arrow above the ground
(b) the range of the arrow
(c) the velocity with which the arrow strikes the ground.

FIGURE 1.19



SOLUTION:

(a) First, the horizontal and vertical components of the initial velocity need to be determined:

$$u_x = u \cos\theta = (30\,m\,s^{-1}) \cos 10° = 29\,m\,s^{-1}$$
$$u_y = u \sin\theta = (30\,m\,s^{-1}) \sin 10° = 5\,m\,s^{-1}$$

Now, it can be seen in the diagram that the arrow will reach its highest point at position B, located a height (y_{max} + 1.5 m) above the ground. At this position, the vertical component of the arrow's velocity will equal zero.

$u_y = 5\,m\,s^{-1};\ v_y = 0;\ a_y = g = -9.8\,m\,s^{-2};\ y_{max} = ?$

$$v_y^2 = u_y^2 + 2a_y y_{max}$$

$$(0)^2 = (5\,m\,s^{-1})^2 + 2(-9.8\,m\,s^{-2})y_{max}$$

$$\Rightarrow y_{max} = \frac{(5)^2}{19.6}$$
$$= 1.3\,m$$

$$h = y_{max} + 1.5\,m = 1.3\,m + 1.5\,m = 2.8\,m$$

The arrow reaches a maximum height of 2.8 m above the ground.

(b) In order to determine the range, the arrow's total time of flight must be determined. While the time taken to reach B from A is equal to the time taken to go from B to C (the symmetrical section of the trajectory), it is apparent that the flight time from B to D is much longer. As a result, two separate sets of flight time calculations must be made: one for the flight time from A to B(t_{AB}) and another set for the flight time from B to D (t_{BD}).

For section AB:

$\boldsymbol{u_y} = 5\,\text{m s}^{-1}; \boldsymbol{v_y} = 0; \boldsymbol{a_y} = \boldsymbol{g} = -9.8\,\text{m s}^{-2}; \boldsymbol{y_{max}} = 1.3\,\text{m}; t_{AB} = ?$

$$\boldsymbol{v_y} = \boldsymbol{u_y} + \boldsymbol{a_y} t_{\text{AB}}$$

$$(0) = \left(5\,\text{m s}^{-1}\right) + \left(-9.8\,\text{m s}^{-2}\right) t_{\text{AB}}$$

$$\Rightarrow t_{\text{AB}} = \frac{5}{9.8} = 0.51\,\text{s}$$

For section BD:

$\boldsymbol{u_y} = 0; \boldsymbol{a_y} = \boldsymbol{g} = -9.8\,\text{m s}^{-2}; \boldsymbol{y} = -2.8\,\text{m}; t_{\text{BD}} = ?$

$$\boldsymbol{y} = \boldsymbol{u_y} t_{\text{BD}} + \frac{1}{2}\boldsymbol{a_y} t_{\text{BD}}{}^2$$

$$(-2.8\ \ \text{m}) = (0)t_{BD} + \frac{1}{2}(-9.8\,\text{m s}^{-2})t_{\text{BD}}{}^2$$

$$\Rightarrow t_{\text{BD}} = \sqrt{\frac{-2.8}{-4.9}} = \pm 0.57\ \ \text{s}$$

In this instance, it is only the positive value that has meaning for us, so $t_{\text{BD}} = 0.57\,\text{s}$

The total time of flight from A to D, t_{AD} can then be found:

$t_{\text{AD}} = t_{\text{AB}} + t_{\text{BD}} = (0.51\,\text{s}) + (0.57\,\text{s}) = 1.08\,\text{s}$

The range $\boldsymbol{x}$ is then calculated:

$\boldsymbol{u_x} = 29\,\text{m s}^{-1}; t = 1.08\,\text{s}; \boldsymbol{x} = ?$

$$\boldsymbol{x} = \boldsymbol{u_x} t = \left(29\,\text{m s}^{-1}\right)(1.08\,\text{s}) = 31.3\,\text{m}$$

The range of the arrow is 31.3 m.

(c) Considering the flight of the arrow between B and D, in the vertical direction:

$\boldsymbol{u_y} = 0; \boldsymbol{y} = -2.8\,\text{m}; \boldsymbol{a_y} = \boldsymbol{g} = -9.8\,\text{m s}^{-2};\ t_{\text{BD}} = 0.57\,\text{s};\ \boldsymbol{v_y} = ?$

$$\boldsymbol{v_y} = \boldsymbol{u_y} + \boldsymbol{a_y} t_{\text{BD}}$$

$$\boldsymbol{v_y} = (0) + \left(-9.8\,\text{m s}^{-2}\right)(0.57)$$

$$\boldsymbol{v_y} = -5.6\,\text{m s}^{-1}$$

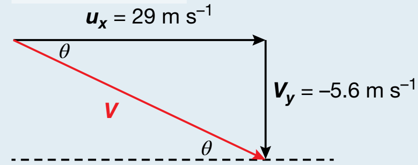
In the horizontal direction, the speed of the arrow remains the same throughout its flight, that is, $\boldsymbol{v_x} = \boldsymbol{u_x} = 29\,\text{m s}^{-1}$.

The velocity $\boldsymbol{v}$ with which the arrow strikes the ground at D will be the vector sum of the horizontal and vertical components:

$$v = \sqrt{(29\,\text{m s}^{-1})^2 + (-5.6\,\text{m s}^{-1})^2}$$
$$= 29.5\,\text{m s}^{-1}$$
$$\Rightarrow \theta = \tan^{-1}\left(\frac{5.6}{29}\right) = 11°$$

The arrow strikes the ground with a velocity of 29.5 m s^{-1} at an angle of 11° to the horizontal.

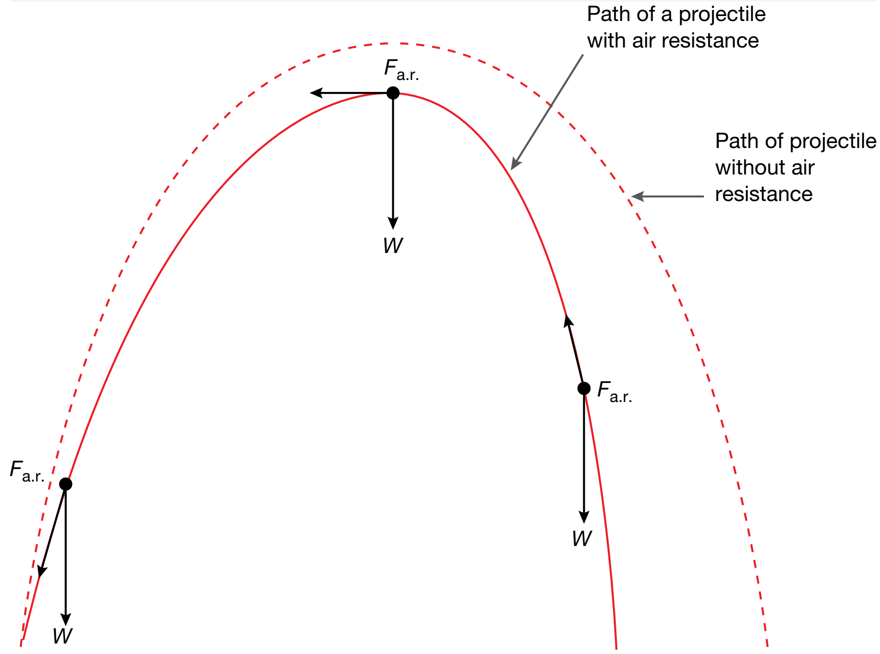
FIGURE 1.20



1.4.4 Projectile motion in the real world

In all of our work on projectile motion we have ignored the effect of air resistance on the motion of the projectile. The reason for this is that it is simply too difficult for us to account for, since it depends on many factors such as the shape, surface area and texture of the projectile, as well as its velocity through the air. In the real world, air resistance acts as a retarding force in both the vertical and horizontal directions. As a result, the path of the projectile is distorted away from a perfect parabola to the shape shown in Figure 1.21.

FIGURE 1.21 Air resistance opposes the velocity of a projectile at any given moment and distorts the trajectory away from a parabolic shape.



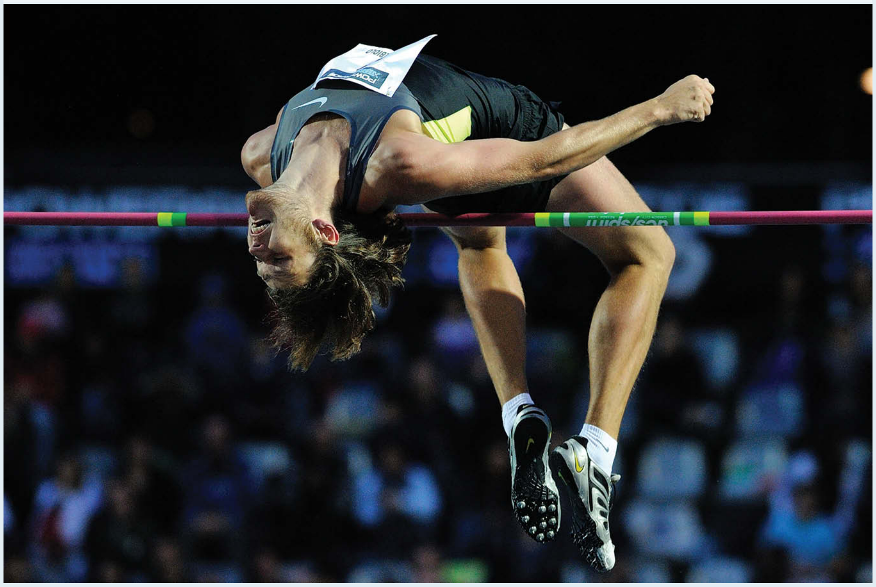
PHYSICS IN FOCUS

Hanging in mid air

Sometimes dancers, basketballers and high jumpers seem to hang in mid air. It is as though the force of gravity had temporarily stopped acting on them. Of course this is not so! It is only the person's centre of mass that moves in a parabolic path. The arrangement of the person's body can change the position of the centre of mass, causing the body to appear to be hanging in mid air even though the centre of mass is still following its original path.

High jumpers can use this effect to increase the height of their jumps. By bending his body as he passes over the bar, a high jumper can cause his centre of mass to be outside his body! This allows his body to pass over the bar, while his centre of mass passes under it. The amount of energy available to raise the high jumper's centre of mass is limited, so he can raise his centre of mass only by a certain amount. This technique allows him to clear a higher bar than other techniques for the same amount of energy.

FIGURE 1.22 A high jumper's centre of mass passes under the bar, while his body passes over the bar!



1.4 Exercise 1

1 What will be the range of a javelin that is thrown by an athlete at an angle of 20° with a velocity of 18 m s^{-1} if the point of the javelin was 1.8 m from the ground when it was launched?

2 A football is kicked when it is lying motionless on the ground, giving it a velocity of 22 m s^{-1} at an angle of 30°.
   (a) How long will the football remain in the air, assuming that it is not intercepted?
   (b) What will be the football's range?
   (c) How high above the ground will the football be 0.6 s after it is kicked?

3 The batter in a softball game hits the ball high into the air at an angle of 68°. If the ball has an initial velocity of 50 km h^{-1}, how far from the batter must the pitcher stand if they are to catch it on the full? Assume that the ball will be caught at the height at which the bat hit it.

4 A golfer wishes to send the ball onto the green 350 m away but knows that it will need to have a trajectory of at least 40° if it is to pass over a nearby section of rough. With what speed will the ball need to travel from the tee to reach the green?

5 A tennis ball is struck 15° above horizontal at a velocity of 25 m s^{-1}. If the ball was 1.2 m above the court surface when it was struck, calculate:
   (a) the maximum height reached by this ball
   (b) the range of the ball assuming that it is not hit by the player's opponent.

eBookplus RESOURCES

Complete this digital doc: Free throw shooter
Searchlight ID: doc-0006

Complete this digital doc: Modelling a stunt driver
Searchlight ID: doc-0007

Complete this digital doc: Investigation: Modelling projectile motion
Searchlight ID: doc-26315

Complete this digital doc: Investigation: Predicting the range of a projectile
Searchlight ID: doc-26316

Complete this digital doc: Investigation: The drop zone
Searchlight ID: doc-26317

1.5 Review

1.5.1 Summary

- In the absence of an atmosphere, all objects on the Earth would fall with the same acceleration regardless of size or mass.
- Objects fall at different rates due to drag (air resistance). The drag force exerted by the air on a falling object is proportional to the density of the air, the object's cross-sectional area perpendicular to its descent, and the square of its velocity.
- When the forces exerted by gravity and air resistance on a falling object are equal, it ceases to accelerate. For the rest of its descent it falls at a constant velocity called the terminal velocity.
- A projectile is any object that is launched into the air.
- There are two forces acting on a projectile in flight: gravity acting downwards and air resistance acting in the opposite direction to that of the motion. In modelling projectile motion, it is helpful to ignore the air resistance.
- To analyse the motion of a projectile, the equations of motion with constant acceleration can be applied to the horizontal and vertical components of the motion separately.
- The vertical motion of a projectile is uniformly accelerated motion and can be analysed using these equations:

$$\boldsymbol{v}_y = \boldsymbol{u}_y + \boldsymbol{a}_y t$$
$$\Delta \boldsymbol{y} = \boldsymbol{u}_y t + \frac{1}{2}\boldsymbol{a}_y t^2$$
$$\boldsymbol{v}_y^2 = \boldsymbol{u}_y^2 + 2\boldsymbol{a}_y \Delta \boldsymbol{y}.$$

- The horizontal motion of a projectile has constant velocity. As acceleration in the horizontal direction is zero, the motion can be analysed using these equations:

$$\boldsymbol{v}_x = \boldsymbol{u}_x$$
$$\boldsymbol{v}_x{}^2 = \boldsymbol{u}_x{}^2$$
$$\Delta \boldsymbol{x} = \boldsymbol{u}_x t.$$

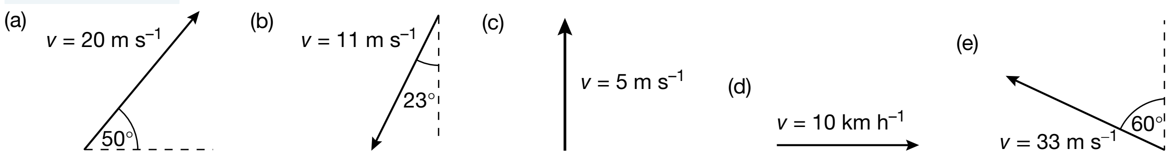
- An object projected horizontally near Earth's surface travels in a semi-parabolic path if air resistance is negligible.
- Objects that are projected upwards at an angle follow a parabolic path that is (ideally) symmetrical. The object reaches its maximum height halfway through its motion, and the vertical component of its velocity at this time is equal to zero.

1.5.2 Questions

(In the following problems, assume negligible air resistance and that $\boldsymbol{g} = -9.8\ \text{m s}^{-2}$ unless otherwise stated.)

1. Explain why it is that the vertical and horizontal components of a projectile's motion are independent of each other. Identify any common variables.
2. Describe the trajectory of a projectile.
3. List any assumptions we are making in our treatment of projectile motion.
4. Describe Galileo's contribution to our knowledge of projectile motion.
5. What is the mathematical significance of vertical and horizontal motions being perpendicular?
6. Describe the effect of air resistance on the trajectory of a projectile.
7. Explain why the horizontal component of velocity remains the same when a projectile's motion is modelled.
8. While many pieces of information relating to the vertical and horizontal parts of a particular projectile's motion are different, the time is always the same. Explain why this is so.
9. In each of the cases shown in Figure 1.23, calculate the magnitude of the vertical and horizontal components of the velocity.

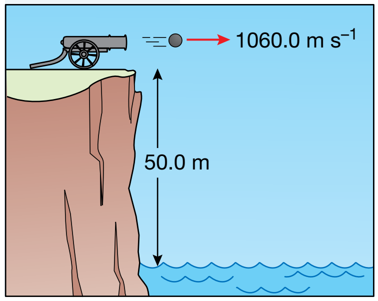
FIGURE 1.23



10. Peter hangs from the diving board by his fingers, with his feet 6 metres above the water. Ignoring air resistance, calculate:
    (a) the time between Peter letting go and his feet entering the water
    (b) his velocity when his feet enter the water
    (c) his distance above the water 0.1 seconds after letting go.
11. A student archer keeps aiming directly at the bullseye, which is 200 m away at the end of the target range, but sees that his arrows always seem to land 1 m below the bullseye. If he is shooting horizontally every time, with what velocity are his arrows leaving the bow?
12. A volleyball player sets the ball for a team mate. In doing so, she taps the ball up at $5.0\ \text{m s}^{-1}$ at an angle of 80.0° above the horizontal. If her fingers tapped the ball at a height of 1.9 m above the floor, calculate the maximum height to which the ball rises above the floor.
13. An 'extreme' cyclist wants to perform a stunt in which he rides up a ramp, launching himself into the air, then flies through a hoop and lands on another ramp. The angle of each ramp is 30.0° and the cyclist is able to reach the launch height of 1.50 m with a launching speed of $30.0\ \text{km h}^{-1}$. Calculate:
    (a) the maximum height above the ground that the lower edge of the hoop could be placed
    (b) how far away the landing ramp should be placed.
14. A football is kicked with a velocity of $35.0\ \text{m s}^{-1}$ at an angle of 60.0°. Calculate:
    (a) the 'hang time' of the ball (time in the air)
    (b) the length of the kick.
15. A basketball player stands 2.50 m from the ring. He faces the backboard, jumps up so that his hands are level with the ring and launches the ball at $5.00\ \text{m s}^{-1}$ at an angle of 50.0° above the horizontal. Calculate whether he will score.

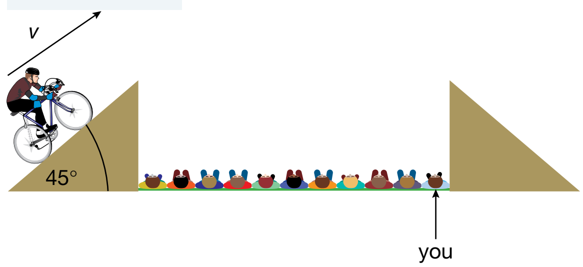
16. A coastal defence cannon fires a shell horizontally from the top of a 50.0 m high cliff, directed out to sea as shown in Figure 1.24, with a velocity of 1060.0 m s^{-1}. Calculate the range of the shell's trajectory.

FIGURE 1.24



17. A ball falls from the roof-top tennis court of an inner city building. This tennis court is 150 m above the street below. Assume the ball has no initial velocity.
    (a) How long would the ball take to hit the street?
    (b) What would the vertical velocity of the ball be just prior to hitting the ground?
18. After taking a catch, Ricky Ponting throws the cricket ball up into the air in jubilation.
    (a) The vertical velocity of the ball as it leaves his hands is 18 m s^{-1}. How long will the ball take to return to its original position?
    (b) What was the ball's maximum vertical displacement?
19. A friend wants to get into the *Guinness Book of Records* by jumping over 11 people on his push bike. He has set up two ramps as shown, and has allowed a space of 0.5 m for each person to lay down in. In practice attempts, he has averaged a speed of 7.0 m s^{-1} at the end of the ramp. Will you lay down as the eleventh person between the ramps? Support your response with appropriate calculations.

FIGURE 1.25



20. A gymnast wants to jump a distance of 2.5 m, leaving the ground at an angle of 28°. With what speed must the gymnast take off?
21. A horse rider wants to jump a 3.0 m wide stream. The horse can approach the stream with a speed of 7 m s^{-1}. At what angle must the horse take off? (Hint: You will need to use trigonometric identities from mathematics, or model the situation using a spreadsheet to solve this problem.)
22. Experienced target shooters say that they do not so much aim at the target as 'feel' for the right spot above the target. Michelle is an experienced target shooter who has scored a perfect bullseye on a target 300 m away. If the target was set at her eye level and the bullet left the rifle at a speed of 600 m s^{-1}, how high above the bullseye was she actually aiming her rifle? (Hint: The trigonometric identity $\sin\theta\cos\theta = \frac{1}{2}\sin 2\theta$ may come in handy.)
23. During a football game, a player kicks the ball from the ground so that it enters the air at an angle of 30° with a speed of 20 m s^{-1}. As it is coming down, a player from the opposing team running in at a constant speed of 6 m s^{-1} catches it when it is 1.2 m above the ground. If the opposing player was running in a straight line towards the kicker when the ball was launched, what was the distance between the two players at that time?

PRACTICAL INVESTIGATIONS

Investigation 1.1: Modelling projectile motion

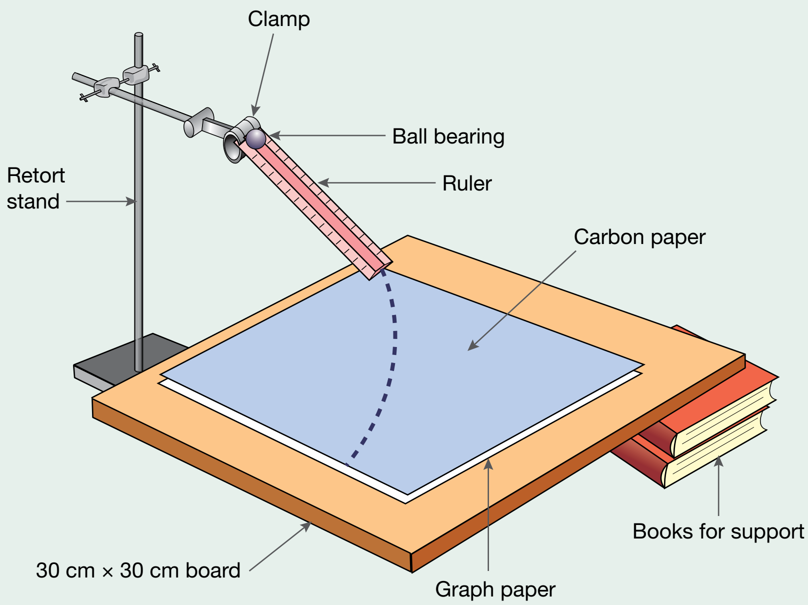
This investigation aims to model projectile motion by studying the motion of a ball bearing projected onto an inclined plane.

You will need the following equipment:

(a) 30 cm × 30 cm board
(b) retort stand and clamp
(c) carbon paper
(d) ball bearing
(e) graph paper
(f) 30 cm ruler (the ramp)

Set up the apparatus as shown in Figure 1.26.

FIGURE 1.26 The path of the projectile (ball bearing) is marked as it rolls down the ramp on the carbon paper.



Set up the inclined plane at an angle of approximately 20° and place the graph paper on it so that the ball will enter onto the inclined plane at a major division on the paper. Clamp the ruler so that the ball bearing rolling from it onto the inclined plane will be projected horizontally. Adjust the angle of the ruler so the path of the ball bearing will fit on the graph paper.

Having adjusted the apparatus, place a piece of carbon paper on the graph paper and record the motion of the ball bearing projected onto the inclined plane.

Remove the carbon paper and highlight the path for easier analysis.

We will assume that the horizontal velocity of the ball bearing's motion remained constant. Therefore, as the ball bearing takes equal times to travel horizontally between the major divisions on the graph paper, we can arbitrarily call one of these major divisions a unit of time.

Beginning at the point where the ball entered the graph paper, label these major divisions 0, 1, 2, 3... time intervals.

Record and tabulate the distance down the slope that the ball bearing travelled during each time interval.

Determine the average speed of the ball bearing down the slope during each time interval. Your answers should be in cm per time unit.

Plot a graph of average speed down the slope versus time and determine a value for the acceleration of the ball down the slope. Your answer will be in cm per $(\text{time unit})^2$.

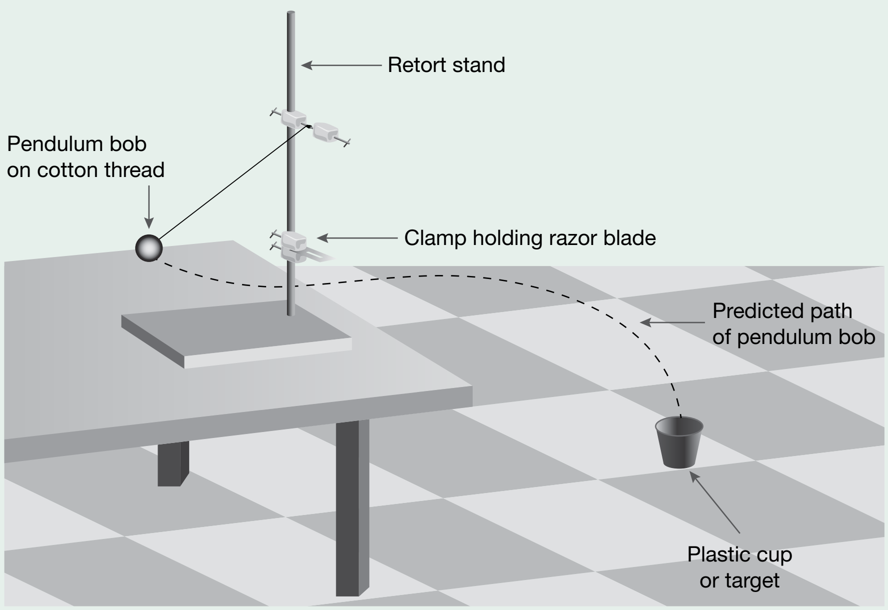
1. What do these graphs indicate about the motion of the ball down the plane?
2. What assumptions have been made in order to obtain these results?
3. How would the path of the ball bearing differ if:
   (a) the inclined plane was raised to a steeper angle while keeping the ramp as it was?
   (b) the angle of the ramp was raised and the inclined plane was kept as it was?
4. The ball moves faster across the bottom of the paper than across the top, which represents an increase in kinetic energy. What is the source of this extra energy? Try to find out why the rolling mass of the ball introduces a problem into this energy conversion.

Investigation 1.2: Predicting the range of a projectile

The aim of this investigation is to predict the range (that is, the horizontal distance travelled) of a projectile with a known initial horizontal velocity, and then to test the prediction.

The method you use to release the projectile with a known velocity will depend on the equipment available. One method is to use a pendulum bob as the projectile. Initially, the pendulum bob is attached to a retort stand with a fine thread, and is held at an angle before being released. The initial height is measured before the pendulum is released. As the pendulum reaches the bottom of its swing, the thread is cut by a carefully placed razor blade. The change in height of the pendulum bob can be used to determine the amount of gravitational potential energy that has been converted into kinetic energy of the pendulum bob. The kinetic energy, in conjunction with its mass, can be used to calculate the velocity of the pendulum at the moment the thread is cut. As the pendulum is moving horizontally at this point, this is the initial horizontal velocity of the pendulum bob. One possible set-up for this investigation is shown in Figure 1.27.

FIGURE 1.27 A possible arrangement of equipment for Investigation 1.2.



An alternative method involves rolling a marble down a ramp and again taking note of the energy changes to calculate the horizontal velocity at the bottom of the ramp. Both of these methods allow an estimate of the initial horizontal velocity of the projectile.

You may have access to technology (such as light gates or motion sensors) that allows you to measure the horizontal velocity of the projectile directly.

Once you know the initial horizontal velocity of the projectile and have measured the vertical distance it will fall, calculate the range of the projectile.

Place a target where you think the projectile will land. The target should be about 5 cm in diameter to take into account the estimated initial velocity, and the fact that air resistance and any spin of the projectile have been ignored in the calculations. Release your projectile and see how accurate your prediction was!

Investigation 1.3: The drop zone

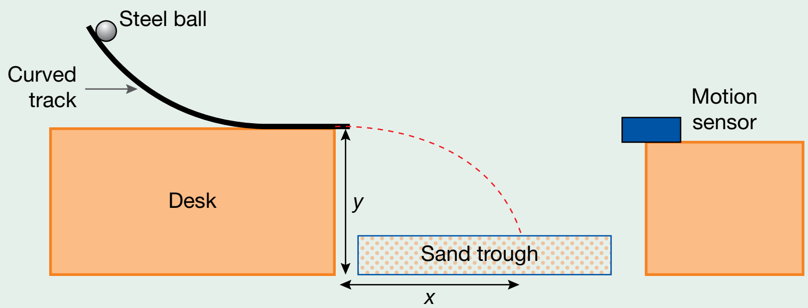
The aim of this investigation is to explore the relationship between the initial speed of a horizontally launched object and its range.

You will need the following equipment:

(a) steel ball bearing
(b) curved track (or flexible plastic racing car track)
(c) metre ruler
(d) long plastic trough or tray filled with sand
(e) data logger with motion sensor.

Set up the apparatus as shown in Figure 1.28. Ensure that the motion sensor is mounted directly opposite the launch end of the track.

FIGURE 1.28



Measure the height y between the launch end of the track and the floor.

Roll the ball down the track and note its landing position in the sand tray.

Measure the distance x between the launch end of the track and the landing position and read the launch speed ($\boldsymbol{u}_x$) from the data logger. Enter these values into a data table.

Adjust the height of the top end of the track to vary the value of $\boldsymbol{u}_x$ and repeat for 3 more launch speeds.

1. Given your measurement of the vertical distance y, and assuming that a = $-9.8\,\text{m}\,\text{s}^{-2}$ and $\boldsymbol{u}_y = 0$, determine the flight time t of the ball.
2. Using the flight time from the previous question and the values of $\boldsymbol{u}_x$, determine the theoretical range x that the ball should have reached in each of the 4 trials.
3. Compare these theoretical values to the experimental values that you obtained in each case.
4. What explanations can you give for any discrepancies between the values?
5. State the relationship between the initial speed of a horizontally launched projectile and its range.

TOPIC 2 Circular motion

2.1 Overview

2.1.1 Module 5: Advanced Mechanics

Circular motion

Inquiry question: Why do objects move in circles?

Students:

- conduct investigations to explain and evaluate, for objects executing uniform circular motion, the relationships that exist between:
  - centripetal force
  - mass
  - speed
  - radius
- analyse the forces acting on an object executing uniform circular motion in a variety of situations, for example:
  - cars moving around horizontal circular bends
  - a mass on a string
  - objects on banked tracks (ACSPH100)
- solve problems, model and make quantitative predictions about objects executing uniform circular motion in a variety of situations, using the following relationships:
  - $a_c = \dfrac{v^2}{r}$
  - $F_c = \dfrac{mv^2}{r}$
  - $\omega = \dfrac{\Delta\theta}{t}$
  - $v = \dfrac{2\pi r}{T}$
- investigate the relationship between the total energy and work done on an object executing uniform circular motion
- investigate the relationship between the rotation of mechanical systems and the applied torque $\tau = rF_{\perp} = rF\sin\theta$

FIGURE 2.1 By leaning into the curve, the motorcyclist is able to balance the forces of centripetal force, gravity and friction in such a way that he can safely negotiate it at a higher speed than if he remained upright.

2.2 Uniform circular motion

2.2.1 Moving in a circle

Humans seem to spend a lot of time going around in circles. Traffic at roundabouts, children on merry-go-rounds, cyclists in velodromes. If you stop to think about it, you are always going around in circles as a result of Earth's rotation. Objects that travel in a circle with a constant speed are described as undergoing **uniform circular motion**.

While performing uniform circular motion, objects will take the same length of time to make each revolution. The term **period** (T) is used to describe the time taken to travel a full circle. The distance covered during the period depends upon the radius of the circle of travel and is equal to the circle's circumference, that is:

$$d = 2\pi r$$

It is clear that the object's average speed as it moves will be equal to the distance around the circle's circumference divided by the period:

$$\text{average speed} = \frac{d}{T} = \frac{2\pi r}{T}$$

2.2 SAMPLE PROBLEM 1

A hammer thrower spins in a circle with the hammer at the end of a 1.2 m long chain. What is the hammer's average speed if it makes 4 revolutions in 5 seconds?

FIGURE 2.2 A hammer thrower spins in a circle with the hammer at the end of a long chain.

SOLUTION:

$$T = \frac{\text{total time for revolutions}}{\text{number of revolutions}} = \frac{5\text{ s}}{4} = 1.25\text{ s}$$

$$d = 2\pi r = 2\pi\,(1.2\text{ m}) = 7.5\text{ m}$$

$$\text{average speed} = \frac{7.5\text{ m}}{1.25\text{ s}} = 6\text{ m s}^{-1}$$

The hammer travels with an average speed of 6 m s^{-1}.

While the hammer in 2.2 Sample Problem 1 may be swung so that it travels with a constant speed, it can never have a constant velocity. As you will recall from your earlier studies, velocity is a vector quantity equal to an object's change in displacement over a time interval. The average velocity of an object at the end of time t depends upon the magnitude and direction of the object's displacement from its starting point (s) at that time:

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

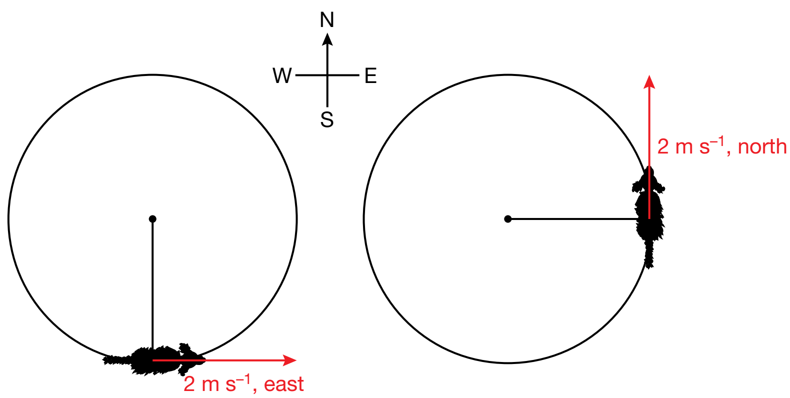
The athlete's hammer, a child on a carousel horse and an athlete running laps of a circular track are constantly returning to their starting positions and, as a result, are considered to have zero displacement. This means that their average velocities will also equal zero.

2.2.2 Instantaneous velocity

However, an object's instantaneous velocity is not zero. Consider the case of a dog that has been chained to a pole in a backyard and is running in circles around the pole at a constant speed at the chain's limit.

While the dog's average velocity for a single lap is zero, his instantaneous velocity at a single moment in time will have a magnitude equal to his speed and a direction according to the way he is facing. Provided that the dog maintains a constant speed, the magnitude of his velocity would not change, but the direction would be continually changing. If the dog is travelling east, his instantaneous velocity is in an easterly direction. A short time later, he will be travelling north, so his instantaneous velocity is in a northerly direction.

FIGURE 2.3 A dog travelling in a circle at a constant speed of 2 m s^{-1} has an ever-changing instantaneous velocity because his direction of motion changes with time



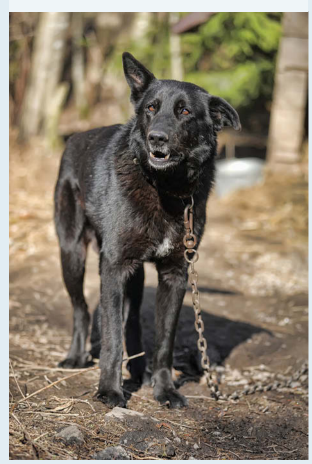
2.2 SAMPLE PROBLEM 2

Ralph has been a bad dog and has been chained up. To amuse himself, he runs in circles.

Ralph's chain is 7.0 m long and attached to a small post in the middle of the garden. It takes an average of 9 s to complete one lap.

(a) What is Ralph's average speed?
(b) What is Ralph's average velocity after three laps?
(c) What is Ralph's instantaneous velocity at point A? (Assume he travels at a constant rate around the circle.)

FIGURE 2.4



SOLUTION:

(a) $r = 7.0\,\text{m}$; $T = 9\,\text{s}$; $\boldsymbol{v}_{av} = ?$

$$\boldsymbol{v}_{av} = \frac{2\pi r}{T} = \frac{2\pi \times 7.0\,\text{m}}{9\,\text{s}} = 5\,\text{m s}^{-1}$$

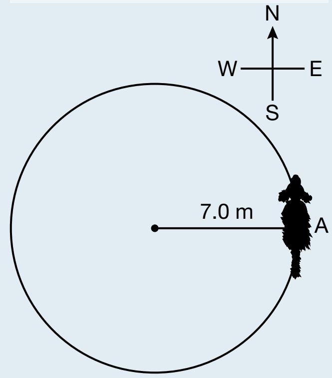
Ralph travels with an average speed of 5 m s^{-1}.

(b) After three laps, Ralph is in exactly the same place as he started, so his displacement is zero. No matter how long he took to run these laps, his average velocity would still be zero, as

$$v_{av} = \frac{s}{t} = \frac{0}{t} = 0$$

(c) Ralph's speed is a constant $5\ \mathrm{m\,s^{-1}}$ as he travels around the circle. At the instant in question, the magnitude of his instantaneous velocity is also $5\ \mathrm{m\,s^{-1}}$. This means Ralph's velocity is $5\ \mathrm{m\,s^{-1}}$ north.

FIGURE 2.5



2.2.3 Centripetal acceleration

Any object moving in a circle has a continually changing velocity as its direction its constantly changing.

As all objects with changing velocities are experiencing an acceleration, this means all objects that are moving in a circle are accelerating.

An acceleration can be caused only by an unbalanced force, so a non-zero net force is needed to move an object in a circle. The hammer thrower in our earlier example must apply a force to the hammer to keep it moving in a circle. When the hammer is released, it moves off with the velocity it had at the instant of release.

Let's consider the motion of the hammer at two different locations, A and B, as shown in Figure 2.7. The instantaneous velocity at each of these two points can be represented by the vectors $\boldsymbol{v_A}$ and $\boldsymbol{v_B}$. These vectors have the same magnitude but are acting in different directions and the direction of the velocity vector is always tangential to the circle in which it is moving.

As acceleration is equal to the change in velocity divided by the time taken for the velocity change, the acceleration of the hammer between points A and B can be found with the following equation:

$$\boldsymbol{a} = \frac{\Delta \boldsymbol{v}}{t} = \frac{\boldsymbol{v}_\mathbf{B} - \boldsymbol{v}_\mathbf{A}}{t}$$

As $\boldsymbol{v}_\mathbf{A}$ and $\boldsymbol{v}_\mathbf{B}$ are both vector quantities, we can find $\boldsymbol{v}_\mathbf{B} - \boldsymbol{v}_\mathbf{A}$ by adding the vector $\boldsymbol{v}_\mathbf{B}$ to the vector $-\boldsymbol{v}_\mathbf{A}$. Remember, $-\boldsymbol{v}_\mathbf{A}$ has the same magnitude as $\boldsymbol{v}_\mathbf{A}$ but acts in the opposite direction.

FIGURE 2.6 As long as the thrower keeps turning, the hammer moves in a circle. When the hammer is released, it moves in a straight line.

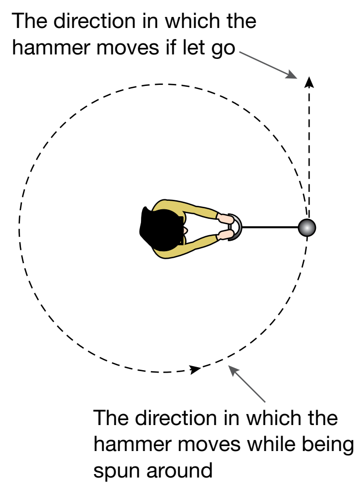


FIGURE 2.7 Velocity vectors for a hammer moving in an anticlockwise circle.

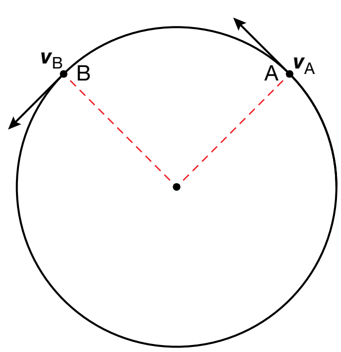
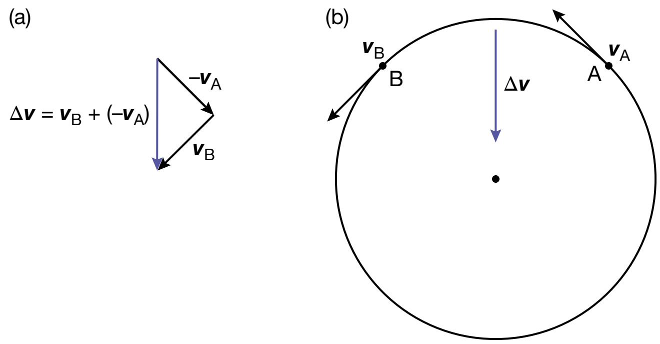


FIGURE 2.8 **(a)** Vector addition of $\boldsymbol{V}_{\mathbf{B}}+(-\boldsymbol{V}_{\mathbf{A}})$ **(b)** The change in velocity is towards the centre of the circle.



Notice that when the $\Delta \boldsymbol{v}$ vector is transferred back to the original circle halfway between the two points in time, it is pointing towards the centre of the circle. Such calculations become more accurate when very small time intervals are used; however, a large time interval has been used here to make the diagram clear.

No matter which time interval is chosen, the acceleration vector always points towards the centre of the circle. This centre-directed acceleration experienced by all objects moving with uniform circular motion is called the ***centripetal acceleration***, $\boldsymbol{a}_{\mathbf{c}}$ (the word 'centripetal' means 'centre-seeking' in Latin).

2.2.4 Calculating centripetal acceleration

Using vector geometry, it is possible to derive an expression for the magnitude of an object's centripetal acceleration.

In moving from point A to point B, the object moves through an angle θ as shown in Figure 2.9. Both A and B are separated from the centre of the circle O by a distance r equal to the circle's radius.

OAB is seen to form an isosceles triangle with OA = OB. The third side is formed by a line, or chord, joining A with B. When the angle θ is very small, the length of this chord is virtually the same as the length of the arc that also joins these two points. As this distance is covered by an object travelling with an average speed v in a time t, the length of AB can be calculated using $\boldsymbol{s} = v\, t$.

FIGURE 2.9 An object moving from A to B moves through an angle θ subtended at the centre of the circle.

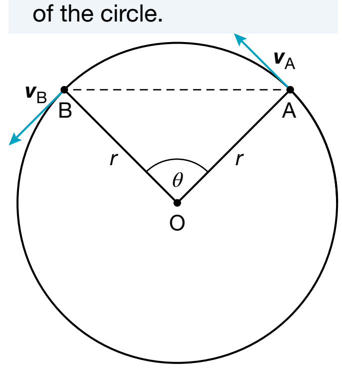
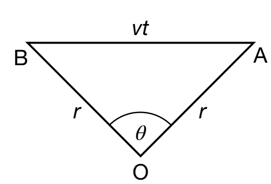


FIGURE 2.10 OAB is an isosceles triangle with OA = AB = r and AB = vt.



Turning our attention to the vector addition of $-\boldsymbol{v}_{\mathbf{A}}$ and $\boldsymbol{v}_{\mathbf{B}}$ shown in Figure 2.8a, we find that the angle made by the two vectors is equal to θ, the angular separation of A and B.

As the object is moving with uniform circular motion, the lengths of the vectors $\boldsymbol{v}_{\mathbf{B}}$ and $-\boldsymbol{v}_{\mathbf{A}}$ are identical (both equal to the object's average speed v) and so form two sides of an isosceles triangle as shown in Figure 2.12.

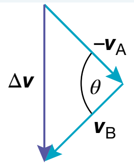


FIGURE 2.11 In the vector addition $\boldsymbol{\Delta v} = -\boldsymbol{v}_A + \boldsymbol{v}_B$, the angle θ is opposite the vector $\boldsymbol{\Delta v}$.

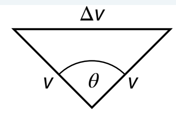


FIGURE 2.12 The vectors $\boldsymbol{v}_B$ and $-\boldsymbol{v}_A$ both have a length of v and, with Δv, form an isosceles triangle

As the triangles shown in both figures 2.10 and 2.12 are isosceles triangles with the same angle θ between their equal sides, we can describe them as being similar triangles — they can be thought of as the same triangle drawn on two different scales.

As the triangles are similar, the ratio of their sides must be constant, so:

$$\frac{\Delta \boldsymbol{v}}{\boldsymbol{v}t} = \frac{\boldsymbol{v}}{r}$$

Multiplying both sides by $\boldsymbol{v}$:

$$\frac{\Delta \boldsymbol{v}}{t} = \frac{\boldsymbol{v}^2}{r}$$

As $\boldsymbol{a} = \dfrac{\Delta \boldsymbol{v}}{t}$

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

Thus, we find that the magnitude of the centripetal acceleration $\boldsymbol{a}_c$ of an object moving with uniform circular motion is described by the equation:

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

where v is the object's average speed and r is the radius of the circle described by the object.

2.2 SAMPLE PROBLEM 3

What is the centripetal acceleration of a hammer that is swung in a 1.7 m radius circle if it makes 0.8 revolutions in a second?

SOLUTION:

First, the average speed of the hammer must be found:

$$T = \frac{1}{0.8\ \text{s}^{-1}} = 1.25\ \text{s};\ r = 1.7\ \text{m};\ v = ?$$

$$v = \frac{2\pi r}{T} = \frac{2\pi \times 1.7\ \text{m}}{1.25\ \text{s}} = 8.5\ \text{m s}^{-1}$$

Now, we may calculate the centripetal acceleration:

$$v = 8.5\ \text{m s}^{-1};\ r = 1.7\ \text{m};\ a_c = ?$$

$$a_c = \frac{(8.5\ \text{m s}^{-1})^2}{1.7\ \text{m}} = 43\ \text{m s}^{-2}$$

The hammer's centripetal acceleration is $43\ \text{m s}^{-2}$ directed towards the centre of the circle.

2.2.5 Centripetal force

According to Newton's Second Law of Motion, an object only accelerates if there is a net unbalanced force acting on it. This means that, for there to be a centripetal acceleration, there must be a net accelerating force acting, and an object that is travelling in a circle does so because there is a net force acting on it that constantly either pulls it or pushes it in towards the centre. This force is known as the **centripetal force**, $\boldsymbol{F_c}$.

We know that the relationship between acceleration and force is described by the equation:

$$\boldsymbol{F_{net}} = m\boldsymbol{a}$$

In the same way, we can describe the centripetal force in terms of mass and centripetal acceleration:

$$\boldsymbol{F_c} = m\boldsymbol{a_c}$$

By substitution of terms for $\boldsymbol{a_c}$, we may also describe the centripetal force in terms of the mass m, the radius r and average speed v:

$$\boldsymbol{F_c} = m\left(\frac{v^2}{r}\right)$$

The centripetal force will act in the same direction as the centripetal acceleration — that is, towards the centre of motion.

2.2 Exercise 1

1 What will be the average speed of a cyclist who takes 12 seconds to travel once around a circular track that has a radius of 50 m?

2 A discus thrower wishes the discus to leave her hand with a speed of 12 m s^{-1}. How fast will she need to rotate in order to do this if the distance from the centre of the discus to the centre axis of her body is 0.7 m?

3 A racing car travels around a circular track at a constant speed of 60 km h^{-1}. How long will it take the car to make one complete revolution if the track has a radius of 420 m?

4 A battery-operated toy car completes a single lap of a circular track in 15 s with an average speed of 1.3 m s^{-1}. Assume that the speed of the toy car is constant.
(a) What is the radius of the track?
(b) What is the magnitude of the toy car's instantaneous velocity halfway through the lap?
(c) What is the average velocity of the toy car after half of the lap has been completed?
(d) What is the average velocity of the toy car over the entire lap?

5 What is the centripetal acceleration acting on a car that is 'spinning a donut' (that is, travelling in a very tight continuous circle) if the radius of its path is 4 m and it is travelling at a speed of 25 km h^{-1}?

6 A tennis player swings his racquet in a circular arc when he hits a ball. What will be the acceleration of the racquet head if it is a distance of 1.1 m from the centre of rotation and it is travelling at a speed of 16 m s^{-1}?

7 A cyclist riding in a velodrome is travelling at 65 km h^{-1}. If he and his bike have a total mass of 70 kg and the velodrome is a circular shape with a diameter of 200 m, what is the centripetal force acting on the cyclist?

8 The centripetal force acting on a hammer while it is swung in a 1.5 m radius circle is 300 N. If the mass of the hammer is 7.3 kg:
(a) what is its average speed?
(b) how long does the hammer take to make one revolution?
(c) how many revolutions does the hammer make in 2 seconds?

9 A car is driven around a roundabout at a constant speed of 20 km h^{-1} (5.6 m s^{-1}). The roundabout has a radius of 3.5 m and the car has a mass of 1200 kg.
(a) What is the magnitude and direction of the acceleration of the car?
(b) What is the magnitude and direction of the force on the car?

10 Kwong (mass 60 kg) rides the Gravitron at the amusement park. This ride moves Kwong in a circle of radius 3.5 m, at a rate of one rotation every 2.5 s.
(a) What is Kwong's acceleration?
(b) What is the net force acting on Kwong? (Include a magnitude and a direction in your answer.)
(c) Draw a labelled diagram showing all the forces acting on Kwong.

2.3 Providing centripetal force

2.3.1 Tension

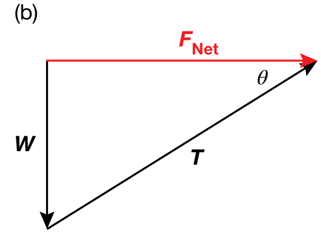
We have learned that a centripetal force is one that causes an object to follow a circular path rather than continue travelling in a straight line. This happens because there is a force acting on the object that either pulls it or pushes it in the direction of the circle's centre. In many cases, this force is **tension** — a pulling force acting through a medium such as a rope or chain. For example, when you suspend a large mass from the end of a string, the downward pull of gravity on the mass causes the bonds between the particles in the string threads to be stretched further apart. The tension in the string pulls upwards on the mass so that the net force acting on it is zero.

Consider a child being swung in a circle by an adult as shown in Figure 2.13a. The muscles and tendons (and, to a certain extent, bones and ligaments) in the arms and shoulders of both the adult and the child are being stretched; the resulting tension force in these fibres pulls the child towards the adult in a line directed along their arms.

FIGURE 2.13 **(a)**The combined effect of the tension force and the child's weight provides an inward net force that keeps the child travelling in a circle around the adult. **(b)** The net force providing the centripetal force is the vector sum of ***T*** and ***W***.

(a)

(b)



The forces of tension and gravity provide a non-zero net force:

$$\boldsymbol{F}_{net} = \boldsymbol{T} + \boldsymbol{W}$$

Resolving the vertical components of these forces:

$$\boldsymbol{F}_{net} \sin 0° = \boldsymbol{T} \sin \theta + \boldsymbol{W} \sin 270°$$

$$\Rightarrow \boldsymbol{W} = \boldsymbol{T} \sin \theta$$

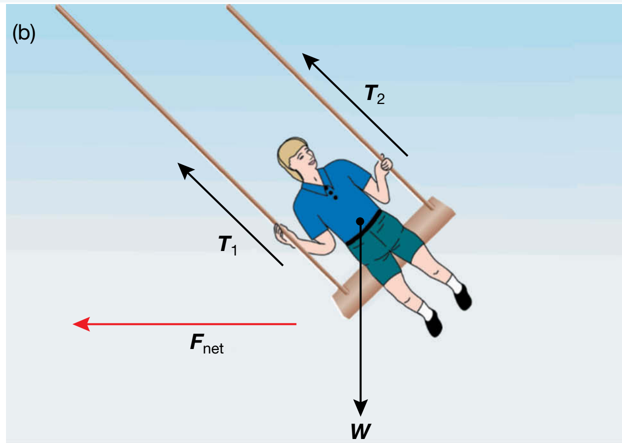
Resolving the horizontal components:

$$\boldsymbol{F}_{net} \cos 0° = \boldsymbol{T} \cos \theta + \boldsymbol{W} \cos 270°$$

$$\Rightarrow \boldsymbol{F}_{net} = \boldsymbol{T} \cos \theta$$

It can be seen that $\boldsymbol{F}_{net}$ is directed horizontally towards the centre of the circle in which the child moves and, so, provides the centripetal force.

FIGURE 2.14 **(a)** Tension contributes to the net force in many amusement park rides. **(b)** The net force acting on a rider.



2.3 SAMPLE PROBLEM 1

A conical pendulum is made of a 10 g mass suspended from a string of length 20 cm. The mass makes a flat horizontal circle of radius 5 cm as it moves. Determine:

(a) the angle θ between the string and the vertical
(b) the tension in the string
(c) the centripetal force acting on the pendulum mass
(d) the average speed of the pendulum mass.

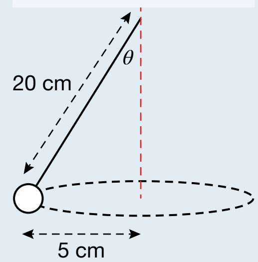
SOLUTION:

(a) Drawing a diagram of the pendulum as shown in Figure 2.15 reveals that

$$\theta = \sin^{-1}\frac{5\,\text{cm}}{20\,\text{cm}} = 14°$$

The string makes an angle of 14° with the vertical as the pendulum mass rotates.

FIGURE 2.15



(b) First, the weight of the pendulum is determined:

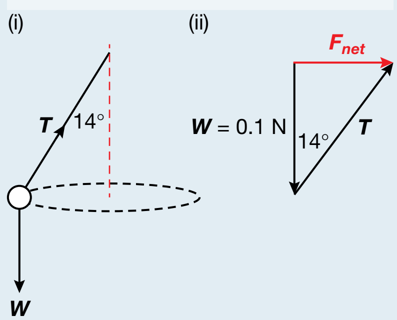
$\boldsymbol{W} = m\boldsymbol{g} = (0.010\,\text{kg})\left(9.8\,\text{m}\,\text{s}^{-2}\right) = 0.1\,\text{N}$, downwards

As the pendulum mass moves in a flat horizontal circle, it can be surmised that the net force acting on it must be acting horizontally towards the centre of that circle (providing the centripetal force).

Also, as the weight and the string tension are the only forces acting on the pendulum,

$\boldsymbol{F}_{net} = \boldsymbol{W} + \boldsymbol{T}$

FIGURE 2.16



Considering the vertical components of the vectors in Figure 2.16 (ii):

$$\boldsymbol{F}_{\text{net}} \sin 0° = (0.1\text{ N}) \sin 270° + T \cos 14°$$

$$0 = -0.1\text{ N} + T \cos 14°$$

$$\Rightarrow T = \frac{0.1\text{ N}}{\cos 14°} = 0.1\text{ N, rounded to one decimal place.}$$

The tension in the string is 0.1 N.

(c) The centripetal force is equal to $\boldsymbol{F}_{\text{net}}$.
Considering the horizontal components of the vectors in Figure 2.16 (ii):

$$\boldsymbol{F}_{\text{net}} \cos 0° = (0.1\text{ N}) \cos 270° + (0.1\text{ N}) \sin 14°$$

$$\boldsymbol{F}_{\text{net}} = (0.1\text{ N}) \sin 14° = 0.02\text{ N}$$

As $\boldsymbol{F}_c = \boldsymbol{F}_{\text{net}}$, the centripetal force acting on the pendulum mass is 0.02 N acting towards the centre of the circle.

(d) $\boldsymbol{F}_c = 0.02$ N inwards; m = 0.01 kg; $\boldsymbol{r} = 0.05$ min inwards; $v = ?$

$$\boldsymbol{F}_c = \frac{mv^2}{r}$$

$$\Rightarrow 0.02\text{ N} = \frac{0.01\text{ kg} \times \boldsymbol{v}^2}{0.05\text{ m}}$$

$$\Rightarrow \boldsymbol{v} = \sqrt{0.1} = 0.3\text{ m s}^{-1} \text{ (Note that only the positive root is used)}$$

The pendulum mass has an average speed of 0.3 m s^{-1}

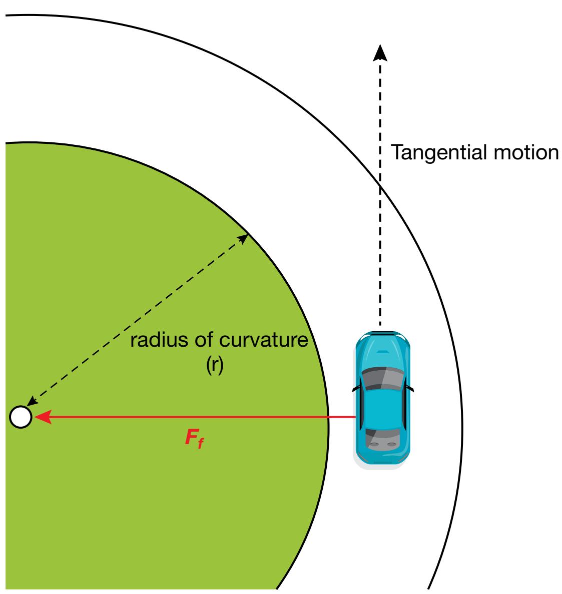
2.3.2 Friction

While it is fairly easy to understand how the tension force in a string or a chain pulls an object into a circular path, it is a little less obvious what's happening when you watch a runner moving around a circular track or a car cornering at speed. In these cases, it is friction that provides the centripetal force.

First, let's look at the frictional forces that act on a car travelling through a curve on a flat road. As the car travels around the curve, inertia (the tendency of an object to maintain its state of motion) dictates that the car moves outwards from the centre of curvature and instead travels in a straight line tangential to the curve. This is similar to the way that a weight swung in a flat circle on the end of a string will continue to move in a straight line tangential to the circle if the string were to snap.

Sideways (or 'lateral') friction between the tyres and the road opposes this outward motion, acting inwards towards the centre of the curve. As this sideways friction makes up the whole of the magnitude of the net force on the vehicle, it provides the centripetal force that keeps the car moving around the curve.

FIGURE 2.17 The lateral frictional forces acting on the car provide the centripetal force required to keep the car travelling in a circle rather than continuing to travel tangentially.



If the coefficient of friction between the road and the tyres is low — such as when the road is icy or wet — the sideways frictional force and, thus the centripetal force, will also be low.

As $\boldsymbol{F}_c = \frac{mv^2}{r}$, a lower centripetal force dictates that the car must travel at a lower speed if it is to negotiate the curve safely rather than sliding outwards.

2.3 SAMPLE PROBLEM 2

A 1200 kg car approaches a tight corner on a wet day. If the corner has a radius of curvature of 10 m and the coefficient of friction for the tyres on a wet road is 0.4, what is the maximum speed at which the car can safely turn the corner?

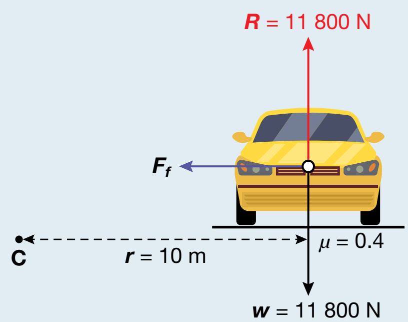
SOLUTION:

$m = 1200\text{ kg}, r = 10\text{ m}, \mu = 0.4, \boldsymbol{F}_c = ?$

$\boldsymbol{W} = m\boldsymbol{g} = (1200\text{ kg})(9.8\text{ m s}^{-2}) = 11\,800\text{ N downwards}$

$\boldsymbol{R} = -\boldsymbol{W} = 11\,800\text{ N upwards}$

FIGURE 2.18



$$\begin{aligned}\boldsymbol{F}_f &= \mu\boldsymbol{R}\\ &= (0.4)(11\,800\text{ N})\\ &= 4720\text{ N inwards towards the centre of the corner}\end{aligned}$$

The centripetal force must not exceed the frictional force if the car is to corner safely, therefore:

$$\boldsymbol{F}_f \geq \boldsymbol{F}_c$$

$$\boldsymbol{F}_f \geq \frac{mv^2}{r}$$

$$(4720\text{ N}) \geq \frac{1200\text{ kg} \times v^2}{10\text{ m}}$$

$$\Rightarrow v \leq \sqrt{\frac{4720\text{ N} \times 10\text{ m}}{1200\text{ kg}}}$$

$$v \leq 6.3\text{ m s}^{-1}$$

Therefore, the car should negotiate the corner at a speed no greater than 6.3 m s^{-1} (or 23 km h^{-1}).

Track athletes, cyclists and motorcyclists also rely on sideways frictional forces to enable them to move around corners. To increase the size of the lateral frictional force, which will therefore allow them to corner more quickly, they often lean into the corner. The lean also means that they are pushing on the surface at an angle, so the reaction force $\boldsymbol{R}$ is no longer normal to the ground and has a component towards the centre of their circular motion.

If we take θ to be the angle that the cyclist makes with the road surface (the 'angle of lean'), the components of the reaction force parallel and perpendicular to the road surface are:

$\boldsymbol{R}_{\parallel} = \boldsymbol{R} \sin \theta$

$\boldsymbol{R}_{\perp} = \boldsymbol{R} \cos \theta$

The net force acting horizontally towards the centre of motion (which provides the centripetal force) is such that:

$\boldsymbol{F}_{\mathbf{net}} = \boldsymbol{F}_{\mathbf{f}} + \boldsymbol{R}_{\parallel}$

Referring to Figure 2.19, we see that

$\boldsymbol{R}_{\perp} = \boldsymbol{W}$

$\boldsymbol{F}_{\mathbf{f}} = \mu \boldsymbol{R}_{\perp}$

$\Rightarrow \boldsymbol{F}_{\mathbf{f}} = \mu \boldsymbol{W}$

As $\boldsymbol{R}_{\perp} = \boldsymbol{W}$ and $\boldsymbol{R}_{\perp} = \boldsymbol{R} \sin \boldsymbol{R}_{\perp} = \boldsymbol{R} \cos \theta$, it follows that

$\boldsymbol{W} = \boldsymbol{R} \cos \theta$

Rearranging,

$$\boldsymbol{R} = \frac{\boldsymbol{W}}{\cos \theta}$$

This allows us to find an expression for $\boldsymbol{R}_{\parallel}$ in terms of $\boldsymbol{W}$:

$\boldsymbol{R}_{\parallel} = \boldsymbol{R} \sin \theta$

$$\boldsymbol{R}_{\parallel} = \left(\frac{\boldsymbol{W}}{\cos \theta}\right) \sin \theta$$

$\Rightarrow \boldsymbol{R}_{\parallel} = \boldsymbol{W} \tan \theta$

Thus,

$\boldsymbol{F}_{\mathbf{net}} = \mu \boldsymbol{W} + \boldsymbol{W} \tan \theta$

The greater the angle the cyclist leans from the perpendicular, the greater the magnitude of the net force acting to move him and his bike around the corner and so the faster the speed at which he is able to safely round the corner.

FIGURE 2.19 Leaning into a corner increases the size of the net force, allowing a higher speed while cornering. The sideways friction is greater, and the reaction force of the ground has a component towards the centre of the circular motion.

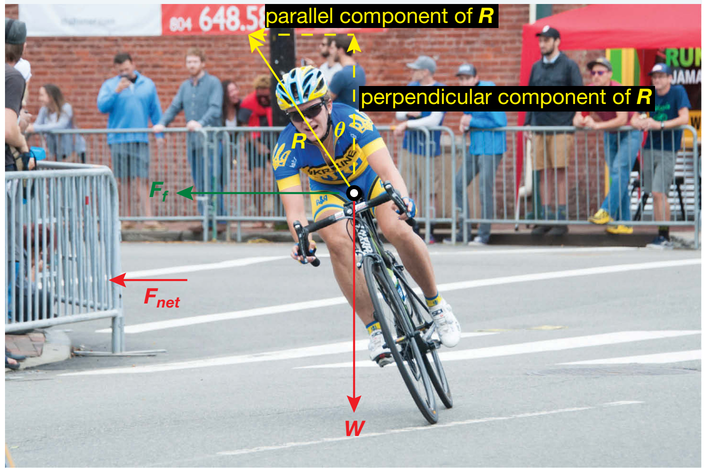


FIGURE 2.20 The extremely small angle that the motorcyclists make with the track allows them to corner at much higher speeds.

2.3.3 Banking

Leaning into the curve helps cyclists to maintain higher speeds when cornering because, by doing so, they enable part of the reaction force to contribute to the net force acting on them. This increases the magnitude of the net force and, therefore, the magnitude of the centripetal force keeping them in the curve.

Another way in which this can be achieved is to tilt the surface on which the cyclist rides.

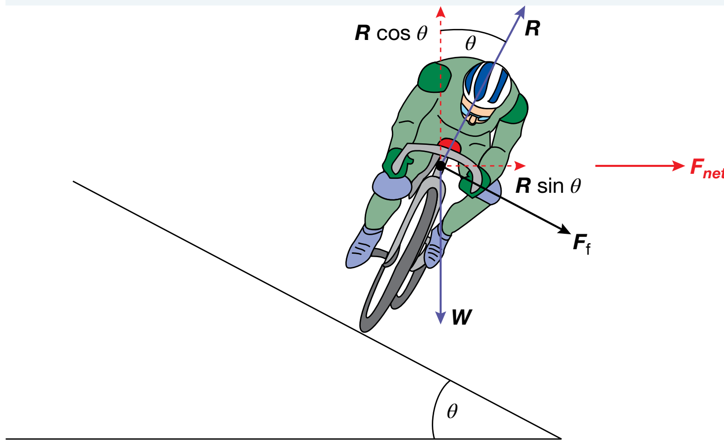
Velodromes — the indoor arenas in which Olympic cyclists compete — are constructed so that the track slants down from the outside edge to the inside edge. This angling of the surface is called **banking** and maximises the speeds at which cyclists are able to travel in sprint races.

Figure 2.22 shows that the net force acting horizontally towards the centre of the curve is the sum of the horizontal components of the normal reaction force $\boldsymbol{R}$ and the frictional force $\boldsymbol{F}_f$:

$$\boldsymbol{F}_{net} = \boldsymbol{R}\sin\theta + \boldsymbol{F}_f\cos\theta$$

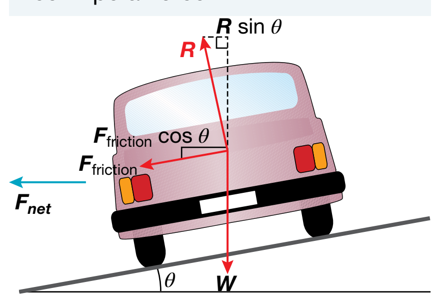
FIGURE 2.21 Cyclists racing in the velodrome at the Rio Olympics in 2016. The banking of the track allows the cyclists to safely reach much higher speeds than if they were racing on a flat surface.

FIGURE 2.22 The forces acting on a cyclist riding on a banked track.



Banking is also applied to curved roads allowing cars to negotiate them safely even under wet conditions. While the banking in the tight curves of a velodrome can be as steep as 45°, road banking by comparison is only a fraction of this.

FIGURE 2.23 Banking the road allows a component of the normal reaction to contribute to the centripetal force.



2.3 SAMPLE PROBLEM 3

A car of mass 1280 kg travels around a bend with a radius of 12.0 m. The total sideways friction on the wheels is 16 400 N. Calculate the maximum constant speed at which the car can be driven around the bend without skidding off the road if

(a) the road is not banked
(b) the road is banked at an angle of 10°

SOLUTION:

(a) The car will maintain the circular motion around the bend if:

$$\boldsymbol{F}_{net} = \frac{mv^2}{r}$$

where v = maximum speed.

If v were to exceed this speed, $F_{\text{net}} < \frac{mv^2}{r}$, the circular motion could not be maintained and the vehicle would skid.

$$\boldsymbol{F}_{net} = \boldsymbol{F}_f = 16\,400\text{ N} = \frac{1280\text{ kg} \times v^2}{12.0\text{ m}}$$
$$\Rightarrow v^2 = \frac{16\,400\text{ N} \times 12.0}{1280\text{ kg}}$$
$$v = 12.4\text{ m s}^{-1}$$

The maximum constant speed at which the vehicle can be driven around the unbanked bend is 12.4 m s^{-1}.

(b) Again, in order to maintain circular motion, $F_{net} > \frac{mv^2}{r}$

$\boldsymbol{F}_{net} = \boldsymbol{R}\sin\theta + \boldsymbol{F}_f\cos\theta$

As $\boldsymbol{R}\cos\theta = \boldsymbol{W} + \boldsymbol{F}_f\sin\theta$ (by considering the vertically acting components)

$$\boldsymbol{R}\cos 10° = \left(1280\text{ kg} \times 9.8\text{ m s}^{-2}\right) + (16\,400\text{ N})\sin 10°$$
$$\Rightarrow \boldsymbol{R} = 15\,630\text{ N}$$
$$\boldsymbol{F}_{net} = (15\,630\text{ N})\sin 10° + (16\,400\text{ N})\cos 10° = 18\,900\text{ N}$$

$$18\,900\text{ N} > \frac{1280\text{ kg} \times v^2}{12.0\text{ m}}$$
$$\Rightarrow v < \sqrt{\frac{18\,900\text{ N} \times 12.0\text{ m}}{1280\text{ kg}}}$$
$$v < 13.3\text{ m s}^{-1}$$

The maximum constant speed at which the vehicle can be driven around the unbanked bend is 13.3 m s^{-1}.

2.3.4 Riding inside circular motion

What happens to people and objects inside larger objects that are travelling in circles? The answer to this question depends on several factors.

Let's think about passengers inside a bus. The sideways frictional forces exerted by the road on the bus tyres act towards the centre of the circle, which increases the net force on the bus and keeps the bus moving around the circle. If the passengers are also to move in a circle (therefore keeping the same position in the bus) they, too, need to have a net force towards the centre of the circle. Without such a force, they would continue to move in a straight line and probably hit the side of the bus! Usually the friction between the seat and a passenger's legs is sufficient to prevent this happening.

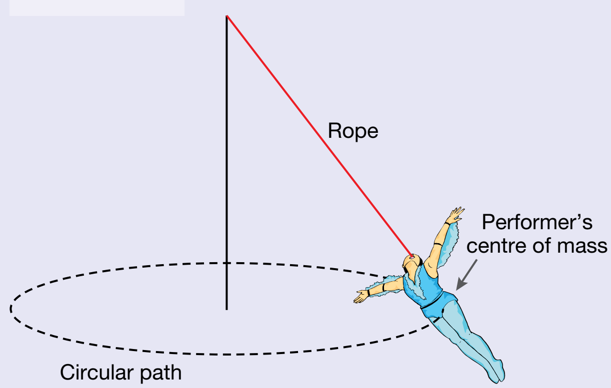
However, if the bus is moving quickly, friction alone may not be adequate.

In such cases, passengers may grab hold of the seat in front, thus adding a force of tension through their arms. Hopefully, the sum of the frictional force of the seat on a passenger's legs and the horizontal component of tensile force through the passenger's arms will provide a large enough centripetal force to keep that person moving in the same circle as the bus!

2.3 Exercise 1

1. Explain why an object moving in a circle with constant speed is accelerating without changing its speed.
2. Many runners who regularly do 400 m or more on a circular track start to experience damage to tendons and muscles on the inside of their legs, even though the large leg muscles used for running tend to be on the top or back of the leg. Runners who do 100 m races (which tend to be straight courses) do not. Explain why this is the case.
3. What angle will be formed by the chain of a chair-o-plane seat that is occupied by a 60 kg student if the tension in the chain is 680 N?
4. A motorcyclist travelling at 35 m s^{-1} around a circular track with a radius of 120 m leans in towards the centre of the circle. What will be the angle between the motorcyclist and the perpendicular?
5. A cyclist leans in towards the centre of a 150 m circular track so that there is an angle of 12° between her body and the perpendicular. What is her speed?
6. The coefficient of friction between the tyres of a rally car and the road surface is 0.58. What will be the maximum safe speed at which the car can turn a corner that has a 20 m radius of curvature?
7. The banking angle in a velodrome is 42°. What is the maximum speed that a cyclist can travel at for the banking to provide all of the centripetal force for the cyclist if he rounds a curve of radius 20 m? (In this case, friction makes no contribution to the net force.)
8. During a ballet performance, a dancer performs a full running circle. To do this, she leans in towards the centre of the circle. If her speed is 4 m s^{-1} and the circle has a radius of 5 m, what angle does her body make with the vertical?
9. When travelling around a roundabout, John notices that the fluffy dice suspended from his rear-vision mirror swing out. If John is travelling at 8.0 m s^{-1} and the roundabout has a radius of 5.0 m, what angle will the string connected to the fluffy dice (mass 100 g) make with the vertical?
10. A 50 kg circus performer grips a vertical rope with her teeth and sets herself moving in a circle with a radius of 5.0 m at a constant horizontal speed of 3.0 m s^{-1}
    (a) What angle does the rope make with the vertical?
    (b) What is the magnitude of the tension in the rope?

FIGURE 2.24



2.4 Non-uniform circular motion

2.4.1 In the half-pipe

So far, we have considered only what happens when the circular motion is carried out at a constant speed. However, in many situations this is not the case. When the circle is vertical, the effects of gravity can cause the object to go slower at the top of the circle than at the bottom. Such situations can be examined either by analysing the energy transformations that take place or by applying Newton's laws of motion.

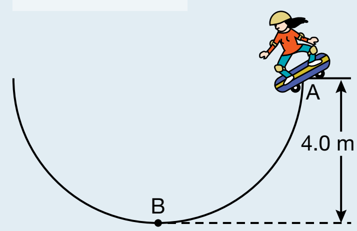
When a skateboarder enters a half-pipe from the top, that person has a certain amount of potential energy, but a velocity close to zero. At the bottom of the half-pipe, some of the gravitational potential energy of the skateboarder has been transformed into kinetic energy. As long as the person's change in height is known, it is possible to calculate the speed at that point.

2.4 SAMPLE PROBLEM 1

A skateboarder (mass 60 kg) enters the half-pipe at point A, as shown in Figure 2.25. (Assume the frictional forces are negligible.)

(a) What is the skateboarder's speed at point B?
(b) What is the net force on the skateboarder at B?
(c) What is the normal reaction force on the skateboarder at B?

FIGURE 2.25



SOLUTION:

(a) At point A, the skateboarder has potential energy, but no kinetic energy. At point B, all the potential energy has been converted to kinetic energy. Once the kinetic energy is known, it is easy to calculate the velocity of the skateboarder.

$m = 60\text{ kg}, h_A = 4.0\text{ m}, h_B = 0, g = 9.8\text{ m s}^{-2}$

decrease of potential energy from A to B = increase of kinetic energy from A to B

$$-(U_B - U_A) = E_{k\,B} - E_{k\,A}$$

$$-(\boldsymbol{mg}h_B - \boldsymbol{mg}h_A) = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2$$

$$-mg\ (h_B - h_A) = \frac{1}{2}m\ \left(v_B^2 - v_A^2\right)$$

Cancelling m from both sides:

$$-g\ (h_B - h_A) = \frac{1}{2}\left(v_B^2 - v_A^2\right)$$

$$-9.8\ (0\text{ m} - 4.0\text{ m}) = \frac{1}{2}\left(v_B^2 - 0\right)$$

$$v_B^2 = 78.4\text{ m}^2\text{s}^{-2}$$

$v_B = 8.854\text{ m s}^{-1}$, rounded to 8.9 m s^{-1}.
The skateboarder's speed at B is 8.9 m s^{-1}.

(b) The formula $\boldsymbol{F_{net}} = \frac{mv^2}{r}$ can be used for any point of the centripetal motion. It must be remembered, however, that the force will be different at each point as the velocity is constantly changing.

$m = 60\ \text{kg};\ r = 4.0\ \text{m};\ v_B = 8.9\ \text{m s}^{-1}$

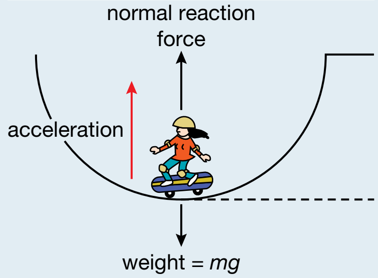
$$\boldsymbol{F_{net}} = \frac{m{v_B}^2}{r} = \frac{60\,\text{kg} \times (8.9\,\text{m s}^{-1})^2}{4.0\,\text{m}}$$

$$= 1176\,\text{N, rounded to } 1200\,\text{N}$$

The net force acting on the skateboarder at point B is 1200 N upwards.

(c) As there is more than one force acting on the skateboarder, it helps to draw a diagram. (See Figure 2.26.)

FIGURE 2.26 Forces acting on the skateboarder.



$$\boldsymbol{F_{net}} = R - \text{W} = R - mg$$

$$\Rightarrow \boldsymbol{R} = \boldsymbol{F_{net}} + \boldsymbol{mg}$$

$$= (1200\,\text{N}) + (60\,\text{kg})\left(9.8\,\text{m s}^{-2}\right)$$

$$= 1788\,\text{N, rounded to } 1800\,\text{N}$$

The normal reaction force acting on the skateboarder at point B is 1800 N upwards. This is larger than the normal reaction force if the skateboarder were stationary. This causes the skateboarder to experience a sensation of heaviness as they come to the bottom of the halfpipe.

2.4.2 Amusing circles

The experience of heaviness described in the previous section, when the reaction force is greater than the weight force, occurs on a roller coaster when the roller-coaster car travels through a dip at the bottom of a vertical arc. When the car is at the top of a vertical arc, the passengers experience a feeling of being lighter. How can this be explained?

When the roller-coaster car is on the top of the track, the reaction force is upwards, and the weight force and the net force are downwards. So,

$$\boldsymbol{F_{net}} = \boldsymbol{mg} - \boldsymbol{R}$$

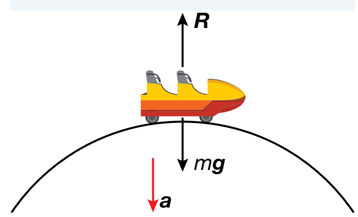
For circular motion, the acceleration is centripetal and is given by the expression

$$\boldsymbol{a_c} = \frac{v^2}{r}$$

As $F_{net} = \boldsymbol{F_c} = m\,\boldsymbol{a_c}$

$$m\frac{v^2}{r} = \boldsymbol{mg} - \boldsymbol{R}$$

FIGURE 2.27 Forces acting at the top of a roller-coaster hill.



The reaction force is a push by the track on the wheels of the roller-coaster car. The track can only push up on the wheels; it cannot pull down on the wheels to provide a downward force. So as the speed increases, there is a limit on how small the reaction force can be. That smallest value is zero.

What would the passenger feel? And what is happening to the roller-coaster car? When the reaction force is zero, the passenger will feel as if they are floating just above the seat. They will feel no compression in the bones of their backside. For the car, at this point it has lost contact with the track. Any attempt to put on the brakes will not slow down the car, as the frictional contact with the track depends on the size of the reaction force. No reaction force means no friction.

Modern roller-coaster cars have two sets of wheels, one set above the track and one set below the track, so that if the car is moving too fast, the track can supply a downward reaction force on the lower set of wheels. The safety features of roller coasters cannot be applied to cars on the road.

If a car goes too fast over a hump on the road, the situation is potentially very dangerous. Loss of contact with the road means that turning the steering wheel to avoid an obstacle or an oncoming car will have no effect whatsoever. The car will continue on in the same direction.

2.4 SAMPLE PROBLEM 2

A passenger is in a roller-coaster car at the top of a circular arc of radius 9.0 m.

(a) At what speed would the reaction force on the passenger equal half their weight force?
(b) What happens to the reaction force as the speed increases?
(c) What would the passenger experience?
(d) At what speed would the passenger feel weightless?

SOLUTION:

(a) Let $R = \frac{W}{2} \Rightarrow R = \frac{m\boldsymbol{g}}{2}$

Using $m\frac{v^2}{r} = m\boldsymbol{g} - R$

$$m\frac{v^2}{r} = m\boldsymbol{g} - \frac{m\boldsymbol{g}}{2}$$

$$\frac{v^2}{r} = \frac{\boldsymbol{g}}{2}$$

$$\Rightarrow v = \sqrt{\frac{\boldsymbol{g}r}{2}} = \sqrt{\frac{9.8\,\text{m}\,\text{s}^{-2} \times 9.0\,\text{m}}{2}} = 6.6\,\text{m}\,\text{s}^{-1}$$

The passenger will feel half as light at a speed of $6.6\,\text{m}\,\text{s}^{-1}$

(b) Rearranging $m\frac{v^2}{r} = m\boldsymbol{g} - R$ gives $R = m\boldsymbol{g} - m\frac{v^2}{r}$

The weight force ($m\boldsymbol{g}$) is constant so, as the speed (v) increases, the reaction force (R) gets smaller.

(c) The reaction force is less than the weight force, so the passenger will feel lighter.

(d) The passenger will feel weightless when the normal reaction force is equal to zero, that is when

$$m\frac{v^2}{r} = \text{m}\boldsymbol{g}$$

Rearranging and cancelling m on both sides, we get

$$v = \sqrt{r\boldsymbol{g}} = \sqrt{9.0\,\text{m} \times 9.8\,\text{m}\,\text{s}^2} = 9.4\,\text{m}\,\text{s}^{-1}$$

The passenger will feel weightless if the speed at the top of the ride is $9.4\,\text{m}\,\text{s}^{-1}$.

2.4 Exercise 1

1. Sometimes, if a car is travelling fast when it comes over the crest of a hill, it can 'catch air' — in other words, become airborne. Explain why this happens in terms of the forces acting on the car and why this is dangerous.
2. A roller-coaster car travels through the bottom of a dip of radius 9.0 m at a speed of 13 m s^{-1}.
   (a) What is the net force on a passenger of mass 60 kg?
   (b) What is the normal reaction force on the passenger by the seat?
   (c) Compare the size of the reaction force to the weight force.
3. A car of mass 800 kg slows down to a speed of 4.0 m s^{-1} to travel over a speed hump that forms the arc of a circle of radius 2.4 m. What normal reaction force acts on the car at the top of the speed hump?
4. At what minimum speed would a car of mass 1000 kg have to travel to momentarily leave the road at the top of the speed hump described in Question 3? (To leave the road, the normal reaction would have to have decreased to zero.)
5. A 100 g lead sinker attached to an 80 cm long string is swung in a vertical circle.
   (a) If the sinker travels at a speed of 10 m s^{-1} at the bottom of the circle, what speed does it have at the top of the circle?
   (b) What is the tension in the string when the sinker is at the bottom of the circle?
   (c) Where is the sinker string most likely to break — while it is at the bottom of the circle or at the top? Explain your answer.
   (d) The breaking strain of the string is 70 N m^{-1}. What is the maximum speed at which the sinker can travel before the string snaps?

2.5 Rotational kinematics and dynamics

2.5.1 Angular velocity

Carousels (also called merry-go-rounds) are a big favourite with younger children in amusement parks. On a fixed carousel, where the horses do not move up and down, the motion of the riders can be described simply in terms of uniform circular motion. If we consider a typical carousel in which there are two concentric circles of horses, although the horses in the two rings have the same period of revolution, the horses in the outer ring travel a greater distance in this time. This means that the average speed of the horses in the outer ring is greater than that of the inner-ring horses.

FIGURE 2.28 Carousels are a big favourite with younger children at amusement parks.

However, the horses all move through the same angular displacement $\Delta\theta$ during the same time period t. This rate of change in angular displacement is called the **average angular velocity ω**.

FIGURE 2.29 The arrangement of concentric rings of horses on a carousel.

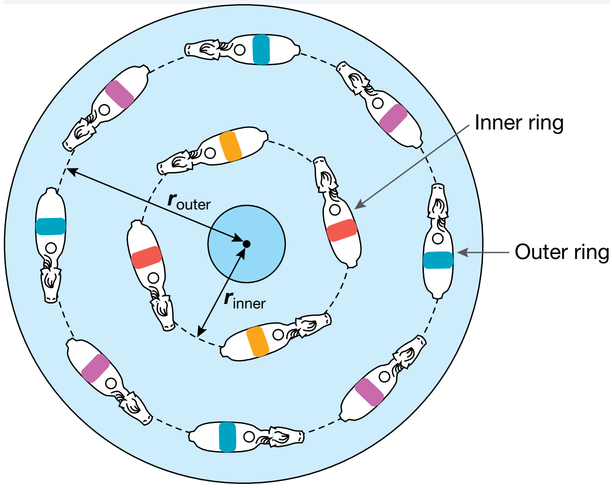
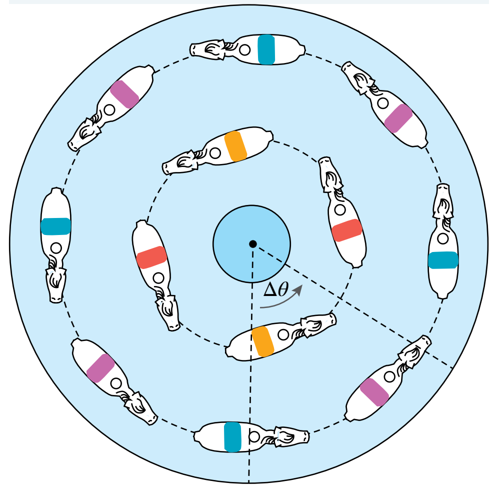


FIGURE 2.30 All the horses move through the same angle $\Delta\theta$ during the same time period t.



While up to now in this course we have used degrees to describe the size of an angle, it is useful in rotational dynamics to describe angles in terms of **radians**.

The angular displacement in radians can be described by the equation

$$\theta = \frac{l}{r}$$

where l is the arc length subtending the angle and r is the radius of the circle. When the arc length is equal in size to the radius, θ is equal to one radian (1 rad).

For a full circle of radius r, the arc length will be equal to the circumference:

$$l = 2\pi r$$

and then

$$\theta = \frac{2\pi r}{r}$$

$$\Rightarrow \theta = 2\pi \text{ radians}$$

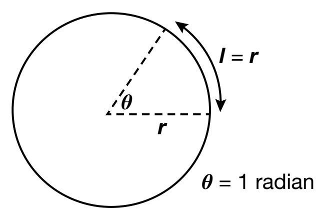
As there are 360° in a full circle,

$$2\pi \text{ radians} = 360°$$

which means that

$$1 \text{ radian} = \frac{360°}{2\pi} \cong 57.3°$$

FIGURE 2.31 One radian is the angle subtended by an arc length l which is equal in magnitude to the radius r of the circle.



2.5 SAMPLE PROBLEM 1

Convert the following angles:
(a) 30° into radians
(b) 4 radians into degrees

SOLUTION:

(a) $\dfrac{30°}{360°} = \dfrac{\theta}{2\pi}$

$$\Rightarrow \theta = \frac{30° \times 2\pi}{360°}$$

$$= 0.52\ \text{rad}$$

Alternatively,

$$\theta = \frac{30°}{57.3°\ \text{rad}^{-1}} = 0.52\ \text{rad}$$

(b) $\dfrac{4}{2\pi} = \dfrac{\theta}{360°}$

$$\Rightarrow \theta = \frac{4 \times 360°}{2\pi}$$

$$= 229.2°$$

We may alternatively convert as follows:

$4\ \text{rad} = 4\ \text{rad} \times 57.3°\ \text{rad}^{-1} = 229.2°$

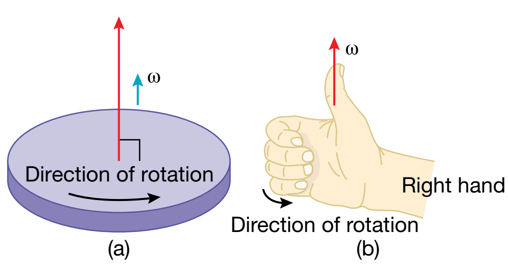
Average angular velocity is described by the equation

$$\omega = \frac{\Delta\theta}{t}$$

where $\Delta\theta$ is the angular displacement in radians of an object moving in a circle and t is the time taken for the displacement. Angular displacement is a vector quantity with direction usually described in terms of being 'clockwise' or 'anticlockwise'.

Angular velocity has the SI units of rad s^{-1} although you may also see the units deg s^{-1} or rpm (revolutions per minute) used. Angular velocity is also a vector quantity; the direction of the angular velocity vector requires the use of a 'right-hand rule' as shown in Figure 2.32. By curling the fingers of the right hand in the direction of rotation, the thumb points in the direction of the angular velocity.

FIGURE 2.32 A right-hand rule is used to determine the direction of the average angular velocity vector.



2.5 SAMPLE PROBLEM 2

An ant has somehow managed to find its way onto a record sitting on a turntable. When the record turns, it rotates at 45 revolutions per minute (45 rpm) in a clockwise direction. The diameter of the record is 18 cm and the ant is positioned 4 cm from the record's centre.

(a) Convert the record's angular speed from rpm to (i) rad s^{-1} (ii) deg s^{-1}.
(b) What is the ant's angular displacement after 2 seconds in (i) radians (ii) degrees?
(c) How many revolutions has the ant made on the record after 2 seconds?

SOLUTION:

(a) i. As 1 revolution = 2π radians and 1 minute = 60 seconds,

$$\frac{45 \text{ revolutions}}{1 \text{ minute}} = \frac{45 \text{ rev} \times 2\pi \text{ rad rev}^{-1}}{1 \text{ min} \times 60 \text{ s min}^{-1}}$$
$$= \frac{3\pi}{2} \text{rad s}^{-1}$$

We use the right-hand rule to determine the direction of the angular velocity vector. As the disc turns clockwise, angular velocity will be directed downwards.

ii. As 1 revolution = 360° and 1 minute = 60 s,

$$\frac{45 \text{ revolutions}}{1 \text{ minute}} = \frac{45 \text{ rev} \times 360 \text{ deg rev}^{-1}}{1 \text{ min} \times 60 \text{ s min}^{-1}}$$
$$= \frac{16\,200 \text{ deg}}{60 \text{ s}}$$
$$= 270 \text{ deg s}^{-1}\text{, downwards}$$

(b) $\omega = \dfrac{\Delta\theta}{t} \Rightarrow \Delta\theta = \omega t$

i. $\Delta\theta = \omega t$

$$= \left(\frac{3\pi}{2} \text{rad s}^{-1}\right)(2 \text{ s})$$
$$= 3\pi \text{ rad, clockwise}$$

ii. $\Delta\theta = \omega t$
$$= (270 \text{ deg s}^{-1})(2 \text{ s})$$
$$= 540°\text{, clockwise}$$

(c) As there are 360° in a circle, number of revolutions $= \dfrac{540°}{360°} = 1.5$

The ant has made 1.5 revolutions after 2 s

2.5.2 Relating angular and tangential velocities

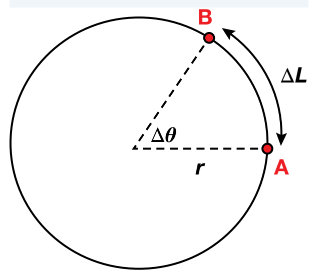
Objects moving with uniform circular motion can be described in terms of both tangential and angular velocity and an expression can be derived that relates them.

The velocity of an object in uniform circular motion as it moves from position A to position B is equal to the rate at which it travels the distance between the two points. This distance is not a straight line but an arc that we shall describe as ΔL. We can then write:

$$v = \frac{\Delta L}{\Delta t}$$

where Δt is the time interval for the object to travel from A to B.

FIGURE 2.33 An object moving in a circle of radius r from positon A to position B.



Let $\Delta\theta$ be the angle subtended by the arc between A and B.

The length of the arc ΔL depends upon the radius and the fraction of the circumference subtending $\Delta\theta$:

$$\Delta L = \frac{\Delta\theta}{2\pi} \times C$$
$$= \frac{\Delta\theta}{2\pi} \times 2\pi r$$

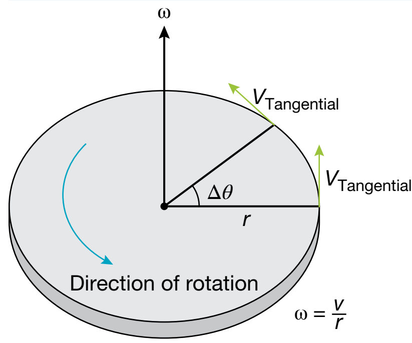
Thus, we find:

$$\Delta L = \Delta\theta\, r$$

As $v = \dfrac{\Delta L}{\Delta t}$ and $\omega = \dfrac{\Delta\theta}{\Delta t}$

$$\Rightarrow v\,\Delta t = (\omega\,\Delta t)\ r$$
$$\Rightarrow v = \omega\, r$$

FIGURE 2.34 The relationship between angular and tangential velocity.



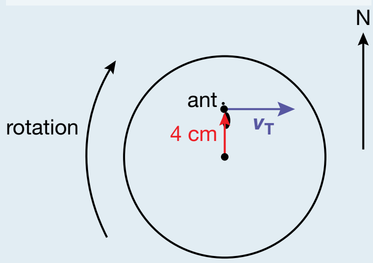
2.5 SAMPLE PROBLEM 3

(a) Determine (i) the tangential velocity (ii) the angular velocity of the ant in 2.5 Sample Problem 2 when the line between the centre spindle and the ant is aligned in a northerly direction.

(b) The ant walks to the edge of the record, which is 9 cm from the centre spindle. Will the magnitudes of his tangential and angular velocities increase, decrease or remain the same?

SOLUTION:

FIGURE 2.35



(a) i. The ant is located 4 centimetres from the centre of rotation, so $r = 4\,\text{cm} = 0.04\,\text{m}$.

The record makes 45 revolutions in 1 minute, so we can calculate the period of the ant's circular motion:

$$T = \frac{60\,\text{s}}{45} = 1.3\,\text{s}$$

$$\Rightarrow v = \frac{2\pi r}{T} = \frac{2\pi \times 0.04\,\text{m}}{1.3\,\text{s}} = 0.19\,\text{m s}^{-1}$$

The tangential velocity will be directed at right angles to the radius at that point — that is to the east.

The ant's tangential velocity will be $0.19\,\text{m s}^{-1}$, east

ii. Rearranging the relationship $v = \omega\, r$ to get

$$\omega = \frac{v}{r} = \frac{0.19\,\text{m s}^{-1}}{0.04\,\text{m}} = 4.7\,\text{rad s}^{-1}$$

As the record is rotating clockwise, the direction of angular velocity will be downwards.

The ant's angular velocity is 4.7 rad s^{-1}, downwards. (Note that this is the same value as obtained in Sample Problem 2 (a) although different means were used to obtain it!)

(b) In moving to the edge, the radius of the circle through which the ant moves has increased so he moves a greater distance in the 1.3 seconds the record takes each revolution. As a result, the ant's tangential velocity has increased.

The new position of the ant makes no difference to his angular velocity as every point on the record moves through the same angle at the same rate.

2.5.3 Turning effect and forces

A strong wind blowing a tree over in a storm, a door opening when you pull on its handle and using a spanner to take the wheel nuts off when a tyre is changed are all examples of forces causing something to rotate.

The rotational effect that a force has on an object is referred to as **torque**, represented in equations by the Greek letter $\boldsymbol{\tau}$ ('tau'). The amount of torque that a force generates around a pivot point depends upon the magnitude of the force, the direction in which it is applied and how far from the pivot point the force is applied.

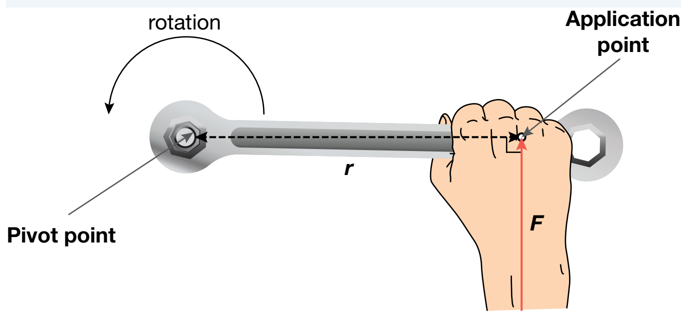
If a force $\boldsymbol{F}$ is applied at right angles to the line between the pivot point and the application point, then torque is calculated simply as:

$\boldsymbol{\tau} = \boldsymbol{Fr}$

where $\boldsymbol{r}$ is the distance between the pivot point and the application point.

The units for torque are newton-metres (N m). Torque is a vector quantity and the direction of the torque is described by considering the direction ('clockwise' or 'anticlockwise') of the rotation it produces in much the same way as the direction of the angular velocity vector (see Figure 2.36). Using the right-hand rule with the fingers curled in the direction of the rotation, the thumb points in the direction of the torque. We can see that the direction of the torque is perpendicular to the plane containing F and r.

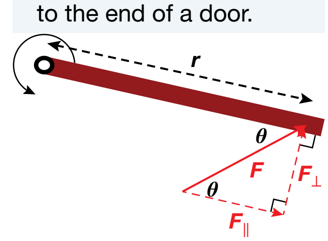
FIGURE 2.36 A force $\boldsymbol{F}$ exerted perpendicularly on a spanner at a distance $\boldsymbol{r}$ from the bolt (pivot point) produces an anticlockwise rotation of the bolt, tightening it. The torque that the force has produced is equal to the product $\boldsymbol{Fr}$ and is directed out of the page.



If the force is applied at a non-perpendicular angle, only the force component acting at right angles to r produces torque around the pivot point. This component is designated $\boldsymbol{F}_{\perp}$.

In Figure 2.37, a force $\boldsymbol{F}$ is applied at an angle θ to a door in order to close it. It is only the component of this force that is perpendicular to the door ($\boldsymbol{F}_{\perp}$) that contributes to the torque that turns the door on its hinge, that is

FIGURE 2.37 A force $\boldsymbol{F}$ is applied at an angle θ to the end of a door.



$\boldsymbol{\tau} = \boldsymbol{F}_{\perp} r$

As $\boldsymbol{F}_{\perp} = \boldsymbol{F} \sin \theta$,

$\boldsymbol{\tau} = (\boldsymbol{F} \sin \theta)\, r$

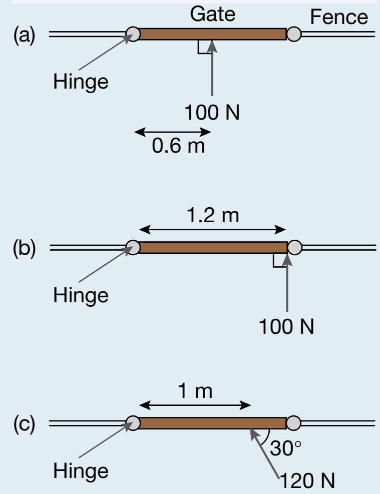
or

$\boldsymbol{\tau} = \boldsymbol{F}\boldsymbol{r} \sin \theta$

2.5 SAMPLE PROBLEM 4

A 1.2 m wide gate in the school fence will only open if there is a torque of 100 N m acting on it. In which of the following cases will the force applied to the gate result in it opening?

FIGURE 2.38



SOLUTION:

In each case, the direction of the torque produced will be out of the page.

(a) $\boldsymbol{r} = 0.6\text{ m}, \boldsymbol{F} = 100\text{ N}, \theta = 90°, \boldsymbol{t} = ?$

$$\boldsymbol{\tau} = \boldsymbol{F}\, r \sin\theta$$
$$= (100\text{ N})\ (0.6\text{ m})\ \sin 90°$$
$$= 60\text{ N m}$$

This is less than the required 100 N m. Therefore, the gate will not open.

(b) $\boldsymbol{r} = 1.2\text{ m}, \boldsymbol{F} = 100\text{ N}, \theta = 90°, \boldsymbol{\tau} = ?$

$$\boldsymbol{\tau} = \boldsymbol{F}\,\boldsymbol{r} \sin\theta$$
$$= (100\text{ N})\,(1.2\text{ m})\ \sin 90°$$
$$= 120\text{ N m}$$

The gate will open.

(c) $\boldsymbol{r} = 1.0\text{ m}, \boldsymbol{F} = 120\text{ N}, \theta = 30°$

$$\boldsymbol{\tau} = \boldsymbol{F}\,\boldsymbol{r} \sin\theta$$
$$= (120\text{ N})\ (1.0\text{ m})\ \sin 30°$$
$$= 60\text{ N m}$$

The gate will not open.

2.5 Exercise 1

1 Convert the following degree angles into radian angles:
(a) 30° (b) 180° (c) 350° (d) 460°

2 A wheel is oriented horizontally and turning in a clockwise direction when viewed from above. When a point on the wheel is to the east of the axle, in what direction is the
(a) tangential velocity
(b) angular velocity vector acting at that point?

3 A jeweller uses a 2 cm diameter tungsten carbide grinding wheel that is spinning at 15 000 rpm.
(a) What is the magnitude of the average angular velocity of the wheel in rad s^{-1}?
(b) Calculate the magnitude of the wheel's tangential velocity at its edge.
(c) If she changes to a 3 cm diameter wheel, what angular speed (in rpm) will it need to run at if the edge of the new wheel is to have the same tangential velocity as the 2 cm wheel?

4 A bolt securing a beam to a pillar requires a torque of 150 N m to loosen it. What is the minimum force that you will need to apply to the handle of a 30 cm long spanner if you are to loosen the bolt?

5 A force of 30 N is applied to the edge of a wheel that has a radius of 45 cm. If the resulting torque produced in the wheel is 5 N m, at what angle was the force applied to the radius?

6 The handle of a torque wrench is hollow so an extension rod can be inserted. If you can exert only 30 N of force, how far along the extension rod from the handle should you place your hand to achieve a torque of 30 N m?

2.6 Review

2.6.1 Summary

- The time taken for an object moving with uniform circular motion to make one revolution of the circle is called the period T.
- The average speed of an object moving with uniform circular motion is expressed by the equation, average speed $= \frac{2\pi r}{T}$ where r is the radius of the circle in which the object moves and T is the period.
- The average velocity of an object moving in a circle is equal to zero as its displacement for a complete revolution is zero.
- The instantaneous velocity (or tangential velocity) of an object moving uniformly in a circle has a magnitude equal to that of the object's average speed. The direction of this velocity is tangential to the circle in which it moves.
- Angular displacement $\Delta\theta$ is the angle through which an object moves between two points on the circumference of a circle. It is measured in radians. There are 2π radians in a full circle; 1 radian $\cong 57.3°$
- The angular velocity ω of an object is equal to its rate of angular displacement, $\omega = \frac{\Delta\theta}{t}$.
- The units of angular velocity are rad $\mathrm{s^{-1}}$. Angular velocity is a vector that is directed away from or into the axis of rotation and is at right angles to the plane of rotation.
- An object moving in a circle at constant speed has centripetal acceleration $\boldsymbol{a_c}$, which is directed towards the centre of the circle; this is described by the equation $a_c = \frac{v^2}{r}$ where v is the object's tangential velocity and r is the radius of the circle in which it travels.
- The centripetal force $\boldsymbol{F_c}$ can be calculated by the equations $\boldsymbol{F_c} = m\boldsymbol{a_c}$ or $\boldsymbol{F_c} = \frac{mv^2}{r}$.
- Torque $\boldsymbol{\tau}$ describes the turning effect of a force $\boldsymbol{\tau} = \boldsymbol{F}r\sin\theta$ where $\boldsymbol{F}$ is the force applied to an object at a position $\boldsymbol{r}$ from the pivot point and θ is the angle between $\boldsymbol{F}$ and $\boldsymbol{r}$. The units for torque are N m. Torque is a vector quantity; the vector is directed into or out of the axis of rotation and is perpendicular to the plane described by $\boldsymbol{F}_\perp$ and $\boldsymbol{r}$.

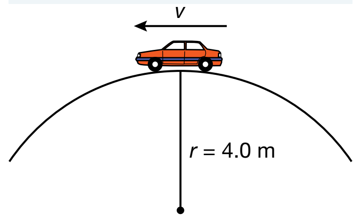
2.6.2 Questions

1. A 400 g rock is tied to the end of a 2 m long string and whirled until it has a speed of 12.5 m $\mathrm{s^{-1}}$. Calculate the centripetal force and acceleration experienced by the rock.
2. A 900 kg motorcycle, travelling at 70 km $\mathrm{h^{-1}}$, rounds a bend in the road with a radius of 17.5 m. Calculate the centripetal force required from the friction between the tyres and the road.
3. A carousel horse located 2 m from the centre of rotation makes 1 revolution every 5 seconds. What will be its:
   (a) period?
   (b) average speed?
   (c) acceleration?
   (d) angular velocity?
4. A ceiling fan in a classroom rotates steadily at 1000 rpm in a clockwise direction (assume you are looking up at the fan). Someone has stuck a piece of chewing gum to one of the blades at a distance of 40 cm from the centre of the fan.
   (a) How long does it take for the gum to make one full revolution of the fan?
   (b) What is the fan's speed in rad $\mathrm{s^{-1}}$?
   (c) What is the magnitude of the gum's instantaneous velocity at any moment in time?
   (d) If the gum were to detach from the fan, describe the path that it would travel on its way to the floor.
5. A 60 kg diver stands at the very edge of a 2 m long diving board. What is the torque acting on the diving board's anchorage point?

6. Explain why door handles are placed at the edge of a door rather than in the middle of the door.
7. In what direction is the torque acting when you exert a force to unscrew the cap of a bottle? Assume that the bottle lid unscrews in an anticlockwise direction and that the bottle is being held upright on a bench.
8. A jogger, of mass 65 kg, runs around a circular track of radius 120 m with an average speed of 6.0 km h^{-1}.
   (a) What is the centripetal acceleration of the jogger?
   (b) What is the net force acting on the jogger?
9. At the school fete, Lucy and Natasha have a ride on the merry-go-round. The merry-go-round completes one turn every 35 s. Natasha's horse is 2.5 m from the centre of the ride, while Lucy's horse is a further 70 cm out. Which girl would experience the greatest centripetal acceleration? Support your answer with calculations.
10. At a children's amusement park, the miniature train ride completes a circuit of radius 350 m, maintaining a constant speed of 15 km h^{-1}.
    (a) What is the centripetal acceleration of the train?
    (b) What is the net force acting on a 35 kg child riding on the train?
    (c) What is the net force acting on the 1500 kg train?
    (d) Explain why the net forces acting on the child and the train are different and yet the train and the child are moving along the same path.
11. The toy car in a slot car set runs on a circular track. The track has a radius of 65 cm, and the 0.12 kg car completes one circuit in 5.2 s.
    (a) What is the centripetal acceleration of the car?
    (b) What is the net force acting on the car?
    (c) Draw a labelled diagram showing all the forces acting on the car. Also include the direction and magnitude of the net force on your diagram.
12. When a mass moves in a circle, it is subject to a net force. This force acts at right angles to the direction of motion of the mass at any point in time. Use Newton's laws to explain why the mass does not need a propelling force to act in the direction of its motion.
13. Explain why motorcyclists lean into bends.
14. A rubber stopper of mass 50.0 g is whirled in a horizontal circle on the end of a 1.50 m length of string. The time taken for ten complete revolutions of the stopper is 8.00 s. The string makes an angle of 6.03° with the horizontal. Calculate:
    (a) the speed of the stopper
    (b) the centripetal acceleration of the stopper
    (c) the net force acting on the stopper
    (d) the magnitude of the tension in the string.
15. A ball is tied to the end of a string and whirled in a horizontal circle of radius 2.0 m. The string makes an angle of 10° with the horizontal. The tension in the string is 12 N.
    (a) Calculate the magnitude of the centripetal force acting on the ball.
    (b) If the mass of the ball is 200 g, what is its speed?
    (c) What is the period of revolution of the ball?
16. Carl is riding around a corner on his bike at a constant speed of 15 km h^{-1}. The corner approximates part of a circle of radius 4.5 m. The combined mass of Carl and his bike is 90 kg. Carl keeps the bike in a vertical plane.
    (a) What is the net force acting on Carl and his bike?
    (b) What is the sideways frictional force acting on the tyres of the bike?
    (c) Carl rides onto a patch of oil on the road; the sideways frictional forces are now 90% of their original magnitude. If Carl maintains a constant speed, what will happen to the radius of the circular path he is taking?

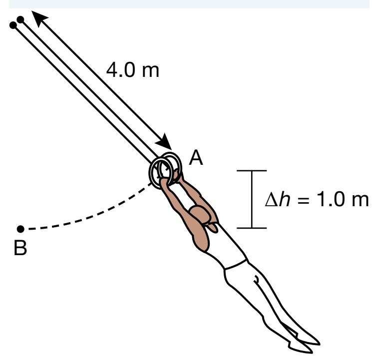
17. A cyclist rounds a bend. The surface of the road is horizontal. The cyclist is forced to lean at an angle of 20° to the vertical to 'only just' take the bend successfully. The total sideways frictional force on the tyres is 360 N. The cycle has a mass of 20 kg. What is the mass of the cyclist?
18. A road is to be banked so that any vehicle can take the bend at a speed of $30\ \text{m s}^{-1}$ without having to rely on sideways friction. The radius of curvature of the road is 12 m. At what angle should it be banked?
19. A car of mass 800 kg travels over the crest of a hill that forms the arc of a circle, as shown in Figure 2.39.
    (a) Draw a labelled diagram showing all the forces acting on the car.
    (b) The car travels just fast enough for the car to leave the ground momentarily at the crest of the hill. This means the normal reaction force is zero at this point.
        i. What is the net force acting on the car at this point?
        ii. What is the speed of the car at this point?

FIGURE 2.39



20. A gymnast, of mass 65 kg, who is swinging on the rings follows the path shown in Figure 2.40.
    (a) What is the speed of the gymnast at point B, if he is at rest at point A?
    (b) What is the centripetal force acting on the gymnast at point B?
    (c) Draw a labelled diagram of the forces acting on the gymnast at point B. Include the magnitude of all forces.

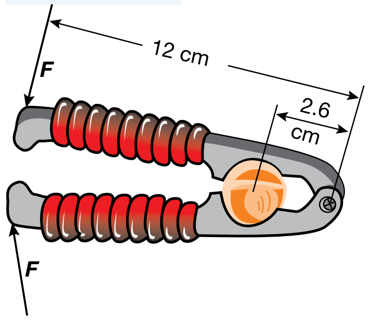
FIGURE 2.40



21. A door that is 820 mm wide can be opened provided that a torque of 10 N m is applied at the hinge. How much force would be needed to open the door if it was applied perpendicularly to the door at a distance of 400 mm from the hinge?

22. To crack a walnut, a force of 40 N must act on its shell from both sides. How much force will you need to exert perpendicularly to the handles of the nutcracker in Figure 2.41 below in order to crack the walnut?

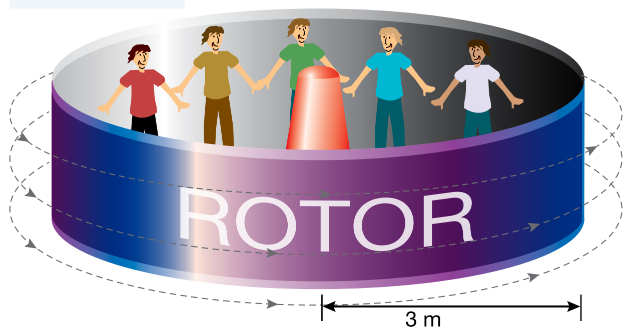
FIGURE 2.41



23. A jet pilot flies his aircraft in a vertical loop. If the jet is flying at 720 km h^{-1} at the lowest point of the loop, find:
   (a) The smallest radius that the loop can have if the acceleration at the bottom is not to exceed 6 g.
   (b) The apparent weight experienced by the 70 kg pilot.

24. On a ride called the Rotor, the riders lean against a vertical wall as shown in Figure 2.42.
   (a) If the Rotor has a radius of 3 m, with what period will it rotate if the horizontal centripetal acceleration acting on the riders is to equal 4g?
   (b) At what speed will the Rotor be rotating at this time?
   (c) During the ride, the floor drops away but the riders stay pressed against the walls. Why don't they fall?

FIGURE 2.42



25. Alistair is riding on a chair-o-plane when the supporting chain suddenly snaps. At this time, his chair was located 4 m from the centre of rotation and was 1.5 m above the ground. If the chain made an angle of 30° with the central pole when the chain snapped, how far horizontally will he travel before he lands?

eBook*plus* RESOURCES

Complete this digital doc: Investigation: Exploring circular motion
Searchlight ID: doc-26318

PRACTICAL INVESTIGATIONS

Investigation 2.1 Exploring circular motion

Aim

To examine some of the factors affecting the motion of an object undergoing uniform circular motion, and then to determine the quantitative relationship between the variables of force, velocity and radius.

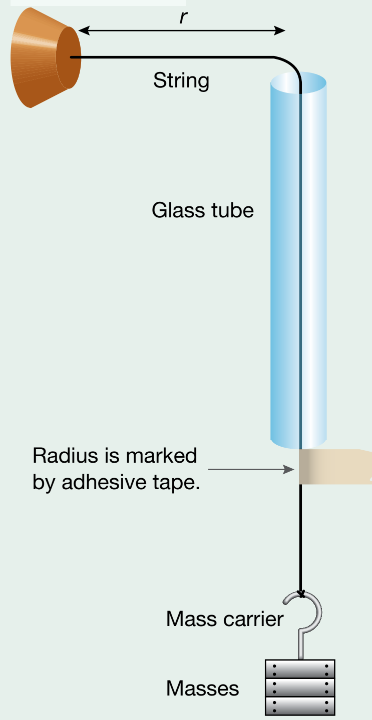
Apparatus

- rubber stopper
- glass tube
- 50 g mass carrier
- stopwatch
- string
- sticky tape
- metre rule
- 50 g slot masses

Method

1. Record the mass of the rubber stopper being used as a bob.
2. Attach the rubber stopper to a length of string approximately 1.5 m long, then thread the loose end of the string through the glass tube.
3. Attach the mass carrier to the loose end of the string as shown in Figure 2.43.
4. Place a piece of sticky tape on the string at the point shown in the diagram so that the distance, r, is 40 cm.
5. Hold the glass tube and move it in a small circle so as to get the rubber bob moving in a circular path. The mass carrier will provide the centripetal force to keep the bob moving in its circular path. Adjust your frequency of rotation so that the sticky tape just touches the bottom of the glass tube. This will keep the radius of the bob's orbit steady.
6. Record the time for the bob to complete 10 revolutions at a constant speed then calculate and record the period. Do this three times and then use the average of these as the correct period. Use the radius and period to calculate the average velocity of the bob.
7. Repeat steps 4 to 6 for radii of 0.60 m, 0.80 m, and 1.0 m.
8. Repeat steps 3 to 7 using masses of 100 g, then 150 g, and finally 200 g

FIGURE 2.43



Results

Tabulate your results as shown in the table below.

TABLE 2.1

Force	RADIUS	PERIOD (10 REVOLUTIONS)			MEAN PERIOD	ORBITAL VELOCITY v
50 g × gravity	1.0 m					
	0.8 m					
	0.6 m					
	0.4 m					
100 g × gravity	1.0 m					
	0.8 m					
	0.6 m					
	0.4 m					

Analysis

1. From the results above, calculate the average velocity of the bob and complete the table.
2. For each of the radii used with 50 g, construct a graph of v^2 versus r.
3. Repeat this for the 100 g, 150 g and 200 g results.

Questions

1. What is the relationship that these graphs indicate?
2. What does the slope of your v^2 versus r graph represent?
3. What role does gravity play in the results in this experiment?

WORKING SCIENTIFICALLY 2.1

In a yoyo trick called 'walking the dog', the yoyo goes to the end of the string and spins in place before being drawn back up. Design a method to measure the average speed of a yoyo in this spinning position as accurately as possible. Use this method in an experiment to determine what effect the tightness in the twist of the yoyo string has on yoyo spin speed.

WORKING SCIENTIFICALLY 2.2

A centrifuge is a device that is often used in laboratories to separate mixtures of materials that have different masses. A low-tech version of this is used by campers: by swinging a billy full of tea leaves and boiling water at arm's length in a vertical circle, the tea leaves are (theoretically) pushed to the bottom of the billy so fewer of them escape with the tea when it's being poured into a mug. Does this really work or this just a bush myth? Is there a minimum number of circles that the billy needs to make for this to work? Does this method work equally well for coffee grounds? Design and perform an investigation to find out the answer to one of these questions. Note: care should be taken not to use boiling water as there is a very real danger of scalding. Room temperature water should work just fine!

WORKING SCIENTIFICALLY 2.3

Design and build an automatic lazy Susan that is powered by no more than 4 AA batteries, is controlled by a switch allowing it to be turned on and off and incorporates three different speeds. The device should be designed in such a way that, even at the highest speed, a bowl with a mass of 100 g does not drift off.

WORKING SCIENTIFICALLY 2.4

Which of the doors in your school is the hardest to open once it is unlocked? Design a method that will allow you to measure the amount of force exerted to swing open an unlocked door. Then, use this method to determine which door in your school requires the greatest amount of torque (not force) to open.

TOPIC 3
Motion and gravitational fields

3.1 Overview

3.1.1 Module 5: Advanced Mechanics

Motion in gravitational fields

Inquiry question: How does the force of gravity determine the motion of planets and satellites?
Students:

- apply qualitatively and quantitatively Newton's Law of Universal Gravitation to:
  - determine the force of gravity between two objects: $F = \dfrac{GMm}{r^2}$
  - investigate the factors that affect the gravitational field strength: $g = \dfrac{GM}{r^2}$
  - predict the gravitational field strength at any point in a gravitational field, including at the surface of a planet (ACSPH094, ACSPH095, ACSPH097)
- investigate the orbital motion of planets and artificial satellites when applying the relationships between the following quantities:
  - gravitational force
  - centripetal force
  - centripetal acceleration
  - mass
  - orbital radius
  - orbital velocity
  - orbital period
- predict quantitatively the orbital properties of planets and satellites in a variety of situations, including near the Earth and geostationary orbits, and relate these to their uses (ACSPH101)
- investigate the relationship of Kepler's Laws of Planetary Motion to the forces acting on, and the total energy of, planets in circular and non-circular orbits using: (ACSPH101)
  - $v = \dfrac{2\pi r}{T}$
  - $\dfrac{r^3}{T^2} = \dfrac{GM}{4\pi^2}$
- derive quantitatively and apply the concepts of gravitational force and gravitational potential energy in radial gravitational fields to a variety of situations, including but not limited to:
  - the concept of escape velocity: $v_{esc} = \sqrt{\dfrac{2GM}{r}}$
  - total potential energy of a planet or satellite in its orbit: $U = -\dfrac{GMm}{r}$
  - total energy of a planet or satellite in its orbit: $U + K = -\dfrac{GMm}{2r}$
  - energy changes that occur when satellites move between orbits (ACSPH096)
  - Kepler's Laws of Planetary Motion (ACSPH101)

FIGURE 3.1 Understanding gravitational forces has allowed us to put satellites in orbit around the Earth.

3.2 Explaining the solar system

3.2.1 The heliocentric solar system

Up until the dawn of the Renaissance, the widely held view of the solar system's mechanics placed the Earth firmly in the centre, with the Sun, Moon and other planets moving in circular orbits around it. This view was overturned by the work of several key figures during what is now considered to be the Golden Age of Science.

In 1542, Nicolas Copernicus (1473–1543) published 'On the Revolution of the Heavenly Orbs', outlining an explanation for the observations of planetary motion with the Sun at the centre. In his explanation, the planets moved in circular orbits about the Sun. Copernicus's model became increasingly preferred over the geocentric model of Ptolemy because it made astronomical and astrological calculations easier. The publication had a significant scientific, social and political impact during the latter part of the sixteenth century.

Galileo Galilei (1564–1642) was a strong advocate for the view that the Copernican model was more than 'a set of mathematical contrivances, merely to provide a correct basis for calculation' and instead represented physical reality. (This had also been Copernicus's view, but he could not express this in print.) Galileo thought that astronomy could now ask questions about the structure, fabric and operation of the heavens, but as with so many of his scientific interests, Galileo did not pursue these questions further.

Johannes Kepler (1571–1630) decided his purpose in life was to reveal the fundamental coherence of a planetary system with the Sun at its centre. In 1600–1601 he was working as an assistant to Tycho Brahe (1546–1601), a Danish astronomer who had been compiling very precise measurements of the planets' positions for over twenty years. Brahe's measurements were so accurate that they are comparable to those achieved with today's more sophisticated technology.

Kepler was seeking to find patterns and relationships between the motions of the various planets. He used Brahe's data to calculate the positions of the planets as they would be observed by someone outside the solar system, rather than from the revolving platform of the Earth. Initially he was looking for circular orbits, but Brahe's precise data did not fit such orbits.

3.2.2 Kepler's First Law — The Law of Ellipses

After Kepler used a variety of other geometric shapes to model the orbits of the planets, in 1604 he formulated what is known as Kepler's First Law:

Each planet moves, not in a circle, but in an ellipse, with the sun, off centre, at a focus.

Ellipses can be thought of as elongated or stretched circles; the degree of stretch is known as **eccentricity**. The eccentricity of an ellipse is defined as the ratio $\frac{c}{2a}$, where c is the distance between the two foci of the ellipse and a is the semi-major axis (see Figure 3.2). A circle is an ellipse with an eccentricity of zero.

FIGURE 3.2 The geometry of an ellipse.

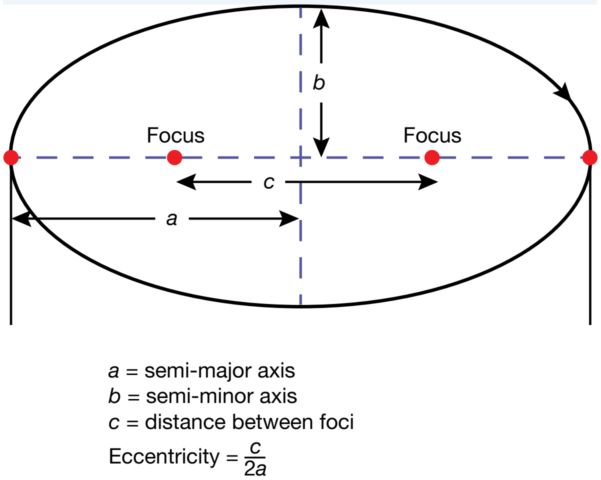
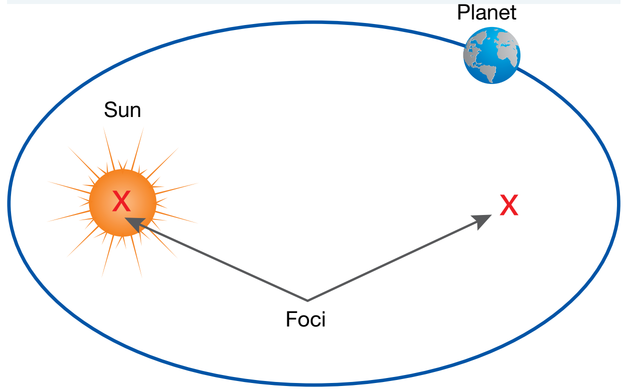


FIGURE 3.3 Kepler stated that, for a planet's elliptical orbit, the Sun is at one focus and nothing is located at the other focus.



Most of the planets in our solar system have orbits with small eccentricities, as shown in Table 3.1. In contrast to the planets, most comets and some asteroids have very eccentric orbits. Our own Moon has an orbit with an eccentricity of 0.0549, although this can vary due to a number of influences, not the least of which is the strong gravitational attraction of the Sun.

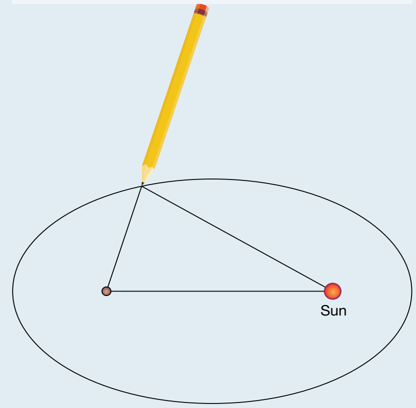
TABLE 3.1 Eccentricities of planetary orbits.

Planet	Eccentricity
Mercury	0.2056
Venus	0.0068
Earth	0.0167
Mars	0.0934
Jupiter	0.0483
Saturn	0.0560
Uranus	0.0461
Neptune	0.0097

TRY THIS

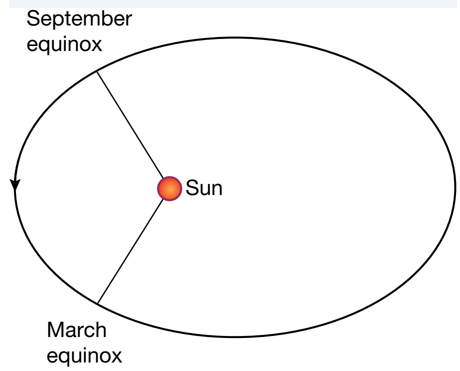
An ellipse is like a stretched circle. The shape can be drawn by placing two pins on the page several centimetres apart, with a loose piece of string tied between the pins. If a pencil is placed against the string to keep it tight and then the pencil is moved around the page, the drawn shape is an ellipse with a focus at each of the pins. The closer the two foci, the more like a circle the ellipse becomes.

FIGURE 3.4 Drawing an ellipse.



Kepler's theory of the elliptical nature of planetary orbits was supported by the timing of the Earth's solar equinoxes. The equinoxes are the two days in the year when the Sun is directly above the equator and the durations of night and day are equal. They occur when the line drawn from the Sun to the Earth is at right angles to the Earth's orbit. If the orbit of the Earth around the Sun was a perfect circle, then the time periods elapsing between the equinoxes would be of equal length. This is not the case. In fact, the time from the March equinox until the September equinox is several days longer than the time period from the September equinox to the next March equinox. This difference is consistent with an elliptical orbit where, as the Sun is located at one of two foci, the two points at which the equinoxes occur will not be directly opposite each other.

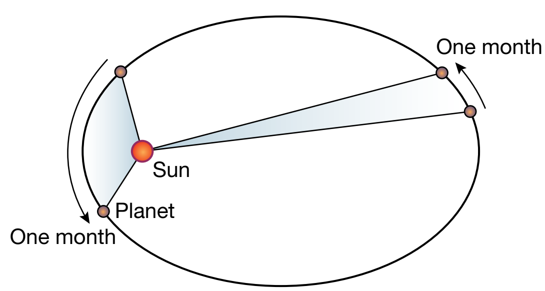
FIGURE 3.5 Location of the two equinoxes in the Earth's orbit; note that the elliptical path of the Earth's orbit is greatly exaggerated in this diagram for clarity.



3.2.3 Kepler's Second Law – The Law of Equal Areas

Kepler also looked at the speeds of the planets in their orbits. His analysis of the data showed that speeds of the planets were not constant. The planets were slower when they were further away from the Sun and faster when closer. He also found that their angular speed — the number of degrees a line from the Sun to a planet sweeps through every day — was not constant. These observations supported his theory as to the elliptical nature of planetary orbits. Further analysis revealed that, while the planets covered different distances in equal time periods, each planet swept out the same *area* per unit time. This is summarised in Kepler's Second Law, published in 1609:

FIGURE 3.6 Kepler's Second Law. The regions swept out by a planet during each month are equal in area.



The linear speed and angular speed of a planet are not constant, but the areal speed of each planet is constant. That is, a line joining the sun to a planet sweeps out equal areas in equal times.

3.2.4 Kepler's Third Law – The Law of Periods

Kepler was keen to find a mathematical relationship between the period of a planet's orbit around the Sun and its average radius that gave the same result for each planet. He tried numerous possibilities and eventually, in 1619, he found a relationship that was consistent with the data. This is Kepler's Third Law:

For all planets, the cube of the average radius is proportional to the square of the orbital period; that is, $\frac{R^3}{T^2}$ *is a constant for all planets going around the Sun.*

Kepler was also able to show that the relationship held for the orbits of the moons of Jupiter.

Kepler had constructed as detailed a description of the solar system as was possible without a mechanism to explain the motion of the planets, although he did understand gravity as a reciprocal attraction. Kepler wrote, 'Gravity is the mutual tendency between bodies towards unity or contact (of which the magnetic force also is), so that the Earth draws a stone much more than the stone draws the Earth … '

TABLE 3.2 The solar system: some useful data.

Body	Mass (kg)	Radius of body (m)	Mean radius of orbit (m)	Period of revolution (s)
Sun	1.99×10^{30}	6.96×10^{8}	Not applicable	Not applicable
Earth	5.97×10^{24}	6.37×10^{6}	1.50×10^{11}	3.16×10^{7}
• Moon	7.35×10^{22}	1.74×10^{6}	3.84×10^{8}	2.36×10^{6}
Mercury	3.30×10^{23}	2.44×10^{6}	5.79×10^{10}	7.60×10^{6}
Venus	4.87×10^{24}	6.05×10^{6}	1.08×10^{11}	1.94×10^{7}
Mars	6.42×10^{23}	3.40×10^{6}	2.28×10^{11}	5.94×10^{7}
Jupiter	1.90×10^{27}	7.15×10^{7}	7.78×10^{11}	3.74×10^{8}
Saturn	5.68×10^{26}	6.03×10^{7}	1.43×10^{12}	9.29×10^{8}
Uranus	8.68×10^{25}	2.59×10^{7}	2.87×10^{12}	2.64×10^{9}
Neptune	1.02×10^{26}	2.48×10^{7}	4.50×10^{12}	5.17×10^{9}
Pluto*	1.46×10^{22}	1.18×10^{6}	5.90×10^{12}	7.82×10^{9}

*Pluto is no longer classified as a planet. Scientists have recently hypothesised that a ninth planet may exist, but it has not yet been directly observed.

3.2 SAMPLE PROBLEM 1

Use the information in Table 3.2 to determine the average time taken for a satellite to orbit the Earth at an altitude of 500 km from its surface.

SOLUTION:

As the satellite and the Moon both orbit the Earth, they will have the same ratio for the value of $\frac{R^3}{T^2}$ where R is the orbital radius for each object's orbit. Note that this is measured between their centres of mass.

$$R_{\text{satellite}} = \text{satellite altitude} + \text{radius of the Earth} = 5 \times 10^5\ \text{m} + 6.37 \times 10^6\ \text{m} = 6.87 \times 10^6\ \text{m}$$

$$\mathrm{R_{Moon}} = 3.84 \times 10^8 \text{m}$$

$$\mathrm{T_{Moon}} = 2.36 \times 10^6 \text{s}$$

Given that

$$\frac{R^3_{\text{Moon}}}{T^2_{\text{Moon}}} = \frac{R^3_{\text{satellite}}}{T^2_{\text{satellite}}}$$

$$\Rightarrow T_{\text{satellite}} = \sqrt{\frac{R^3_{\text{satellite}} \times T^2_{\text{Moon}}}{R^3_{\text{Moon}}}}$$

$$= \sqrt{\frac{(6.87 \times 10^6 \text{m})^3 \times (2.36 \times 10^6 \text{s})^2}{(3.84 \times 10^8 \text{m})^3}}$$

$$= 5.65 \times 10^3 \text{s}$$

The satellite has an orbital period of 5.65×10^3 s (approximately 1 hour and 34 minutes).

3.2 Exercise 1

1 Which of the following statements is true?
- (a) The orbits of planets around the Sun are in the shape of slightly squashed circles.
- (b) Planets move faster in their orbits the further away they are from the Sun.
- (c) The Sun is in the exact geometric centre of our solar system.
- (d) A planet that is twice as far away from the Sun as the Earth will take twice as long as the Earth to complete its orbit.

2 An AU (Astronomical Unit) is defined as being equal to the average orbital radius of the Earth from the Sun. Given that the Venus year is equal to 0.615 Earth years, determine its distance from the Sun in AU.

3 (a) Use the information in Table 3.2 to determine the value of the ratio $\frac{R^3}{T^2}$ for
- i. Mars, and
- ii. Jupiter . Give your answers to 2 significant figures.

(b) Would you expect this ratio to have the same value for all planets in our solar system? Explain your answer.

4 Ganymede and Io are two of Jupiter's moons. The orbital period of Ganymede around Jupiter is 4.1 times that of Io. If Io has an average orbital radius of 4.2×10^8 m, determine the orbital radius of Ganymede.

eBook*plus* RESOURCES

Watch this eLesson: Kepler's Laws
Searchlight ID: eles-2557

3.3 Newton's Law of Universal Gravitation

3.3.1 The Apple and the Moon

With his three laws and Brahe's observational data, Kepler was able to accurately calculate how the solar system was arranged and how it moved. However, he could not say why it moved this way. This understanding would have to wait another 80 years for the work of Isaac Newton.

Isaac Newton is one of the most important people in the history of science, a genius whose influence has been felt in many varied fields within physics. In 1687, he published an enormously important work titled *Philosophiae Naturalis Principia Mathematica*, which contains theories on the motion of objects, as well as a new type of mathematics — calculus — needed to analyse these motions.

In this work, Newton described how his observation of an apple that had been pulled from the tree by the attractive force of the Earth led him to consider whether the same effect (which he described as 'action at a distance') would be observed if the apple were falling from a very great height — as high as the Moon, say.

Newton had previously developed expressions for the inward acceleration of objects that had uniform circular motion. These expressions will be familiar to you from your earlier study of centripetal motion, specifically:

$$a_c = \frac{v^2}{r} \text{ and } a_c = \frac{4\pi r}{T^2}$$

where a_c is the centripetal acceleration of an object moving uniformly in a circle with radius r with an orbital speed v, and having an orbital period of T.

Newton theorised that the same agency that pulled the apple to the Earth also provided the centripetal acceleration that pulled the Moon into its circular path. Newton compared the acceleration of the Moon and of the apple using the accepted scientific values of the time:

Radius of the Earth, $\mathrm{R_E} = 6371$ km

Earth–Moon distance, $\mathrm{r} = 60\,\mathrm{R_E}$

Period of the Moon, $\mathrm{T} = 27.3$ days $= 2.36 \times 10^6$ s

$$\mathrm{a_{apple}} = 9.8\ \mathrm{m\,s^{-2}}$$

$$\begin{aligned}\mathrm{a_{Moon}} &= \frac{\mathrm{v^2_{Moon}}}{\mathrm{r}} = \left(\frac{2\pi \mathrm{r}}{\mathrm{T}}\right)^2 \times \frac{1}{\mathrm{r}} = \frac{4\pi^2 \mathrm{r}}{\mathrm{T}^2}\\ &= \frac{4\pi^2 \mathrm{r}}{\mathrm{T}^2} = \frac{4\pi^2\,(60\mathrm{R_E})}{\mathrm{T}^2}\\ &= \frac{4\pi^2\left(60\left(6.371 \times 10^6\ \mathrm{m}\right)\right)}{(2.36 \times 10^6\ \mathrm{s})^2}\\ &= 2.71 \times 10^{-3}\ \mathrm{m\ s^{-2}}\end{aligned}$$

$$\frac{\mathrm{a_{apple}}}{\mathrm{a_{Moon}}} = \frac{9.8\ \mathrm{m\,s^{-2}}}{2.71 \times 10^{-3}\ \mathrm{m\,s^{-2}}} = 3616$$

Then, he considered the relative distances of the apple and the Moon from the Earth's centre:

$$\frac{\mathrm{d_{Moon}}}{\mathrm{d_{apple}}} = \frac{r}{\mathrm{R_E}} = \frac{60\mathrm{R_E}}{\mathrm{R_E}} = 60$$

Given that $\sqrt{3616} = 60.1 \approx 60$, it could be seen that

$$\left(\frac{\mathrm{d_{Moon}}}{\mathrm{d_{apple}}}\right)^2 \approx \left(\frac{\mathrm{a_{apple}}}{\mathrm{a_{Moon}}}\right)$$

As an acceleration is the result of a force being applied and the acceleration is directly proportional to the size of the force (Second Law of Motion), this implied that

$$\left(\frac{\mathrm{d_{Moon}}}{\mathrm{d_{apple}}}\right)^2 \propto \left(\frac{F_{\mathrm{apple}}}{\mathrm{F_{Moon}}}\right)$$

where F_{apple} and F_{Moon} are the attractive gravitational forces that the Earth exerts on the apple and the Moon respectively.

In this way, Newton concluded that the size of the gravitational force that the Earth exerts on any object is inversely proportional to that object's distance (r) from the Earth's centre of mass, that is:

$$F_g \propto \frac{1}{r^2}$$

3.3.2 Newton's Law of Universal Gravitation

While other scientists, such as Edmund Halley and Robert Hooke, had theorised that the force that the Sun exerted on a planet or a comet was inversely proportional to the square of the distance between them, Newton was the first to prove it mathematically and to realise that the force involved was gravity — the same force that pulled all terrestrial objects downwards. Newton combined his deductions from Kepler's laws with his own laws of motion to develop an expression for a law of universal gravitation that would allow the gravitational attraction between any two objects — an apple and a planet, a moon and a planet, a planet and the Sun — to be calculated.

Using his second law of motion, $\boldsymbol{F_{net}} = m\boldsymbol{a}$, Newton reasoned that the gravitational force exerted on a planet by the Sun depended on the mass of the planet. By the third law of motion, the force that the Sun exerted on a planet must be equal in magnitude (but opposite in direction) to the force that the planet exerted on the Sun. By the second law of motion, the magnitude of the force that the planet exerted on the Sun depended on the mass of the Sun.

Combining these ideas with the inverse square law for distance, Newton derived a relationship for the magnitude of the gravitational force acting between the planet and the Sun:

$$F_{Sun\ on\ planet} \propto \frac{\mathrm{m}_{Sun} \times \mathrm{m}_{planet}}{r^2}$$

From this, he developed the general equation for the Law of Universal Gravitation:

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

where G is the universal gravitational constant and m_1 and m_2 are the masses of any two objects.

Note that gravitation is always an attractive force. For this reason, you may sometimes see this equation expressed as

$$\boldsymbol{F_g} = -\mathrm{G}\frac{m_1\, m_2}{\boldsymbol{r}^2}$$

(You will see where this negative sign comes from later, in Section 3.5.1.)

You may also see this equation written as:

$$\boldsymbol{F_g} = \mathrm{G}\frac{M\, m}{\boldsymbol{r}^2}$$

where the gravitational force between a very small mass (m) and a much larger mass (M) is being calculated.

Regardless of the size of the masses involved, the gravitational force is exerted equally on both masses.

The value of G could not be determined at the time because the mass of the Earth was not known. It took another 130 years before Henry Cavendish was able to measure the gravitational attraction between two known masses and calculate the value of G to be 6.674×10^{-11} N m^2 kg^{-2}.

The value of G is very small, which indicates that gravitation is quite a weak force. A large mass is needed to produce a gravitational effect that is easily noticeable.

3.3 SAMPLE PROBLEM 1

Given the following data, determine the magnitude of the gravitational attraction between:

(a) the Earth and the Moon

(b) the Earth and the Sun.

mass of the Earth $= 5.97 \times 10^{24}$ kg

mass of the Moon $= 7.35 \times 10^{22}$ kg

mass of the Sun $= 1.99 \times 10^{30}$ kg

Average Earth– Moon distance $= 3.84 \times 10^{8}$ m

Average Earth– Sun distance $= 1.50 \times 10^{11}$ m

(Note: This distance is referred to as one astronomical unit or 1 AU.)

SOLUTION:

(a) $$F_g = \frac{\mathrm{G}m_\mathrm{E}m_\mathrm{M}}{r^2}$$

$$= \frac{6.67 \times 10^{-11}\,\mathrm{N\,m^2kg^{-2}}(5.97 \times 10^{24}\,\mathrm{kg})\,(7.35 \times 10^{22}\,\mathrm{kg})}{(3.84 \times 10^{8}\,\mathrm{m})^2}$$

$$= 1.98 \times 10^{20}\,\mathrm{N}$$

That is, the magnitude of the gravitational force of attraction between the Earth and the Moon is approximately 1.98×10^{20} N.

(b) $$F_g = \frac{\mathrm{G}m_\mathrm{E}m_\mathrm{S}}{r^2}$$

$$= \frac{6.67 \times 10^{-11}\,\mathrm{N\,m^2kg^{-2}}(5.97 \times 10^{24}\,\mathrm{kg})\,(1.99 \times 10^{30}\,\mathrm{kg})}{(1.50 \times 10^{11}\,\mathrm{m})^2}$$

$$= 3.52 \times 10^{22}\,\mathrm{N}$$

That is, the magnitude of the gravitational force of attraction between the Earth and the Sun is approximately 3.52×10^{22} N, or about 180 times greater than the Earth–Moon attraction.

3.3.3 Universal Gravitation and Kepler's Law of Periods

Newton's law of universal gravitation placed the physics of planetary orbits on a firm mathematical footing, as well as explaining why the planets and moons move as they do. From this law, Kepler's Law of Periods can be derived.

Kepler had originally stated his Law of Periods in the form $\frac{R^3}{T^2} = \mathrm{k}$, but he was not able to determine an expression for the constant k. When Isaac Newton was devising his Law of Universal Gravitation, he found that he was able to derive such an expression. The derivation begins by equating the gravitation and centripetal forces:

$$\boldsymbol{F_g} = \boldsymbol{F_c}$$

$$\mathrm{G}\frac{Mm}{\boldsymbol{r}^2} = \frac{m\boldsymbol{v}^2}{\boldsymbol{r}}$$

where M is the mass of the Sun, m is the mass of the planet, $\boldsymbol{r}$ is the distance between the centres of the planet and the Sun and $\boldsymbol{v}$ is the planet's orbital speed.

As $v = \frac{2\pi r}{T}$, where T is the planet's period,

$$G\frac{Mm}{\boldsymbol{r}^2} = \frac{m}{\boldsymbol{r}}\left(\frac{2\pi r}{T}\right)^2$$

$$\Rightarrow \frac{GMm}{\boldsymbol{r}^2} = \frac{4\pi^2 m\boldsymbol{r}}{T^2}$$

$$\Rightarrow \frac{GM}{4\pi^2} = \frac{\boldsymbol{r}^3}{T^2} = \mathrm{k}$$

Thus, Kepler's constant k depends only on the mass of the Sun and thus has the same value for all planets orbiting the Sun.

Similarly, the Moon and all other satellites orbiting the Earth will have the same value for $\frac{R^3}{T^2}$ although in this case the value will equal $\frac{G\ M_{Earth}}{4\pi^2}$.

3.3 SAMPLE PROBLEM 2

Calculate the value of $\frac{R^3}{T^2}$ for the Moon using the data in Sample Problem 1 and then use that value to calculate the mass of the Earth.

SOLUTION:

Radius of Moon's orbit, $r = 3.84 \times 10^8$ m; period, $T = 2.36 \times 10^6$ s; $G = 6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$; mass of Earth, $M_{Earth} = ?$

$$\begin{aligned} k &= \frac{r^3}{T^2} \\ &= \frac{(3.84 \times 10^8\ \mathrm{m})^3}{(2.36 \times 10^6\ \mathrm{s})^2} \\ &= 1.02 \times 10^{13}\ \mathrm{m^3\,s^{-2}} \end{aligned}$$

$$\text{As } k = \frac{G\ m_{Earth}}{4\pi^2},$$

$$\begin{aligned} m_{Earth} &= \frac{4\pi^2 k}{G} \\ &= \frac{4\pi^2\left(1.02 \times 10^{13}\ \mathrm{m^3\,s^{-2}}\right)}{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}} \\ &= 6.02 \times 10^{24}\ \mathrm{kg} \end{aligned}$$

3.3 Exercise 1

In the following exercises, refer to Table 3.2 for planetary values.

1. Calculate the force due to gravity of the Earth on a 70 kg person standing on the equator.
2. Calculate the force due to gravity by:
   (a) the Earth on the Moon
   (b) the Moon on Earth.
3. Determine the gravitational force of attraction between two 250 g apples that are lying on a desk 1.0 m apart.
4. Determine the gravitational force of attraction between Jupiter (1.90×10^{27} kg) and its moon Ganymede (1.48×10^{23} kg). The orbital radius of Ganymede is 107 million km.
5. Calculate the gravitational force between Ganymede and a 250 g apple lying on Ganymede's surface. The radius of Ganymede is 2631 km.
6. What is the net gravitational force acting on the Moon during a lunar eclipse? (Assume that the centres of the Sun, Moon and Earth are in a straight line.)
7. Two spheres with masses of 8 kg and 6 kg are separated by a distance of 50 cm. At what position would a 2 kg sphere need to be placed between them to experience a net gravitational force of zero? (Ignore the local gravitational field.)
8. Two small space rocks deep in interstellar space are separated by a distance of 1 m. If one of the space rocks has a mass of 200 g while the other has a mass of 150 g, how long will it take for them to be pulled together by gravitational attraction? Ignore the radius of the rocks themselves and assume that they are initially at rest.

3.4 Objects in orbit

3.4.1 Determining orbital speed

Newton's Law of Universal Gravitation can also be used to calculate the average orbital speed (v_o) of a planet around the Sun:

$$F_c = F_g$$

$$\frac{m_{\text{planet}} v_o^2}{r} = \frac{\text{G M}_{\text{Sun}} m_{\text{planet}}}{r^2}$$

Dividing both sides by m_{planet}:

$$\frac{v_o^2}{r} = \frac{\text{G M}_{\text{Sun}}}{r^2}$$

$$\Rightarrow v_o^2 = \frac{\text{G M}_{\text{Sun}}}{r}$$

$$\Rightarrow v_o = \sqrt{\frac{\text{G M}_{\text{Sun}}}{r}}$$

Note that the value of a satellite's orbit depends on:

- the mass of the planet being orbited
- the radius of the orbit. For a satellite orbiting a planet, this is equal to the radius of the planet plus the altitude of the orbit.

Hence, for the case of a satellite orbiting the Earth, the formula becomes:

$$\boldsymbol{v}_o = \sqrt{\frac{\mathrm{G\,M_E}}{\boldsymbol{R}_{\mathrm{Earth}} + \mathrm{altitude}}}$$

where

$\boldsymbol{v}_o$ = orbital velocity (m s^{-1}); $\mathrm{M_E}$ = mass of the Earth = 5.97×10^{24} kg; $\mathrm{R_E}$ = radius of the Earth = 6.38×10^6 m; altitude = height of orbit above the ground (m).

It is clear from this formula that altitude is the only variable that determines the orbital velocity required for a specific orbit. Further, the greater the radius of the orbit, the lower that velocity is.

3.4 SAMPLE PROBLEM 1

Calculate the orbital speeds of three different satellites orbiting the Earth at altitudes of

(a) 250 km
(b) 400 km
(c) 40 000 km.

SOLUTION:

$\mathrm{M_E} = 5.97 \times 10^{24}$ kg; $\mathrm{R_{Earth}} = 6.38 \times 10^6$ m; $\mathrm{G} = 6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}$

(a) At an altitude of 250 km:

$$\begin{aligned} \boldsymbol{v}_0 &= \sqrt{\frac{\mathrm{G\,M_E}}{\mathrm{R_{Earth}} + \mathrm{altitude}}} \\ &= \sqrt{\frac{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}\left(5.97 \times 10^{24}\ \mathrm{kg}\right)}{\left(6.38 \times 10^6\ \mathrm{m}\right) + \left(2.50 \times 10^5\ \mathrm{m}\right)}} \\ &= 7.75 \times 10^3\ \mathrm{m\,s^{-1}} \approx 27\,900\ \mathrm{km\,h^{-1}} \end{aligned}$$

(b) At an altitude of 400 km:

$$\begin{aligned} \boldsymbol{v}_0 &= \sqrt{\frac{\mathrm{G\,M_E}}{\mathrm{R_{Earth}} + \mathrm{altitude}}} \\ &= \sqrt{\frac{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}\left(5.97 \times 10^{24}\ \mathrm{kg}\right)}{\left(6.38 \times 10^6\ \mathrm{m}\right) + \left(4.00 \times 10^5\ \mathrm{m}\right)}} \\ &= 7.66 \times 10^3\ \mathrm{m\,s^{-1}} \approx 27\,600\ \mathrm{km\,h^{-1}} \end{aligned}$$

(c) At an altitude of 40 000 km:

$$\begin{aligned} \boldsymbol{v}_0 &= \sqrt{\frac{\mathrm{G\,M_E}}{\boldsymbol{R}_{\mathrm{Earth}} + \mathrm{altitude}}} \\ &= \sqrt{\frac{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}\left(5.97 \times 10^{24}\ \mathrm{kg}\right)}{\left(6.38 \times 10^6\ \mathrm{m}\right) + \left(4.00 \times 10^7\ \mathrm{m}\right)}} \\ &= 2.93 \times 10^3\ \mathrm{m\ s^{-1}} \approx 10\,550\ \mathrm{km\,h^{-1}} \end{aligned}$$

3.4.2 Elliptical orbits

The preceding theory assumes that we are dealing with circular orbits but that is usually not the case. As we have seen earlier in this chapter, the most common orbital shape is an ellipse, which can be round or elongated with the degree of elongation described by its eccentricity e: $e = \frac{c}{2a}$ where c is the distance between the two foci of the ellipse and a is the semi-major axis (see Figure 3.7).

Most satellites are placed into near-circular orbits, as shown in Table 3.3, but there are a few notable exceptions, which will be discussed later in this section.

Much of the information shown in Table 3.3 can be calculated if the semi-major axis a is known, and this can be determined from the apogee and perigee distances.

TABLE 3.3 A sample of various Earth satellites.

Satellite name	Purpose	Orbit description	Period (min)	Inclination (degrees)	Perigee altitude (km)	Velocity at perigee (km h^{-1})	Apogee altitude (km)	Velocity at apogee (km h^{-1})	Eccentricity
GENESAT	Biological research and amateur radio beacon	Low Earth orbit	93	40	397	27 590	401	27 580	0.0050
USA 197	Military eye-in-the-sky	Polar low Earth orbit	97	97.8	627	27 140	630	27 120	0.0024
IRIDIUM 95	Satellite phone communication	Low Earth orbit	98	86.6	670	27 050	674	27 040	0.0030
NOAA 18	Weather	Polar low Earth orbit	102	98.8	845	26 740	866	26 660	0.0123
RASCOM 1	African communications	Transfer orbit to geostationary position	638	5.4	587	35 650	35 745	5 900	0.9677
MOLNIYA 3– 53	TV and military communications	Elliptical Molniya orbit	718	64.9	1 047	34 570	39 308	5 620	0.9481
NAVSTAR 59	Global Positioning System	High altitude GPS orbit	718	55.2	20 092	13 980	20 273	13 890	0.0045
OPTUS D2	Australian and NZ television communications	Geostationary orbit	1436	0	35 776	11 060	35 798	11 060	0.0003
SKYNET 5B	Military communications	Geostationary orbit	1436	0.1	35 773	11 060	35 810	11 060	0.0005

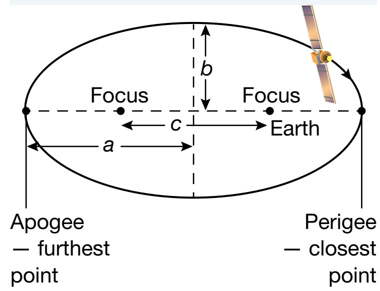
The velocity of a satellite at any point along an elliptical path can be calculated using the following general equation:

$$v = \sqrt{G\,M_{Earth}\left(\frac{2}{r} - \frac{1}{a}\right)}$$

where r is the orbital radius at the point being considered.

The satellite velocities at apogee and perigee in Table 3.3 were calculated using this formula. (You should confirm that for a circular orbit, $a = r$ and that this formula simplifies to the orbital velocity equation.) Notice how the velocities of each satellite are slowest at the apogee and fastest at the perigee, that is, the satellites move quickly when closest to the Earth and slow down as they move further away. This, of course, is just what is described in Kepler's second law.

FIGURE 3.7 The features of the elliptical orbit of a satellite around the Earth.



a = semi-major axis
b = semi-minor axis
c = distance between foci
Eccentricity = $\frac{c}{2a}$

The *Molniya* orbit was developed specifically for this reason. Devised for Russian communications (as most of Russia lies too far north to be satisfactorily covered by a geostationary satellite), this very eccentric orbit places a high apogee over the desired location. A *Molniya* satellite will cruise slowly through this apogee before zipping around through the low perigee and returning quickly to the coverage area.

3.4.3 Types of orbit

Spacecraft or satellites placed into orbit will generally be placed into one of two altitudes — either a ***low Earth orbit*** or a ***geostationary orbit***.

A low Earth orbit is an orbit higher than approximately 250 km, in order to avoid atmospheric drag, and lower than approximately 1000 km, which is the altitude at which the Van Allen radiation belts start to appear. These belts are regions of high radiation trapped by the Earth's magnetic field and pose significant risk to live space travellers as well as to electronic equipment. The space shuttle utilised a low Earth orbit somewhere between 250 km and 400 km depending upon the mission. At 250 km, an orbiting spacecraft has a velocity of 27 900 km h^{-1} and takes just 90 minutes to complete an orbit of the Earth.

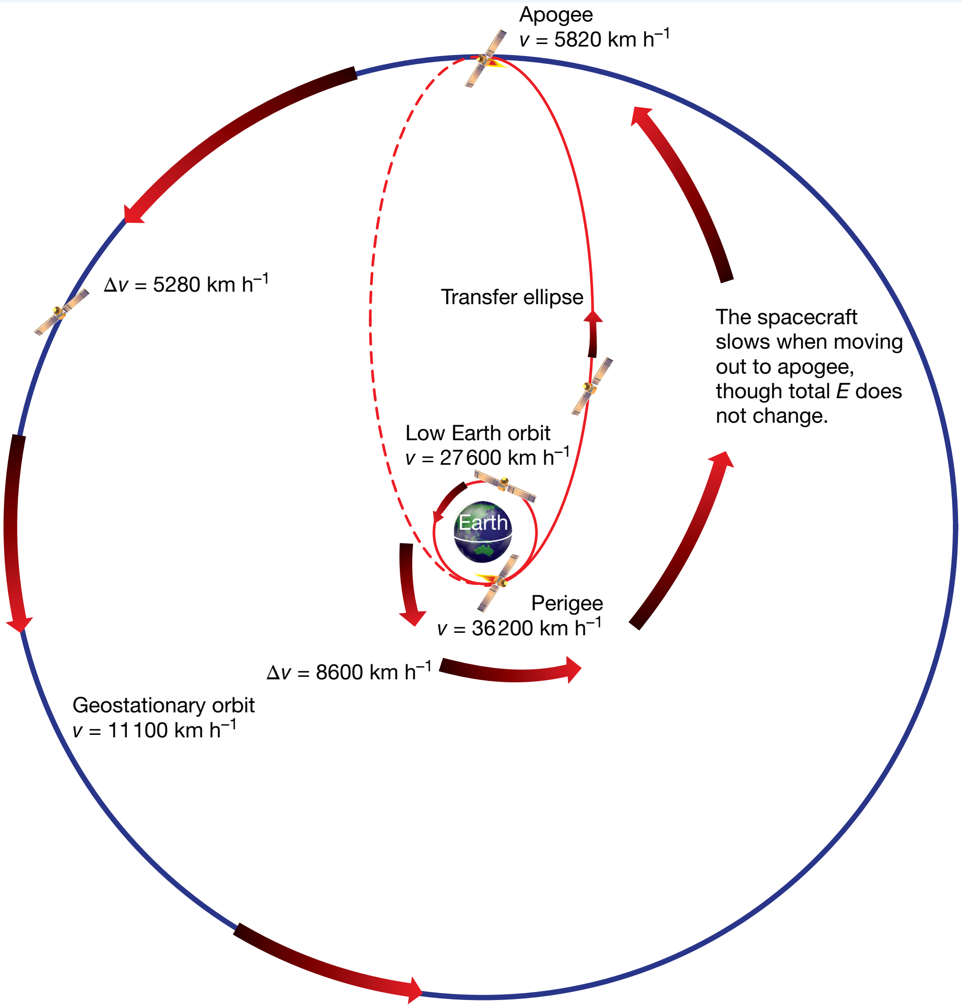
A geostationary orbit is at an altitude at which the period of the orbit precisely matches that of the Earth. If over the equator, such an orbit would allow a satellite to remain 'parked' over a fixed point on the surface of the Earth throughout the day and night. From the Earth, such a satellite appears to be stationary in the sky, always located in the same direction regardless of the time of day. This is particularly useful for communications satellites because a receiving dish need only point to a fixed spot in the sky in order to remain in contact with the satellite.

The altitude of such an orbit can be calculated from Kepler's Law of Periods. Firstly, the period of the orbit must equal the length of one sidereal day; that is, the time it takes the Earth to rotate once on its axis, relative to the stars. This is 3 minutes and 56 seconds less than a 24-hour solar day, so that T is set to be 86 164 s. The radius of the orbit then works out to be 42 168 km, or 6.61 Earth radii. Subtracting the radius of the Earth gives the altitude as approximately 35 800 km. This places the satellite at the upper limits of the Van Allen radiation belts and near the edge of the magnetosphere, making it useful for scientific purposes as well. Australia has the AUSSAT and OPTUS satellites in geostationary orbits.

If a satellite at this height is not positioned over the equator but at some other latitude, it will not remain fixed at one point in the sky. Instead, from the Earth the satellite will appear to trace out a 'figure of eight' path each 24 hours. It still has a period equal to Earth's, however, so this orbit is referred to as geosynchronous.

A **transfer orbit** is a path used to manoeuvre a satellite from one orbit to another. Satellites headed for a geostationary orbit are first placed into a low Earth orbit and then boosted up from there using a transfer orbit, which has a specific orbital energy that lies between that of the lower and higher circular orbits. Orbital manoeuvres utilise Keplerian motion, which is not always intuitive. In order to move a satellite into a different orbit, the satellite's energy must be changed; this is achieved by rapidly altering the kinetic energy. Rockets are fired to change the satellite's velocity by a certain amount, referred to as 'delta-v' (Δv), which will increase or decrease the kinetic energy (and therefore the total energy) to alter the orbit as desired. However, as soon as the satellite begins to change altitude, transformations between the potential energy and the kinetic energy occur, so its speed is continually changing.

FIGURE 3.8 A Hohmann transfer orbit used to raise a satellite from a low Earth orbit of altitude 400 km up to a geostationary orbit of altitude 35 800 km.



The simplest and most fuel-efficient path is a **Hohmann transfer orbit**, as shown in Figure 3.8. This is a transfer ellipse that touches the lower orbit at its perigee and touches the higher orbit at its apogee. The Hohmann transfer involves two relatively quick (called 'impulsive') rocket boosts. In orbital mechanics, the word 'impulsive' describes a quick change in velocity and energy. In order to move to a higher orbit, the first boost increases the satellite's velocity, stretching the circular low Earth orbit out into a transfer ellipse. The perigee is the fastest point on this transfer orbit and, as the satellite moves along the ellipse, it slows again. When it finally reaches the apogee, it will be at the correct altitude for its new orbit, but it will be moving too slowly, the apogee being the slowest point on the ellipse. At this point the rockets are fired again, to increase the velocity to that required for the new higher, stable and circular orbit.

In order to move down from a higher to a lower orbit, the process is reversed, requiring two negative delta-v rocket boosts, that is, retro-firing of the rockets. These two boosts will slow the satellite, hence changing its orbit, first from the higher circular orbit into a transfer ellipse that reaches down to lower altitudes, then from the perigee of the transfer ellipse into a circular low Earth orbit.

3.4 Exercise 1

1 Compare, in words only, low Earth orbits and geostationary orbits.
2 Distinguish between a geostationary orbit and a geosynchronous orbit.
3 Calculate the altitude, period and velocity data to complete the following table.

TABLE 3.4

Type of orbit	Altitude (km)	Period (h)	Velocity ($m\,s^{-1}$)
Low Earth	360		
Geostationary		24	

4 (a) Define the orbital velocity of a satellite.
(b) Describe its relationship to:
i. the gravitational constant G
ii. the mass of the planet it is orbiting
iii. the mass of the satellite
iv. the radius of the orbit
v. the altitude of the satellite.
(c) State this relationship in algebraic form.
5 Calculate the orbital velocity required by the Space Shuttle when at an orbital altitude of 250 km above the surface of the Earth. Assume the mass of the Earth is 5.97×10^{24} kg and its radius is 6380 km.
6 Apollo command modules orbited the Moon at an altitude of 110 km. Calculate the orbital velocity required to do this. Assume that the mass of the Moon is 7.35×10^{22} kg and its radius is 1738 km.
7 Determine the orbital velocity of a GPS satellite orbiting at an altitude of 20 000 km above the Earth.
8 A 5200 kg satellite orbits the Earth at an altitude of 11 000 km. Its velocity is 4800 $m\,s^{-1}$. Determine the gravitational force required to keep this satellite in orbit.

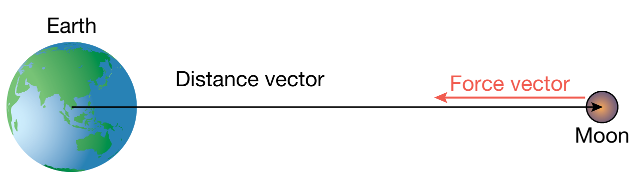
3.5 Gravitational fields

3.5.1 Graphing gravitational force

The gravitational force is an attractive force, whereas the force between electric charges can be either attractive or repulsive.

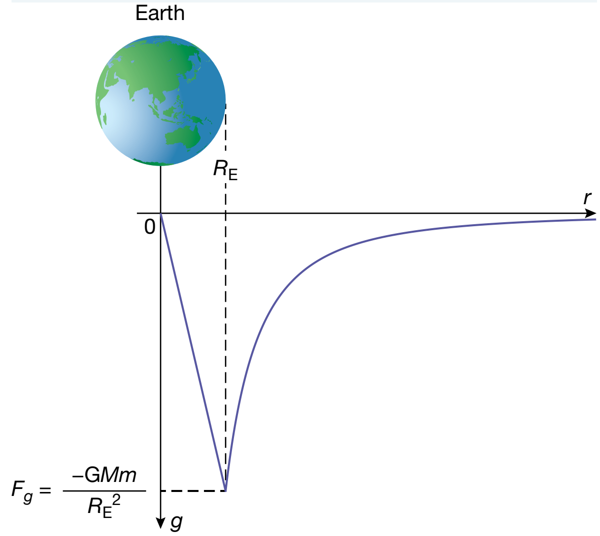
For the force the Earth exerts on the Moon, there is a distance vector from the centre of the Earth to the centre of the Moon, whereas the force vector points in the opposite direction, back to the Earth. For this reason, the gravitational force equation has a negative sign and the force is therefore graphed under the distance axis, as shown in Figure 3.10.

FIGURE 3.9 The Earth exerts gravitational force on the Moon.



The straight blue line in the graph shows how the gravitational force exerted by the Earth on you would decrease if you were to drill down to the centre of the Earth. Newton calculated that if you were inside a hollow sphere, the gravitational force from the mass in the shell would cancel out, no matter where you were inside the sphere. This means that if you were inside the Earth, only the mass in the inner sphere between you and the centre of the Earth would exert a gravitational force on you. This force will get smaller the closer to the centre you go, and at the centre of the Earth the gravitational force will be zero.

FIGURE 3.10 This diagram shows how the Earth's gravitational force varies with distance r from the Earth's centre.

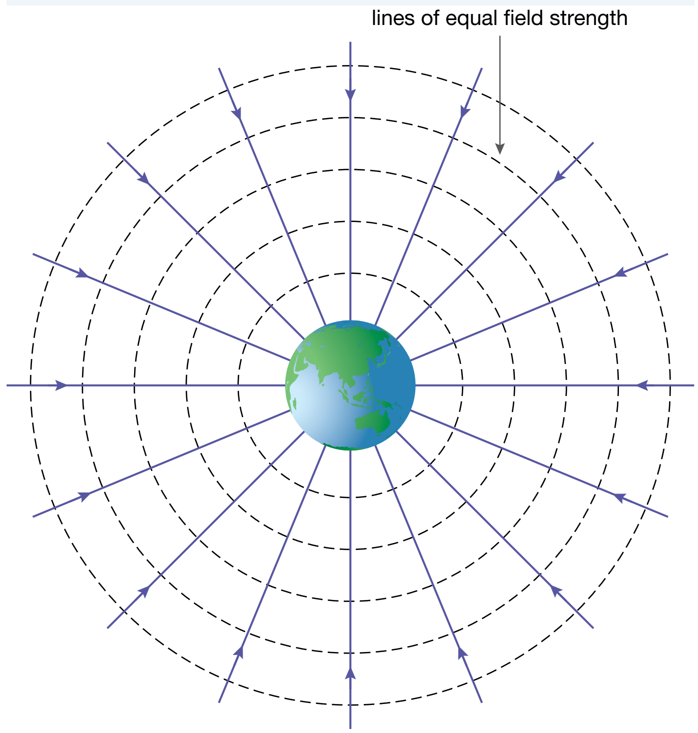


3.5.2 Drawing the gravitational field

Newton's Law of Universal Gravitation describes the force between two masses. However, the solar system has many masses, all attracting one another. The Sun, the heaviest object in the solar system, determines the orbits of all the other masses, but each planet can cause minor variations in the orbital paths of the other planets. Precise calculation of the path of a planet or comet becomes a complicated exercise with many gravitational forces needing to be considered.

Physicists after Newton realised it was easier to determine for each point in space the total force that would be experienced by a unit mass, that is, 1 kilogram, at that point. This idea slowly developed and in 1849 Michael Faraday, in explaining the interactions between electric charges and between magnets, formalised the concept, calling it a 'field'.

FIGURE 3.11 Diagram of the Earth's gravitational field.



A field is more precisely defined as a physical quantity that has a value at each point in space. For example, a weather map showing the pressure across Australia could be described as a diagram of a pressure field. This is an example of a **scalar field**. In contrast, gravitational, electric and magnetic fields are **vector fields**; they give a value to the strength of the field at each point in space, and also a direction for that field at that point. For example, the arrows in the diagram of the Earth's gravitational field show the direction of the field, and the density of the lines (how close together the lines are) indicates the strength of the field.

On a small scale, such as the interior of a room, the field lines — or lines of force — appear parallel and point down since this is the direction of the force that would be experienced by a mass placed within the field.

Of course, any large object near the Earth, such as the Moon, will have a gravitational field of its own, and the two fields will combine to form a more complex field, such as that shown in Figure 3.13. Note that there is a point between the two, but somewhat closer to the Moon, at which the strength of the field is zero. In other words, the gravitational attraction of the Earth and that of the Moon are precisely equal but opposite in direction. Such points exist between any two masses but become noticeable when considering planets and stars that are close enough to be gravitationally bound together.

FIGURE 3.12 The gravitational field within a room on Earth.

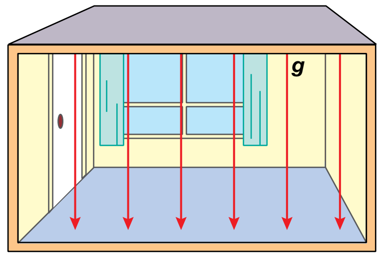
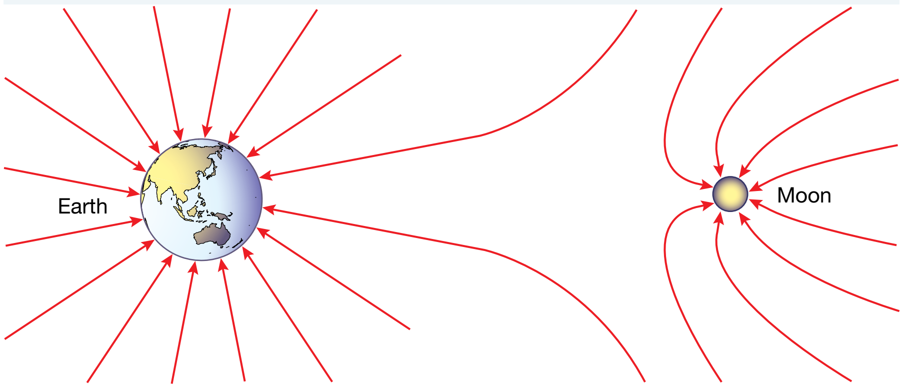


FIGURE 3.13 The gravitational field around the Earth and Moon. The overall shape depends upon the relative strengths of the two fields involved.



3.5.3 Calculating gravitational field strength

A value for the strength of the gravitational field around a mass M can be determined from the value of the force on a unit mass in the field. If the mass m in Newton's Universal Law of Gravitation equation is assigned a value of 1 kg, then the force expression will give the strength of the gravitational field:

$$\boldsymbol{g} = -\frac{GM}{r^2}$$

The unit of gravitational field strength is Newtons per kilogram, $N\,kg^{-1}$.

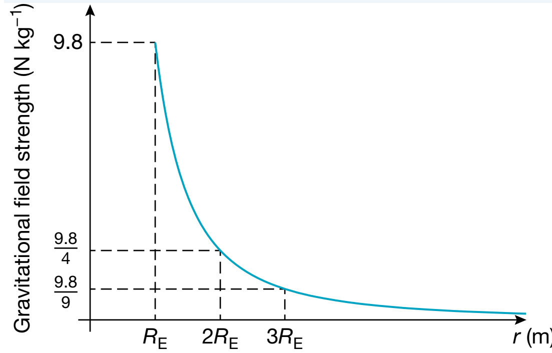
The strength of the gravitational field at the Earth's surface can be calculated using the values for the mass and radius of the Earth from Table 3.2:

$$\begin{aligned}\boldsymbol{g} &= -\frac{6.67 \times 10^{-11}\,N\,m^2\,kg^{-2}\left(5.97 \times 10^{24}kg\right)}{(6.37 \times 10^6\,m)^2} \\ &= -9.80\,N\,kg^{-1}\end{aligned}$$

This is the acceleration due to gravity at the Earth's surface.

The value of the Earth's radius used here, 6380 km (at the equator), is an average value, so the value of $\boldsymbol{g}$ calculated, 9.80 N kg^{-1}, also represents an average value. In fact, due to the spin of the Earth, as well as the Earth being slightly flattened at the poles, $\boldsymbol{g}$ varies from a minimum value at the equator of 9.782 N kg^{-1} to a maximum value of 9.832 N kg^{-1} at the poles. In addition, local variations in $\boldsymbol{g}$ can occur due to variations in density and thickness of the Earth's crust.

FIGURE 3.14 Graph of the magnitude of the strength of the Earth's gravitational field.



The formula for gravitational field strength indicates that the value of $\boldsymbol{g}$ will also vary with altitude above the Earth's surface. By using a value of $\boldsymbol{r}$ equal to the radius of the Earth (R_E) plus altitude, the values of $\boldsymbol{g}$ at different altitudes as shown in Table 3.5 can easily be calculated using a modified version of the formula as follows:

$$\boldsymbol{g} = -\frac{G M_E}{(R_E + \text{altitude})^2}$$

TABLE 3.5 The variation of $\boldsymbol{g}$ with altitude above Earth's surface.

Altitude (km)	g (N kg^{-1})	Comment
0	9.80	Earth's surface
8.8	9.77	Mount Everest
80	9.54	Arbitrary beginning of space
200	9.21	Mercury capsule orbit altitude
300	8.94	Typical space shuttle orbit altitude
40 000	0.19	Communications satellite orbit altitude

It is clear from Table 3.5 that the effect of the Earth's gravitational field is felt quite some distance out into space.

3.5 SAMPLE PROBLEM 1

Determine the gravitational field strength experienced by an astronaut if she were
(a) on the surface of Venus (The mass of Venus is 4.869×10^{24} kg and its radius is 6052 km)
(b) in orbit around Saturn's largest moon, Titan, at an altitude of 150 km (Titan's mass is 1.35×10^{23} kg and its radius is 2575 km).

SOLUTION:

(a) $$\boldsymbol{g} = -\frac{G M_{Venus}}{R_{Venus}}$$
$$= -\frac{6.67 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2}(4.869 \times 10^{24}\ \text{kg})}{(6.052 \times 10^6\ \text{m})^2}$$
$$= -8.87\ \text{N kg}^{-1}$$

The gravitational field strength would be 8.87 N kg^{-1} directed downwards towards the centre of Venus.

(b) $\boldsymbol{g} = -\frac{G\,M_{Titan}}{(R_{Titan} + \text{altitude})}$

$$= -\frac{6.67 \times 10^{-11}\,\text{N m}^2\,\text{kg}^{-2}(1.35 \times 10^{23}\,\text{kg})}{(2.575 \times 10^6\,\text{m} + 1.50 \times 10^5\,\text{m})^2}$$

$$= -1.21\,\text{N kg}^{-1}$$

The gravitational field strength would be $1.21\,\text{N kg}^{-1}$ directed downwards towards the centre of Titan.

At the time Newton developed his Law of Universal Gravitation, he knew it did not provide an explanation for how gravity works, that is, how 'action at a distance' was achieved.

It is inconceivable … that Gravity should be innate, inherent and essential to Matter, so that one body may act upon another at a distance thro' a Vacuum, without the Mediation of any thing else … is to me so great an Absurdity that I believe no Man who has in philosophical Matters a competent Faculty of thinking can ever fall into it. Gravity must be caused by an Agent acting constantly according to certain laws; but whether this Agent be material or immaterial, I have left to the Consideration of my readers. Newton, 1692

The concept of a field now provides an explanation for action at a distance.

3.5 Exercise 1

1. Draw the gravitational field resulting when two spheres of equal mass are placed near each other.
2. How would the gravitational field diagram for the Earth differ from that of Mars?
3. What is the value of the Earth's gravitational field strength:
   (a) in the centre of the Earth
   (b) at a distance $\frac{R_E}{2}$ from the Earth's centre (Hint: refer to Figure 3.10)
   (c) at an altitude of 4000 m above the Earth's surface?
4. What would be the effect on the gravitational field strength experienced on the Earth's surface if the Earth's radius were tripled but the mass remained constant?
5. A 2 kg object is dropped from a height of 10 m on the surface of the Earth. How much longer would the same object take to fall if dropped from the same height on the surface of the Moon? Assume the mass of the Moon = 7.35×10^{22} kg and the radius of the Moon = 1740 km.

3.6 Energy in a gravitational field

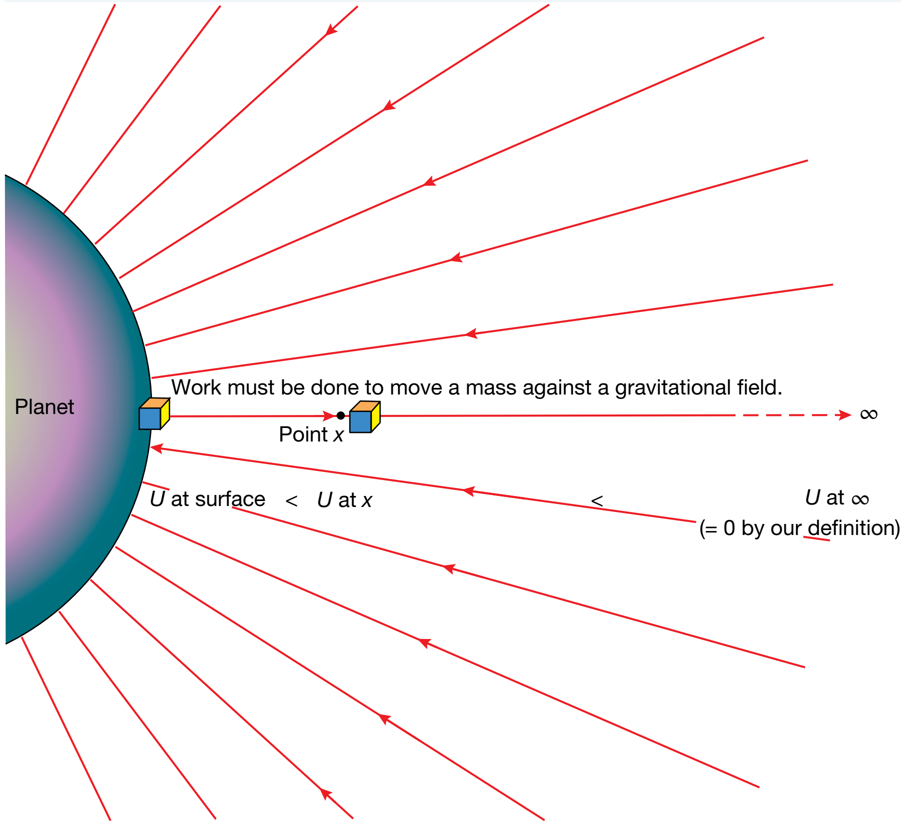
3.6.1 Potential energy in a gravitational field

Gravitational potential energy, U, is the energy of a mass due to its position within a gravitational field. Here on Earth, the U of an object at a height h above the ground is easily found as it is equal to the work done to move the object from the ground:

$$\begin{aligned} U &= \text{work done to move to the point} \\ &= \text{force required} \times \text{distance moved} \\ &= (m\boldsymbol{g}) \times h \\ &= m\boldsymbol{g}\,h \end{aligned}$$

Hence, in this case, $U = m\boldsymbol{g}h$. We chose the ground as our starting point because this is our defined zero level, that is, the place where $U = 0$. Note that since work must be done on the object to lift it, it acquires energy. Hence, at height h, U is greater than zero.

FIGURE 3.15 Different levels of U. If we choose a planet's surface as the zero level, U_x has a positive value. If infinity is chosen as the zero level, U has a negative value.



On a larger, planetary scale, we need to rethink our approach. Due to the inverse square relationship in the Law of Universal Gravitation, the force of attraction between a planet and an object will drop to zero only at an infinite distance from the planet. For this reason, we will now choose infinity (or some point a very large distance away) as our level of zero potential energy.

There is a strange side effect of our choice of zero level. Because gravitation is a force of attraction, work must be done on the object to move it from a point X to infinity; that is, against the field so that it gains potential energy. Therefore,

U_∞ (potential energy at infinity) > U_x(potential energy at point X)

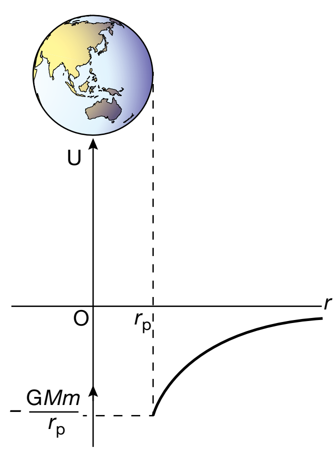
As $U_\infty = 0$, this means that

$0 > U_x$

That is, U_x has a negative value!

Using the same approach as earlier, the gravitational potential energy U of an object at a point X in a gravitational field is equal to the work done to move the object from the zero energy level at infinity (or some point very far away) to point X. It can be shown mathematically that:

FIGURE 3.16 A graph showing how the negative value for gravitational potential energy, U, increases with distance up to a maximum value of zero.



$$U = -\frac{G M m}{r}$$

where M is the mass of the planet, m is the mass of the object and r is the distance separating the masses.

3.6 SAMPLE PROBLEM 1

Given the following data, determine the gravitational potential energy of:
(a) the Moon within the Earth's gravitational field
(b) the Earth within the Sun's gravitational field.

Mass of the Earth $= 5.97 \times 10^{24}$ kg
Mass of the Moon $= 7.35 \times 10^{22}$ kg
Mass of the Sun $= 1.99 \times 10^{30}$ kg
Earth–Moon distance $= 3.84 \times 10^{8}$ m
Earth–Sun distance $= 1.50 \times 10^{11}$ m

SOLUTION:

(a) $$U = -\frac{G M_E m_M}{r}$$

$$= -\frac{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}(5.97 \times 10^{24}\ \mathrm{kg})(7.35 \times 10^{22}\ \mathrm{kg})}{(3.84 \times 10^{8}\ \mathrm{m})}$$

$$= -7.62 \times 10^{28}\ \mathrm{J}$$

That is, the gravitational potential energy of the Moon is approximately -7.62×10^{28} J. Put another way, the work that would be done in moving the Moon from a very large distance away from Earth to its current distance would be -7.62×10^{28} J. The negative sign indicates that this would be work done by the system (not on the system) in moving the Moon from a very large distance away from the Earth to its present orbital distance. This negative work represents potential energy lost by the system as the Moon and the Earth are brought together (converted into other forms of energy, most probably kinetic). Since the U is reduced below the zero level (see Figure 3.16), it is quite appropriate that it should appear as a negative value.

(b) $$U = -\frac{G M_S m_E}{r}$$

$$= -\frac{6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}(1.99 \times 10^{30}\ \mathrm{kg})(5.97 \times 10^{24}\ \mathrm{kg})}{(1.50 \times 10^{11}\ \mathrm{m})}$$

$$= -5.28 \times 10^{33}\ \mathrm{J}$$

That is, the gravitational potential energy of the Earth is approximately -5.28×10^{33} J.

3.6.2 Kinetic energy of an orbiting object

As we have seen in Section 3.4.1, the orbital velocity v_o of an object as it moves around a mass M at a distance r from the mass's centre can be described by the equation:

$$v_o = \sqrt{\frac{GM}{r}}$$

This means that we can develop an equation describing the kinetic energy of an orbiting object as follows:

$$E_k = \frac{1}{2} m v_o^2 \text{ (where m is the mass of the orbiting object)}$$

$$= \frac{1}{2} m \times \left(\sqrt{\frac{GM}{\boldsymbol{r}}}\right)^2$$

$$= \frac{1}{2} m \times \frac{GM}{\boldsymbol{r}}$$

So, $E_k = \dfrac{GMm}{2r}$

Note that the object's kinetic energy is half that of its potential energy but opposite in sign.

3.6 SAMPLE PROBLEM 2

Calculate the kinetic energy of a 1500 kg satellite orbiting the Earth at a distance of 30 000 km from the Earth's centre. Assume the mass of the Earth = 5.97×10^{24} kg.

SOLUTION:

$$E_k = \frac{G M_E m_S}{2r}$$

$$= \frac{6.67 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2}\ (5.97 \times 10^{24}\ \text{kg})\ (1.5 \times 10^3\ \text{kg})}{2 \times (3.00 \times 10^7\ \text{m})}$$

$$= 9.95 \times 10^9\ \text{J}$$

The satellite has a kinetic energy of 9.95×10^9 Joules (or 9.95 GJ)

3.6.3 Orbital energy

Any satellite travelling in a stable circular orbit at a given orbital radius has a characteristic total mechanical energy E. This is the sum of its kinetic energy E_k (due to its orbital velocity) and its gravitational potential energy U (due to its height).

We can derive an expression for the total orbital energy E:

$$E = U + E_k$$

$$= -\frac{GMm}{r} + \frac{GMm}{2r}$$

$$= -\frac{GMm}{2r}$$

This equation looks very similar to the equation for U and also represents a negative energy well. The value of the mechanical energy of a satellite orbiting a planet depends only on the masses involved and the radius of the orbit. A lower orbit produces a more negative value of E and, therefore, less energy, while a higher orbit corresponds to more energy.

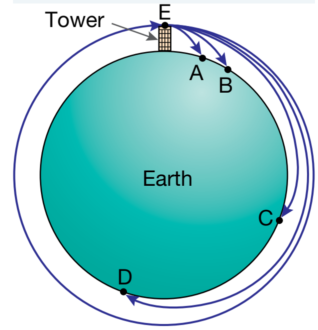
3.6.4 Escape velocity

Isaac Newton wrote that it should be possible to launch a projectile fast enough so that it achieves an orbit around the Earth. His reasoning was that a stone thrown from a tall tower will cover a considerable range before striking the ground. If it is thrown faster, it will travel further before stopping. If thrown faster still, it will have an even greater range. If thrown fast enough then, as the stone falls, the Earth's surface curves away, so that the falling stone never actually lands on the ground and, instead, orbits the Earth.

It was only a thought experiment, of course. He had no way of testing this idea but it does hit upon one important fact — that for any given altitude, there is a specific velocity required for any object to achieve a stable circular orbit.

If this specific velocity is exceeded slightly, then the object will follow an elliptical orbit around the Earth. If the specific velocity is exceeded further still, then the object will follow a parabolic or hyperbolic path away from the Earth. This is the manner in which space probes depart the Earth and head off into space.

FIGURE 3.17 Newton's suggestion for achieving an orbit.



We will now consider a situation similar to Newton's. Imagine throwing a stone directly up. When thrown, the stone will rise to a certain height before falling back to Earth. If thrown faster, it will rise higher. If thrown fast enough, it should rise up and continue to rise, slowing down but never falling back to Earth, and finally coming to rest only when it has completely escaped the Earth's gravitational field. The initial velocity required to achieve this is known as **escape velocity**.

By considering the kinetic and gravitational potential energy of a projectile, it can be shown mathematically that the escape velocity of a planet depends only upon the universal gravitation constant and the mass and the radius of the planet.

Escape velocity ($\boldsymbol{v}_{esc}$) from a large body of mass M is achieved when the total orbital energy of an object is equal to 0, that is:

$$E_k + U = 0$$

This means that, for an object of mass m,

$$\frac{1}{2}m\boldsymbol{v}_{esc}^2 - \frac{GMm}{r} = 0$$

Cancelling for m and rearranging, we see

$$\boldsymbol{v}_{\text{esc}}^2 = \frac{2GM}{r}$$

$$\Rightarrow \boldsymbol{v}_{\text{esc}} = \sqrt{\frac{2GM}{r}}$$

3.6 SAMPLE PROBLEM 3

Calculate the escape velocity for Earth (assume mass of the Earth $= 5.97 \times 10^{24}$ kg and radius of the Earth $= 6.38 \times 10^6$ m).

SOLUTION:

$$\boldsymbol{v}_{\text{esc}} = \sqrt{\frac{2GM_E}{R_E}}$$

$$= \sqrt{\frac{2 \times 6.67 \times 10^{-11}\,\text{N m}^2\text{kg}^{-2}\left(5.97 \times 10^{24}\text{kg}\right)}{\left(6.38 \times 10^6\text{m}\right)}}$$

$$= 1.12 \times 10^4\ \text{m s}^{-1} \approx 40\,000\ \text{km h}^{-1}$$

That is, the escape velocity on Earth is about 40 000 km h^{-1}. This is a considerable velocity, but remember that this is the velocity with which a projectile must be launched directly up in order to completely escape the Earth's gravitational field. It does not apply to a rocket, which continues its thrust well after launch.

3.6 Exercise 1

1 Determine the escape velocity of the planet Venus, given that its mass is 4.87×10^{24} kg and its radius is 6052 km.

2 Use the following data to calculate the gravitational potential energy of
(a) Callisto as it orbits within Jupiter's gravitational field, and
(b) Jupiter as it orbits within the Sun's gravitational field.

Mass of Jupiter = 1.90×10^{27} kg

Mass of Callisto = 1.08×10^{23} kg

Mass of the Sun = 1.99×10^{30} kg

Jupiter–Callisto distance = 1.88×10^{9} m

Jupiter–Sun distance = 7.78×10^{11} m

3 If the mass of the Earth were somehow changed to four times its real value, how would the value of the escape velocity change? Justify your response algebraically.

4 The orbit of Earth around the Sun has a very low eccentricity but is not a perfect circle as has been assumed for simplicity in this section. The Earth's distance from the Sun ranges from 1.47×10^{8} km at its closest approach to 1.52×10^{8} km at its furthest. Determine the corresponding differences in (a) gravitational potential energy, (b) kinetic energy, and (c) orbital energy.

5 A 100 kg satellite orbits the planet Listeria every 4.0 hours. The satellite experiences a gravitational field strength of 8 N kg^{-1} due to Listeria.
(a) What is the radius of the satellite's orbit from Listeria's centre of mass?
(b) Determine the mass of Listeria.
(c) Calculate the satellite's
  i. kinetic energy, and
  ii. gravitational potential energy.

eBook*plus* RESOURCES

Try out this interactivity: One giant leap
Searchlight ID: int-6611

Complete this digital doc: Investigation: Using a pendulum to determine **g**
Searchlight ID: doc-26319

3.7 Review

3.7.1 Summary

- Johannes Kepler deduced three laws that described the motion of the planets using the heliocentric model. The first law describes the elliptical orbits of the planets. The second law describes the speed of the planets along their orbits. The third law relates the period of a planet's orbit (T) to the average radius (r) of the orbit: $\frac{r^3}{T^2} = \text{k}$ where k is a constant for all objects orbiting the same body.

- The Law of Universal Gravitation, proposed by Isaac Newton, describes the force of attraction between any two masses. This force also explains why the solar system behaves as described by Kepler's laws. The law is described by the equation $\boldsymbol{F}_g = \frac{\text{G}Mm}{\boldsymbol{r}^2}$ where $\text{G} = 6.67 \times 10^{-11}\,\text{N m kg}^{-2}$, M and m are masses in kilograms and $\boldsymbol{r}$ is the separation between the masses in metres.
- The acceleration due to gravity, $\boldsymbol{g}$, describes the strength of a gravitational field at a given point around a planetary body and can be expressed in m s^{-2}or N kg^{-1}. It depends upon the mass and radius of the planet orbited as well as the altitude of the point: $\boldsymbol{g} = \frac{\text{G}M}{(r + \text{altitude})^2}$.
- Orbits are generally elliptical where the elongation of the orbit is described by its eccentricity. The simplest orbit is uniform circular motion where the eccentricity is zero.
- Circular motion requires a centripetal force. Gravitation provides the centripetal force for orbital motion.
- Every circular orbit requires a unique orbital velocity v_o for the orbit to be maintained. The orbital velocity depends on the mass and radius of the planet as well as the altitude of the orbit: $\boldsymbol{v}_o = \sqrt{\frac{\text{G}M}{(r + \text{altitude})}}$.
- The gravitational potential energy U of an object at some point within a gravitational field is equivalent to the work done in moving the object from an infinite distance to that point. It is calculated by using the equation $\text{U} = -\frac{\text{G}Mm}{\boldsymbol{r}}$.
- The kinetic energy of an object of mass m in orbit around a mass$M$ at a distance $\boldsymbol{r}$ is described by the equation $\text{E}_\text{k} = \frac{\text{G}Mm}{2\boldsymbol{r}}$.
- The total mechanical (or orbital) energy of an object of mass m in orbit around a mass M at a distance $\boldsymbol{r}$ is described by the equation $\text{E} = -\frac{\text{G}Mm}{2\boldsymbol{r}}$.
- Escape velocity $\boldsymbol{v}_\text{esc}$ is the vertical velocity that a projectile would need to just escape the gravitational field of a planet. It is given by the equation $\boldsymbol{v}_\text{esc} = \sqrt{\frac{2\text{G}M}{\boldsymbol{r}}}$.

3.7.2 Questions

1. What would happen to the strength of the force of attraction between two masses if the distance between them was to halve while the masses themselves were each to double?
2. Draw the gravitational field:
   (a) as it exists within the room in which you are currently sitting
   (b) as it exists around the planet Mars.
3. Draw a diagram of a satellite orbiting the Earth, showing all forces acting on the satellite.
4. Compare and contrast, in words only, low Earth orbits and geostationary orbits.
5. In general terms only, describe the variation in $\boldsymbol{g}$ that would be experienced in a spacecraft travelling directly from the planet Mars to its moon, Phobos, 9380 km away.
6. The gravitational field vector $\boldsymbol{g}$ has an average value on the surface of the Earth of $9.8\,\text{N kg}^{-1}$ or m s^{-2}. Show that the two alternative units quoted are equivalent.
7. If a person were to hurl a rock of mass 0.25 kg horizontally from the top of Mt Everest, at an altitude of 8.8 km above sea level, calculate:
   (a) the velocity required by the rock so that it would orbit the Earth and return to the thrower, still waiting on Mt Everest. Note: The mass of the Earth is 5.97×10^{24} kg, and its mean radius is 6380 km.
   (b) how long the thrower would have to wait for the rock's return.
   (c) the magnitude and direction of the centripetal force acting on the rock.

8. Apollo rockets to the Moon would always begin their mission by launching into a low orbit with an approximate altitude of just 180 km. An orbit of this height can decay quite rapidly. However, they were never there for very long before boosting out of orbit towards the Moon. For their orbit calculate:
   (a) the required velocity in km h^{-1}
   (b) the period of the orbit
   (c) the acceleration of the spacecraft
   (d) the centripetal force acting on the spacecraft.

   Use 110 000 kg as the mass of the spacecraft in orbit.
9. Define Newton's Law of Universal Gravitation.
10. (a) List the variables upon which the force of gravity between two masses depends.
    (b) Note that time is not one of the variables, implying that this force acts instantaneously throughout space, faster even than light. Does this seem reasonable to you? Explain your answer.
11. State what would happen to the strength of the force of attraction between two masses if the distance between them was to halve and the masses themselves were each to double.
12. (a) Discuss the reasons why Newton's Law of Universal Gravitation is important to an understanding of the motions of satellites.
    (b) Describe how the use of this law allows calculation of the motion of satellites.
13. Calculate the magnitude of the force of gravitation between a book of mass 1 kg and a pen of mass 50 g lying just 15 cm from the centre of mass of the book.
14. Calculate the force of gravity between a 72.5 kg astronaut and the Earth (mass = 5.97×10^{24} kg and radius 6.378×10^{6} m), using the Law of Universal Gravitation:
    (a) standing on the ground prior to launch
    (b) at an altitude of 285 km after launch.
15. For each of the orbiting bodies shown in Table 3.6, calculate the orbital velocity from the period and then use it to calculate the centripetal force. Also calculate the value of the gravitational force acting on the body and indicate how well the two forces compare.

TABLE 3.6

Orbiting body	Oribital period	Central body	Oribital radius
Satellite in low Earth orbit $m = 1360$ kg	90.6 minutes	Earth $m = 5.97 \times 10^{24}$ kg	Altitude = 300 km $\therefore r = 6.68 \times 10^{6}$ m
Venus $m = 4.9 \times 10^{24}$ kg	225 Earth days	The sun $m = 1.99 \times 10^{30}$ kg	1.09×10^{11} m
Callisto $m = 1.1 \times 10^{23}$ kg	16.7 Earth days	Jupiter $m = 1.90 \times 10^{27}$ kg	1.88×10^{9} m

16. The moon of Pluto, Charon (pronounced Kair-on), discovered in 1978, is one of the largest moons, in proportion to its planet or dwarf planet (as Pluto is), in the solar system.
    (a) The mass of Charon is 1.62×10^{21} kg while the mass of Pluto is 1.31×10^{22} kg. Calculate the ratio of the mass of Charon to the mass of Pluto.
    (b) The radius of Charon is 593 km and the radius of Pluto is 1151 km. Calculate the ratio of the radius of Charon to the radius of Pluto.
    (c) Calculate the ratio of the density of Charon to the density of Pluto.
    (d) Calculate the ratio of $\boldsymbol{g}$ on Charon to $\boldsymbol{g}$ on Pluto.

17. Calculate the escape velocity of the following planets, using the data shown in Table 3.7.

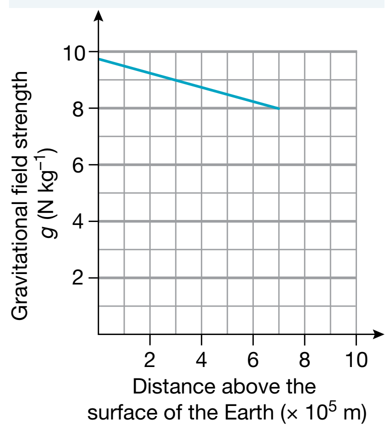
TABLE 3.7

Body	Mass (kg)	Radius (km)	Escape velocity ($m\,s^{-1}$)
Mercury	3.3×10^{23}	2410	
Venus	4.9×10^{24}	6052	
Io	8.9×10^{22}	1821	
Callisto	1.1×10^{23}	2400	

18. A gravitational field strength detector is released into the atmosphere and reports back a reading of $9.70\,N\,kg^{-1}$.
    (a) If the detector has a mass of 10 kg, what is the force of gravity acting on it?
    (b) If the detector is to remain stationary at this height, what upwards force must be exerted on the detector?
    (c) How far is the detector from the centre of Earth?
19. A space probe orbits a distance of 5.0×10^5 m from the centre of an undiscovered planet. It experiences a gravitational field strength of $4.3\,N\,kg^{-1}$. What is the mass of the planet?
20. By how much would your reading on bathroom scales change with the Moon on the opposite side of the Earth to you, compared with being above you?
21. How many Earth radii from the centre of the Earth must an object be for the gravitational force by the Earth on the object to equal the gravitational force that would be exerted by the Moon on the object if the object was on the Moon's surface?
22. A space station orbits at a height of 355 km above Earth and completes one orbit every 92 min.
    (a) What is the centripetal acceleration of the space station?
    (b) What gravitational field strength does the space station experience?
    (c) Your answers to (a) and (b) above should be the same. (i) Explain why. (ii) Explain any discrepancy in your answers.
    (d) If the mass of the space station is 1200 tonnes, what is its weight?
    (e) The mass of an astronaut and the special spacesuit he wears when outside the space station is 270 kg. If he is a distance of 10 m from the centre of mass of the space station, what is the force of attraction between the astronaut and the space station?
23. What is the centripetal acceleration of a person standing on Earth's equator due to Earth's rotation about its axis? (Radius of Earth is 6.38×10^6 m.) Would the centripetal acceleration be greater or less for a person standing in New South Wales? Justify your answer.
24. Neutron stars are thought to rotate at about 1 revolution every second. What is the minimum mass for the neutron star so that a mass on the star's surface is in the same situation as a satellite in orbit, that is, the strength of the gravitational field equals the centripetal acceleration at the surface?
25. The Sun orbits the centre of our galaxy, the Milky Way, at a distance of 2.2×10^{20} m from the centre with a period of 2.5×10^8 years. The mass of all the stars inside the Sun's orbit can be considered as being concentrated at the centre of the galaxy. The mass of the Sun is 2.0×10^{30} kg. If all the stars have the same mass as the Sun, how many stars are in the Milky Way?
26. The asteroid 243 Ida was discovered in 1884. The Galileo spacecraft, on its way to Jupiter, visited the asteroid in 1993. The asteroid was the first to be found to have a natural satellite, that is, its own moon, now called Dactyl. Dactyl orbits Ida at a radius of 100 km and with a period of 27 hours. What is the mass of the asteroid?

27. A satellite is in a circular orbit around the Earth with a radius equal to half of the radius of the Moon's orbit. What is the satellite's period expressed as a fraction of the Moon's period about the Earth?
28. A geostationary satellite remains above the same position on the Earth's surface. Once in orbit, the only force acting on the satellite is that of gravity towards the centre of Earth. Why doesn't the satellite fall straight back down to Earth?
29. A new geostationary satellite is to be launched. At what height above the centre of Earth must the satellite orbit?
30. Can a geostationary satellite remain above Sydney? Why or why not?
31. A space shuttle, orbiting Earth once every 93 mins at a height of 400 km above the surface, deploys a new 800 kg satellite that is to orbit a further 200 km away from Earth.
    (a) Use the graph in Figure 3.18 to estimate the work needed to deploy the satellite from the shuttle.
    (b) Use the mass and radius of Earth to assist you in determining the period of the new satellite.
    (c) Show how the period of the new satellite can be determined without knowledge of the mass of Earth.
    (d) If the new satellite was redesigned so that its mass was halved, how would your answers to (a) and (b) change?

FIGURE 3.18



PRACTICAL INVESTIGATIONS

Investigation 3.1 Using a pendulum to determine g

Aim

To determine the rate of acceleration due to gravity using the motion of a pendulum.

Apparatus

retort stand
bosshead and clamp
approximately 1 metre of string
50 g mass carrier or pendulum bob
stopwatch
metre rule

Theory

When a simple pendulum swings with a small angle, the mass on the end performs a good approximation of the back-and-forth motion called simple harmonic motion. The period of the pendulum, that is, the time taken to complete a single full back-and-forth swing, depends upon just two variables: the length of the string and the rate of acceleration due to gravity. The formula for the period is as shown below:

$$T = 2\pi\sqrt{\frac{l}{g}}$$

where

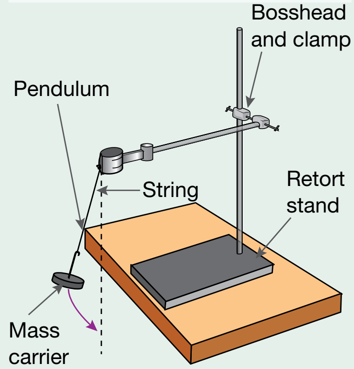
$T =$ period of the pendulum (s)

$l =$ length of the pendulum (m)

$\boldsymbol{g} =$ rate of acceleration due to gravity (m s^{-2}).

Method

FIGURE 3.19 Apparatus for practical activity 3.1.



1. Set up the retort stand and clamp on the edge of a desk as shown in Figure 3.19. Tie on the string and adjust its length to about 90 cm before attaching the 50 g mass carrier or pendulum bob to its end.
2. Using the metre rule, carefully measure the length of the pendulum from the knot at its top to the base of the mass carrier. Enter this length in your results table.
3. Set the pendulum swinging gently (30° maximum deviation from vertical) and use the stopwatch to time 10 complete back-and-forth swings. Be sure to start and stop the stopwatch at an extreme of the motion rather than somewhere in the middle. Enter your time for 10 swings in the results table.
4. Repeat steps 2 and 3 at least five times, after shortening the string by 5 cm each time.

Results

Copy the table below into your practical book to record your results, and then complete the other columns of information.

TABLE 3.8 Results for Investigation 3.1.

Trial	Time for 10 oscillations (s)	Period T (s)	Period squared T^2 (s^2)	Length of pendulum (m)
1				
2				
3				
4				
5				

Draw a graph of period squared versus length of the pendulum. Plot T^2 on the vertical axis and length on the horizontal axis.

Analysis

1. Your graph should display a straight-line relationship. Draw a line of best fit and evaluate the gradient.
2. Rearrange the pendulum equation given earlier to the form $T^2 = kl$, where k is a combination of constants.
3. Compare this formula with the general equation for a straight line: $y = kx$. This comparison shows that if T^2 forms the y-axis and length, l, forms the x-axis, the expression you derived for k in step 2 should correspond to the gradient of the graph you have drawn. Write down your expression:
   gradient = _________________(complete).
4. Use your expression to calculate a value for ***g***, the acceleration due to gravity.

Questions

1. This method usually produces very accurate results. Can you suggest a reason why it should be so reliable?
2. What are the sources of error in this experiment?
3. What could you do to improve the method of this experiment to make it even more accurate?

TOPIC 4
Electric and magnetic fields

4.1 Overview

4.1.1 Module 6: Electromagnetism

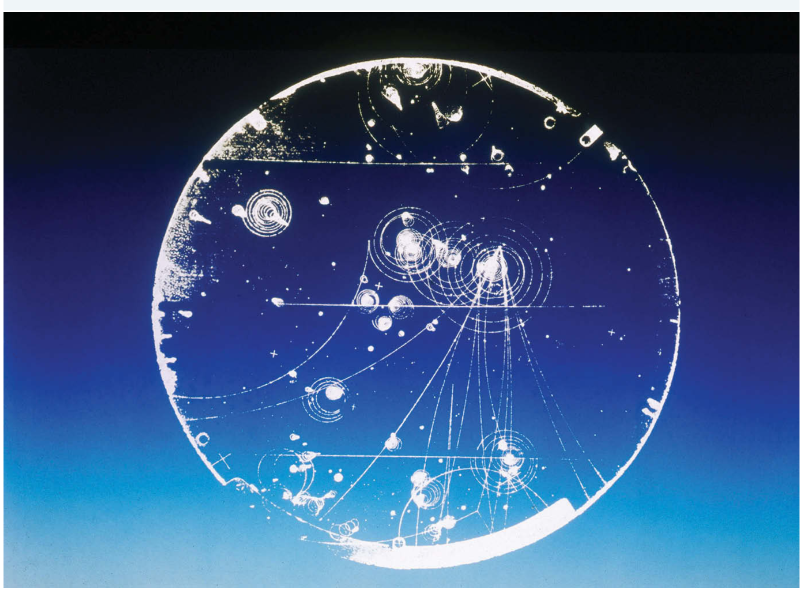
Electric and magnetic fields

Inquiry question: What happens to stationary and moving charged particles when they interact with an electric or magnetic field?

Students:

- investigate and quantitatively derive and analyse the interaction between charged particles and uniform electric fields, including: (ACSPH083)
  - electric field between parallel charged plates: $E = \dfrac{V}{d}$
  - acceleration of charged particles by the electric field: $\boldsymbol{F}_{net} = ma$, $\boldsymbol{F}_{net} = q\boldsymbol{E}$
  - work done on the charge: $W = qv$, $W = qEd$, $K = \dfrac{1}{2}mv^2$
- model qualitatively and quantitatively the trajectories of charged particles in electric fields and compare them with the trajectories of projectiles in a gravitational field
- analyse the interaction between charged particles and uniform magnetic fields, including: (ACSPH083)
  - acceleration, perpendicular to the field, of charged particles
  - the force on the charge ($F = qv_{\perp}B = qvB\sin\theta$)
- compare the interaction of charged particles moving in magnetic fields to:
  - the interaction of charged particles with electric fields
  - other examples of uniform circular motion (ACSPH108)

FIGURE 4.1 The tracks in a bubble chamber show the trajectories of charged particles moving across a magnetic field. The superheated liquid environment of the chamber creates a trail of bubbles behind the moving charges.

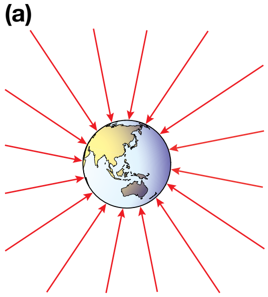
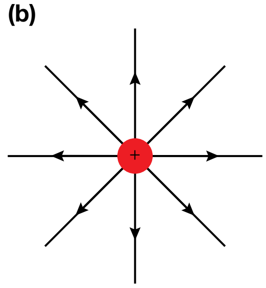
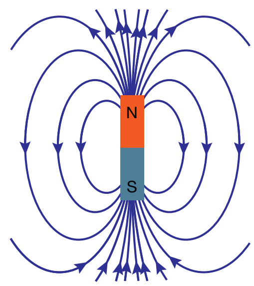


4.2 Charged particles in uniform electric fields

4.2.1 A review of vector fields

This topic is about electric and magnetic fields, but these fields have similarities to gravitational fields (with which we are quite familiar) and so these will be used for comparison. Each of these vector fields can exert a force within a certain space and each has a field vector that indicates the strength and direction of the field at given points. Recall that vector fields can be drawn using lines of force. The direction of a line of force indicates the direction of the field, and the spacing of the lines indicates the strength of the field. The following table briefly compares these three vector fields.

TABLE 4.1 A comparison of three vector fields.

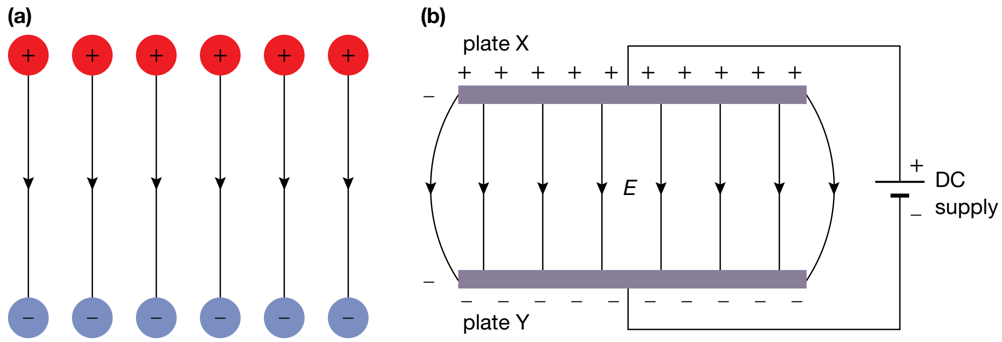
	Gravitational field	Electric field	Magnetic field
Surrounds …	a mass	an electric charge	a magnet or electric current
Diagram: **FIGURE 4.2**	**(a)** 	**(b)** 	**(c)** 
Exerts a force on …	other masses	other charges	other magnets or magnetic materials or moving electric charges
Direction of the field defined by …	the direction of the force on a test mass placed in the field	the direction of the force on a test charge placed in the field	the direction of the force on a test north pole in the field
Field vector	$\boldsymbol{g}$ Units: N kg^{-1} (or m s^{-2})	$\boldsymbol{E}$ Units: N C^{-1} (or V m^{-1})	$\boldsymbol{B}$ Units: Nm A^{-1} (or tesla, T)
A uniform field exists...	inside a room on the Earth's surface	between two parallel charged plates	between the poles of a large horseshoe magnet, or inside a coil carrying an electric current.

4.2.2 Uniform electric fields

If a set of positive and negative charges were lined up in two rows facing each other, the lines of electric field in the space between the rows would be evenly spaced, that is, the value of the strength of the field would be constant. This is called a uniform electrical field.

It is also very easy to set up. Just set two metal plates a few centimetres apart, then connect one plate to the positive terminal of a battery and connect the other plate to the negative terminal of the battery. The battery will transfer electrons from one plate, making it positive, and put them on the other, making that one negative. The battery will keep on doing this until the positive plate is so positive that the battery's voltage, or the energy it gives to each coulomb of electrons, is insufficient to overcome the attraction of the positive charged plate. Similarly, the negatively charged plate will become so negative that the repulsion from this plate prevents further electrons being added.

FIGURE 4.3 Uniform electric fields. (a) A uniform electric field (b) An electric field between two plates.



If a space contains a uniform field, that means that if a charge was placed in that space it would experience a constant electric force, $\boldsymbol{F} = q\boldsymbol{E}$. The direction of the force on a positive charge will be in the direction of the field, and the force on a negative charge will be opposite to the field direction. Also, because the force is constant, the acceleration will be constant, as given by the equation $\boldsymbol{F} = m\boldsymbol{a}$.

As we will see later, the situation with a charged particle in the space between the plates in Figure 4.3b is similar to the vertical motion under gravity. Indeed, if a charged particle is injected with speed into the field from one side, its subsequent motion is similar to projectile motion.

4.2.3 Electric field strength in a uniform field

The emf of a battery, or its voltage, is the amount of energy that the battery gives to each coulomb of charge. A battery of V volts would use up V joules of energy transferring one coulomb of electrons from the top plate through the wires to the bottom plate. Once on the negative plate, this coulomb of electrons would have V joules of electrical potential energy.

If this coulomb of electrons could be released from the negative plate, it would be accelerated by the constant force of the electric field between the plates, gaining kinetic energy like a stone falling in a gravitational field.

This kinetic energy could be found using the standard formula $E_k = \frac{1}{2}mv^2$, but it is simpler to equate the energy gained to the energy required to move the electrons initially.

The gain in kinetic energy of one coulomb of charge = V joules.
The gain in kinetic energy for q coulombs of charge = qV joules.
This is the relationship $W = qV$.

Work done on q coulombs of charge (W) = quantity of charge (q) × voltage drop or potential difference (V)
However, work done (W) also has a definition of motion:

Work done (W) = force $(\boldsymbol{F})$ × displacement (d)

$$W = \boldsymbol{F}d$$

But the force, if it is an electrical force, is given by $\boldsymbol{F} = q\boldsymbol{E}$, so $W = q\boldsymbol{E} \times d$, where d in this instance is the separation of the plates.

Equating the two expressions for work done,

$$q\boldsymbol{E} \times d = q \times V.$$

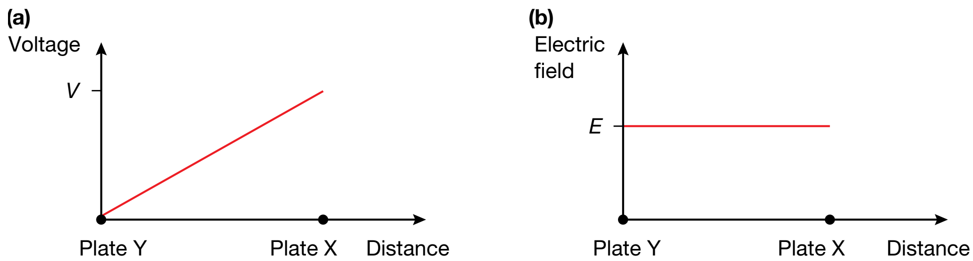
Cancelling the charge, q, gives

$$\boldsymbol{E} = \frac{V}{d}.$$

This provides an alternative unit for electric field of volts per metre or V m^{-1}. So, like gravitational field strength, electric field strength has two equivalent units: either newtons per coulomb or volts per metre. Using volts per metre makes it very easy to determine the strength of a uniform electric field.

Graphing electric field strength

FIGURE 4.4 Electric field strength equals the gradient of the voltage–distance graph.



4.2 SAMPLE PROBLEM 1

What is the strength of the electric field between two plates 5.0 cm apart connected to a 100 V DC supply?

SOLUTION:

$V = 100\text{ V},\ d = 5.0\text{ cm} = 5.0 \times 10^{-2}\text{ m},\ E = ?$

$$\begin{aligned} \boldsymbol{E} &= \frac{V}{d} \\ &= \frac{100\text{ V}}{5.0 \times 10^{-2}\text{ m}} \\ &= 2000\text{ V m}^{-1} \end{aligned}$$

4.2.4 An electric field as a particle accelerator

An electric field can be used to increase the speed and kinetic energy of charged particles. This is the case in all the devices in the following table.

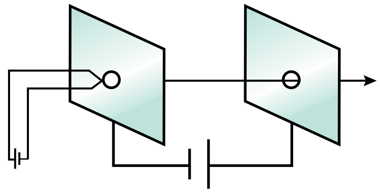
TABLE 4.2 Devices that use electric fields to accelerate charged particles.

Device	Operation	Purpose
Mass spectrometer	Accelerate positive ions of different mass, which then enter a uniform magnetic field and curve around to hit a screen in different spots	To measure the abundance of different elements and isotopes in a sample
Electron microscope	Accelerate electrons, which then pass through electric and magnetic lenses to produce an image	To use an electron beam to examine very small objects
Synchrotron	Accelerate electrons close to the speed of light, then feed them into a storage ring	To produce intense and very narrow beams of mainly X-rays to examine the fine structure of substances such as proteins
Large Hadron Collider	Accelerate protons or lead ions close to the speed of light, then let them collide	To test the predictions of theories of particle physics, e.g. the existence of the Higgs boson

The first part of all these devices is an electron gun, a device that is designed to produce electrons and then give them an initial acceleration.

The diagram shows two metal plates with a small hole cut in the middle of each plate. The plates have been connected to a DC power supply. In the hole of the negative plate is a filament of wire, like the filament in

FIGURE 4.5 The electrons on the hot filament are attracted across to the positive plate and pass through the hole that is in line with the beam.



an incandescent light globe, connected to a low voltage. When the current flows in this circuit, the filament glows red hot. The electrons are, in a sense, 'boiling at the surface' of the filament. The electric field can easily pull the electrons off the surface of the filament.

The hole in the positive plate is in a direct line with the filament, so as the electrons are accelerated across the space between the plates, they go straight through the hole to the next part of the machine. This design is called an electron gun.

4.2 SAMPLE PROBLEM 2

An electron is accelerated from one plate to another. The voltage drop between the plates is 100 V.

(a) How much energy does the electron gain as it moves from the negative plate to the positive plate?

(b) How fast will the electron be travelling when it hits the positive plate, if it left the negative plate with zero velocity?

Use mass of electron = 9.1×10^{-31} kg, charge on electron = 1.6×10^{-19} C.

SOLUTION:

(a)
$$\begin{aligned} W &= Vq \\ &= 100\ \text{V} \times 1.6 \times 10^{-19}\ \text{C} \\ &= 1.6 \times 10^{-17}\ \text{J} \end{aligned}$$

Energy gained is 1.6×10^{-17} J.

(b) Energy is gained as kinetic energy.

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

$$1.6 \times 10^{-17}\ \text{J} = \frac{1}{2} \times 9.1 \times 10^{-31}\ \text{kg} \times v^2$$

$$v^2 = \frac{2 \times 1.6 \times 10^{-17}\ \text{J}}{9.1 \times 10^{-31}\ \text{kg}}$$

$$v = 5.9 \times 10^{6}\ \text{m s}^{-1}$$

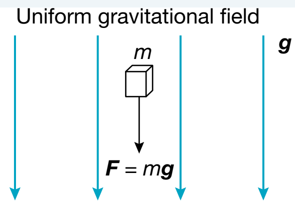
The speed of the electron is 5.9×10^6 m s^{-1} or 5900 km s^{-1}, which is about 2% of the speed of light.

4.2.5 Comparing trajectories

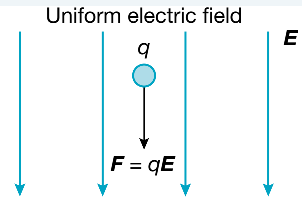
Compare what happens to a mass in a gravitational field to what happens to a charge in an electric field, as shown in Table 4.3 below.

TABLE 4.3 Comparing trajectories within vector fields.

FIGURE 4.6 (a) A mass m placed at a point in a gravitational field $\boldsymbol{g}$ will experience a force $\boldsymbol{F} = m\boldsymbol{g}$. As a result, the mass will accelerate uniformly along the field (that is, parallel to the field lines and in the direction the field lines are pointing) in accordance with Newton's Second Law; hence, $\boldsymbol{a} = \dfrac{\boldsymbol{F}}{m}$.



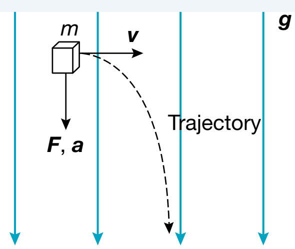
(b) A charge q with mass m placed at a point in an electric field $\boldsymbol{E}$ will experience a force $\boldsymbol{F} = q\boldsymbol{E}$. As a result, the charge will accelerate uniformly along the field, also in accordance with Newton's Second Law, hence $\boldsymbol{a} = \dfrac{\boldsymbol{F}}{m}$.



In both cases, the object will accelerate uniformly along a straight path, acquiring a uniformly increasing velocity $\boldsymbol{v} = \boldsymbol{a}t$ and kinetic energy $E_k = \frac{1}{2}m\boldsymbol{v}^2$.

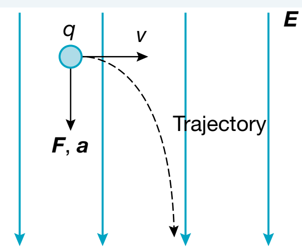
(c) A mass m travelling with velocity $\boldsymbol{v}$ across a gravitational field $\boldsymbol{g}$ will experience a parabolic trajectory, because the only force acting is directed along the field lines ('down'). Hence, its motion has two components:
1. Uniform velocity across the field
2. Uniform acceleration along the field.

The combination of these two components produces the parabolic trajectory of a projectile.



(d) A positive charge q travelling with velocity $\boldsymbol{v}$ across an electric field $\boldsymbol{E}$ will experience a parabolic trajectory, because the only force acting is directed along the field lines. Hence, its motion has two components:
1. Uniform velocity across the field
2. Uniform acceleration along the field.

The combination of these two components produces a parabolic trajectory very similar to projectile motion.



In both cases, the object begins moving across the field but is accelerated down the field, producing a parabolic trajectory.

The following exercises highlight the similarities between these two scenarios. Each will be solved as a projectile motion problem.

4.2 Exercise 1

1 Scenario 1 — a gravitational field exercise
A 0.25 kg ball is projected horizontally at 20 m s^{-1} from the top of a 100 m high cliff. Ignore the effects of air resistance. Data: $\boldsymbol{g} = 9.8\text{ m s}^{-2}$

Begin by drawing a diagram of the scenario. By considering the vertical component of the ball's motion,
(a) determine the time taken for the ball to reach the ground below, and then
(b) calculate the vertical velocity of the ball just prior to hitting the ground
(c) identify the horizontal velocity of the ball at that same point in time
(d) finally, perform a vector addition of the final vertical and horizontal velocities to find the magnitude and direction of the net final velocity of the ball just prior to hitting the ground.

2 Scenario 2 — an electric field exercise
Two horizontal parallel metal plates separated by 5 cm have a voltage of 2.56 V applied between them so that the upper plate is positively charged. A proton is fired horizontally into the space between the plates, just below the upper plate. Assume that the plates are large enough that the proton will remain in that space, and ignore any gravitational effects.

Data: mass of a proton = 1.67×10^{-27} kg; charge of a proton = 1.60×10^{-19} C

As before, draw a diagram of the scenario.
(a) Calculate the electric field strength $\boldsymbol{E}$ that exists between the plates.
(b) Calculate the force that the field applies to the proton and the resulting acceleration of the proton.
By considering the vertical component of the proton's motion,
(c) determine the time taken for the proton to reach the plate below it, and then
(d) calculate the vertical velocity of the proton just prior to hitting the lower plate
(e) Identify the horizontal velocity of the proton at that same point in time
(f) finally, as before, perform a vector addition to find the magnitude and direction of the net final velocity of the proton just prior to hitting the lower plate.

3 A graphing challenge
If you have completed the last two exercises correctly, you will have discovered that the final direction of the trajectories of the ball in scenario 1 and the proton in scenario 2 are the same. A more revealing exercise is to use a spreadsheet application such as Excel to graph the two motions and then compare the graphs side by side.

Begin by setting up three columns for each scenario — one for time, another for horizontal displacement and one for vertical displacement. In each case,

$$\text{horizontal displacement} = \text{initial horizontal velocity} \times \text{time}$$
$$\text{vertical displacement} = 0.5 \times \text{acceleration} \times \text{time}^2$$

Insert the appropriate values to complete the formulae.

For the gravitational scenario, begin with a time value of zero in the first cell, and then increment by 0.1 s for each row until you have reached 4.5 s. Complete the table.

For the electric field scenario, begin with a time value of zero and increment by 0.1 ms for each row until you have reached 4.5 ms. Complete the table.

Finally, for each scenario use a smoothed scatter graph to graph horizontal distance against vertical distance. Your graphs should reflect the shape of the trajectory in each scenario.

4.3 Charged particles in uniform magnetic fields

4.3.1 The nature of the interaction

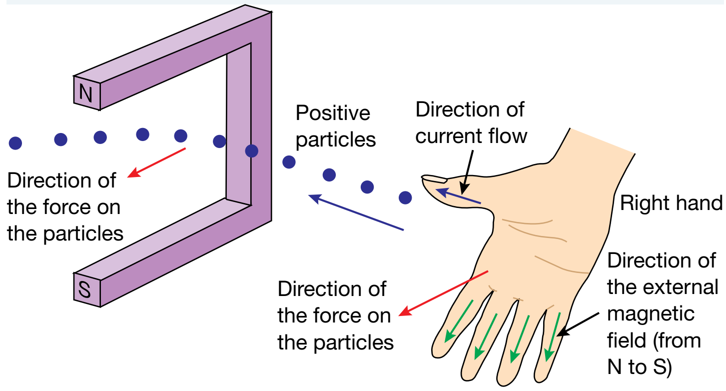
As will be seen in greater depth in Topic 5, moving charged particles such as electrons, protons or alpha particles produce magnetic fields around themselves. If they enter an external magnetic field (caused by another source), the two fields interact.

When the moving charged particles enter the magnetic field at right angles to the field, they experience a force that is at right angles to the velocity and to the direction of the external field. The direction of the force is determined by using the right-hand push rule (also known as the right-hand palm rule) and is demonstrated in Figure 4.7.

To use the right-hand push rule, position your right hand so that:

- the fingers point in the direction of the external field
- the thumb points in the direction of conventional current flow (this means in the direction of the velocity of positive charges or in the opposite direction to the velocity of negative charges)
- the direction of the force on the particles is directly away from the palm of the hand.

FIGURE 4.7 The right-hand push rule for charged particles moving across uniform magnetic fields.



The magnitude of the force experienced by a charge travelling through a magnetic field is given by the following equation:

$$\boldsymbol{F} = q\,\boldsymbol{v}\,\boldsymbol{B}\,\sin\theta$$

where $\boldsymbol{F}$ = the magnitude of the force in newtons, N
q = the charge in coulombs, C
$\boldsymbol{v}$ = the velocity of the charge in m s^{-1}
$\boldsymbol{B}$ = the magnetic field strength in teslas, T
θ = the angle between the velocity and the field lines.

This equation means that the force is a maximum when the velocity of the charge is perpendicular to the magnetic field lines, and reduces to zero when the velocity is parallel to the field lines. The following table shows several possible applications of this equation.

TABLE 4.4 Applying the formula $\boldsymbol{F} = q\boldsymbol{vB}\sin\theta$.

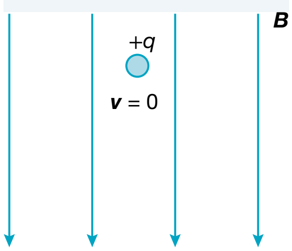
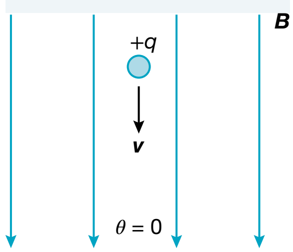
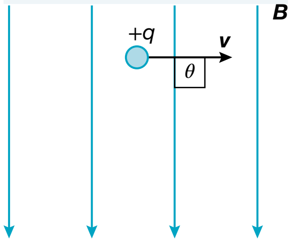
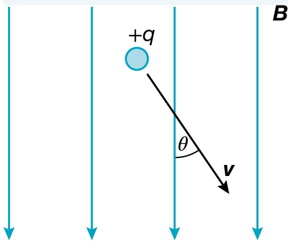
FIGURE 4.8 (a) 	If $\boldsymbol{v} = 0$ then $\boldsymbol{F} = q\boldsymbol{vB}\sin\theta = 0$ A motionless charge in a magnetic field experiences no force.
(b) 	The velocity is parallel to the field lines so that $\theta = 0$, and hence $\sin\theta = 0$. Therefore $\boldsymbol{F} = q\boldsymbol{vB}\sin\theta = 0$. A charge that moves along a magnetic field parallel to the field lines experiences no force.

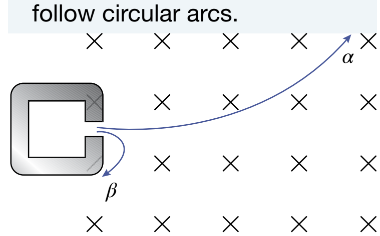
TABLE 4.4 Applying the formula $\boldsymbol{F} = q\boldsymbol{vB} \sin\theta$. *(Continued)*

(c)  	The velocity is perpendicular to the field lines so that $\theta = 90°$, and hence $\sin\theta = 1$. Therefore $\boldsymbol{F} = q\boldsymbol{vB}$ and has its maximum value. The force will be directed into the page.
(d)  	The velocity is at an angle to the field lines so that $0 < \theta < 90°$, and hence $0 < \sin\theta < 1$. Therefore $\boldsymbol{F} = q\boldsymbol{vB} \sin\theta$ and will be directed into the page.

4.3.2 The motion of a charged particle in a magnetic field

The force on a charge moving across a magnetic field is at right angles to its velocity (as well as the magnetic field), which will cause it to change direction. As the velocity direction changes, the force direction will change so that it is perpendicular to the velocity once more. As the velocity continues to change direction, the force direction will continue to change so that it is always perpendicular to the velocity. This situation creates circular motion as this force is acting as a centripetal force, similar to the tension in a rope used to whirl a ball around or the gravitational force that keeps satellites in orbit around a planet.

FIGURE 4.9 Positive alpha particles are deflected up while negative beta particles are deflected down. Both follow circular arcs.



Another example of this motion occurs when alpha and beta radiation particles are released into a magnetic field, as shown in Figure 4.9. An alpha particle is a relatively heavy, positively charged ${}^{4}_{2}He$ ion. If you correctly apply the right-hand push rule as described in Figure 4.7, you should agree that the particle will be pushed upwards along a circular arc. By contrast, a beta particle is a comparatively low mass, negatively charged electron, and so it will be forced downwards along a much tighter circular arc. The smaller mass of the electron causes the smaller radius of its arc.

4.3.3 Calculating the circular motion

It is possible to derive a formula for the radius of the circular motion created when a charged particle q of mass m travelling with velocity $\boldsymbol{v}$ crosses a magnetic field $\boldsymbol{B}$.

The magnitude of the net force on the charged particle as it moves in the magnetic field is:

$\boldsymbol{F}_{net} = m\boldsymbol{a}$.

In this case the only significant force is the magnetic force, $\boldsymbol{F} = q\boldsymbol{vB}$.

$\Rightarrow q\boldsymbol{vB} = m\boldsymbol{a}$

Because the acceleration is centripetal and constant in magnitude, its magnitude can be expressed as $\boldsymbol{a} = \dfrac{\boldsymbol{v}^2}{r}$, where r is the radius of the circular motion.

$$\Rightarrow q\boldsymbol{v}\boldsymbol{B} = \frac{m\boldsymbol{v}^2}{r}$$

The expression for the radius is therefore:

$$r = \frac{m\boldsymbol{v}}{\boldsymbol{B}q}.$$

Hence, a greater mass or velocity will increase the radius, while a greater charge or magnetic field strength will decrease the radius of the circular path of the particle.

4.3 SAMPLE PROBLEM 1

An electron travelling at $5.9 \times 10^6\ \text{m s}^{-1}$ enters a magnetic field of 6.0 mT. What is the radius of its path?

SOLUTION:

$m = 9.1 \times 10^{-31}\ \text{kg}$, $q = 1.6 \times 10^{-19}\ \text{C}$, $\boldsymbol{v} = 5.9 \times 10^6\ \text{m s}^{-1}$, $B = 6.0\ \text{mT}$

$$\begin{aligned} r &= \frac{m\boldsymbol{v}}{Bq} \\ &= \frac{9.1 \times 10^{-31}\ \text{kg} \times 5.9 \times 10^6\ \text{m s}^{-1}}{6.0 \times 10^{-3}\ \text{T} \times 1.6 \times 10^{-19}\ \text{C}} \\ &= 5.6 \times 10^{-3}\text{m} = 5.6\ \text{mm} \end{aligned}$$

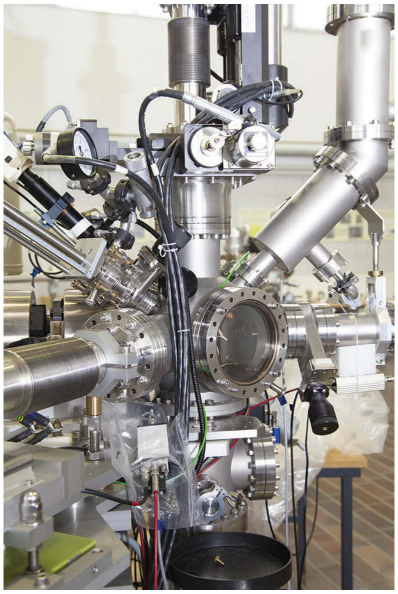
If the velocity $\boldsymbol{v}$, charge q and magnetic field strength $\boldsymbol{B}$ are all known, the radius will allow a calculation of the mass of the particle. The formula can be rearranged to show this:

$$m = \frac{r\boldsymbol{B}q}{\boldsymbol{v}}$$

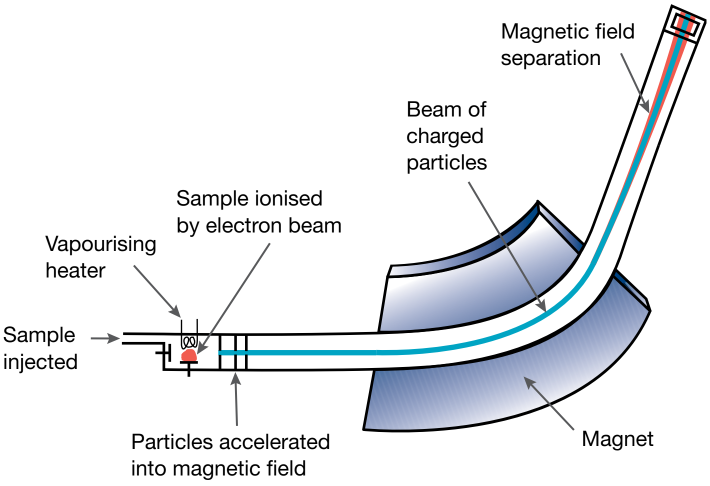
A device that is used to perform this measurement is called a mass spectrometer, shown in Figure 4.10.

FIGURE 4.10 A mass spectrometer is used to determine the mass of atomic particles by ionising a sample particle to give it a charge, passing it through a known magnetic field and measuring the radius of curvature of its path. From this information, the mass can be calculated.

(a)



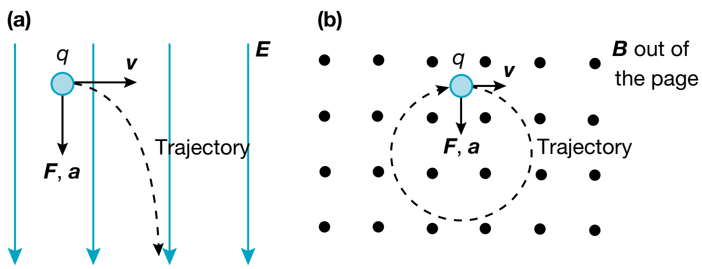
(b)



4.3.4 Comparing the effect of electric and magnetic fields

In this topic we have examined the effect that electric and magnetic fields have on electric charges, and it interesting at this point to compare these two effects. Whereas an electric field exerts a force on a charge that is directed along the field, a magnetic field exerts a force on a charge only if it is already moving across the field, and then the force will be directed perpendicularly to the field and the velocity of the charge. The electric field will produce a parabolic trajectory while the magnetic field produces a circular trajectory.

FIGURE 4.11 A comparison of the effects of electric and magnetic fields on the trajectory of a moving charged particle **(a)** A positive charge q travelling with velocity $\boldsymbol{v}$ across an electric field $\boldsymbol{E}$ will experience a parabolic trajectory. **(b)** A positive charge q travelling with velocity $\boldsymbol{v}$ across a magnetic field $\boldsymbol{B}$ will experience a circular trajectory.



4.3 Exercise 1

1 A stream of electrons with a velocity of 1.2×10^6 m s^{-1} enters a uniform magnetic field of strength 2.6×10^{-3} T. Calculate
(a) the magnitude of the force that the field exerts on each electron
(b) the radius of the path of the electrons.
Data for this exercise

Mass of an electron	9.1×10^{-31} kg
Charge of an electron	1.6×10^{-19} C

2 Calculate the speed of an electron that would move in an arc of radius 5.00 mm in a magnetic field of 5.00 mT.

eBook*plus* RESOURCES

Try out this interactivity: Electrons in magnetic fields
Searchlight ID: int-0124

4.4 Review

4.4.1 Summary

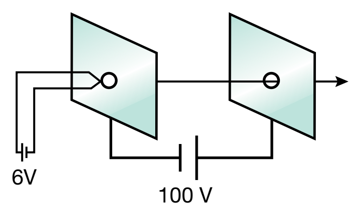
- There are 3 basic vector fields — gravitational fields, electric fields and magnetic fields.
- A uniform electric field can be set up between two charged parallel plates. They can be charged by attaching a DC voltage supply between the plates, so that $E = \dfrac{V}{d}$.
- A uniform magnetic field can exist between the poles of a large horseshoe magnet, or inside of a current-carrying coil, or solenoid.
- An electric charge in an electric field will experience a force $\boldsymbol{F} = q\boldsymbol{E}$ parallel with the field lines. If the charge is positive, the force will be in the direction of the field lines; if the charge is negative, the direction of the force will be opposite to the direction of the field lines.

- As the charge accelerates along the field lines due to $\boldsymbol{F} = m\boldsymbol{a}$, the field does work on the charge, found using $W = qV$ or $W = q\boldsymbol{E}d$. Work is done to increase the particle's kinetic energy $E_k = \frac{1}{2}m\boldsymbol{v}^2$.
- If the charge is initially travelling across the electric field lines, its trajectory will be parabolic, similar to the path followed by a projectile in a gravitational field.
- Electron guns use electric fields to accelerate and eject high-velocity electrons into devices such as mass spectrometers, electron microscopes and cathode ray tubes.
- A charged particle moving across a magnetic field will experience a force $\boldsymbol{F} = qv_{\perp}\boldsymbol{B} = q\boldsymbol{vB}\sin\theta$. The direction of the force is perpendicular to both the field lines and the velocity of the charged particle, and can be predicted using the right-hand push rule.
- This force acts as a centripetal force and causes the charged particle to follow a circular trajectory through the magnetic field.
- The radius of the circular path followed by the charged particle depends upon the mass, charge and velocity of the particle as well as the strength of the magnetic field.

4.4.2 Questions

1. Two metal plates, X and Y, are set up 10 cm apart. The X plate is connected to the positive terminal of a 60 V battery and the Y plate is connected to the negative terminal. A small positively charged sphere is suspended midway between the plates and it experiences a force of 4.0×10^{-3} newtons.
   (a) What would be the size of the force on the sphere if it was placed 7.5 cm from plate X?
   (b) The sphere is placed back in the middle and the plates are moved apart to a separation of 15 cm. What is the size of the force now?
   (c) The plates are returned to a separation of 10 cm but the battery is changed. The force is now 6.0×10^{-3} newtons. What is the voltage of the new battery?
2. Electrons from a hot filament are emitted into the space between two parallel plates and are accelerated across the space between them.

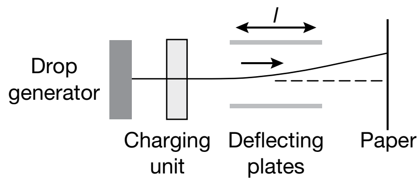
FIGURE 4.12



   (a) Which battery supplies the field to accelerate the electrons?
   (b) How much energy would be gained by an electron in crossing the space between the plates?
   (c) How would your answer to (b) change if the plate separation was halved?
   (d) How would your answer to (b) change if the terminals of the 6 V battery were reversed?
   (e) How would your answer to (b) change if the terminals of the 100 V battery were reversed?
   (f) How would the size of the electric field between the plates, and thus the electric force on the electron, change if the plate separation was halved?
   (g) Explain how your answers to (c) and (f) are connected.
3. (a) Calculate the acceleration of an electron in a uniform electric field of strength 1.0×10^6 N C^{-1}.
   (b) Starting from rest, how long would it take for the speed of the electron to reach 10% of the speed of light? (Ignore relativistic effects.)
   (c) What distance would the electron travel in that time?
   (d) If the answer to (c) was the actual spacing of the plates producing the electric field, what was the voltage drop or potential difference across the plates?

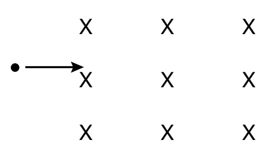
4. In an inkjet printer, small drops of ink are given a controlled charge and fired between two charged plates. The electric field deflects each drop and thus controls where the drop lands on the page.
   Let m = the mass of the drop, q = the charge of the drop, v = the speed of the drop, l = the horizontal length of the plate crossed by the drop, and E = electric field strength.
   (a) Develop an expression for the deflection of the drop. *Hint:* This is like a projectile motion question.
   (b) With the values $m = 1.0 \times 10^{-10}$ kg, $v = 20\,\text{m s}^{-1}$, $l = 1.0$ cm and $E = 1.2 \times 10^{6}\,\text{N C}^{-1}$, calculate the charge required on the drop to produce a deflection of 1.2 mm

FIGURE 4.13



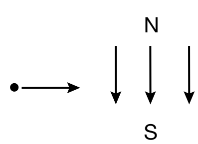
5. An electron travelling east at $1.2 \times 10^{5}\,\text{m s}^{-1}$ enters a region of uniform magnetic field of strength 2.4 T.
   (a) Calculate the size of the magnetic force acting on the electron.
   (b) Describe the path taken by the electron, giving a reason for your answer.
   (c) Calculate the magnitude of the acceleration of the electron.
6. (a) What is the size of the magnetic force on an electron entering a magnetic field of 250 mT at a speed of $5.0 \times 10^{6}\,\text{m s}^{-1}$?
   (b) Use the mass of the electron to determine its centripetal acceleration.
   (c) If a proton entered the same field with the same speed, what would be its centripetal acceleration?
7. Determine the direction of the magnetic force in the following situations, using your preferred hand rule. Use the following terminology in your answers: up the page, down the page, left, right, into the page, out of the page.
   (a) Magnetic field into the page, electron entering from left

FIGURE 4.14



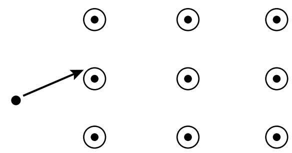
   (b) Magnetic field down the page, electron entering from left

FIGURE 4.15



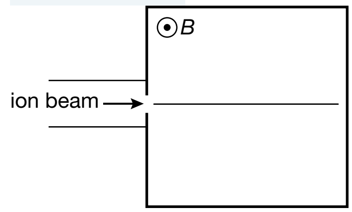
(c) Magnetic field out of the page, proton entering obliquely from left

FIGURE 4.16



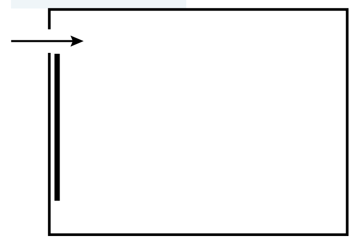
8. An ion beam consisting of three different types of charged particle is directed eastwards into a region having a uniform magnetic field *B* directed out of the page. The particles making up the beam are (i) an electron, (ii) a proton and (iii) a helium nucleus or alpha particle. Copy Figure 4.17 and draw the paths that the electron, proton and helium nucleus could take.

FIGURE 4.17



9. In a mass spectrometer, positively charged ions are curved in a semicircle by a magnetic field to hit a detector at different points depending on the radius and mass. The ions enter the chamber at the top left corner, and curve around to hit the detector (see Figure 4.18). What should be the direction of the magnetic field for the spectrometer to work properly? Use the answers from question 8.

FIGURE 4.18



10. Calculate the radius of curvature of the following particles travelling at 10% of the speed of light in a magnetic field of 4.0 T.
    (a) An electron
    (b) A proton
    (c) A helium nucleus
11. What magnetic field strength would cause an electron travelling at 10% of the speed of light to move in a circle of 10 cm?

eBook*plus* RESOURCES

Complete this digital doc: Investigation: The effect of magnets on cathode rays
Searchlight ID: doc-26586

PRACTICAL INVESTIGATIONS

Investigation 4.1: The effect of magnets on cathode rays

Aim
To observe the trajectory of electrons passing through a magnetic field.

Apparatus
Cathode ray demonstration tube or Crookes tube, complete with power supply
A strong magnet

Theory

This activity will probably be performed as a demonstration by your teacher as the voltages connected to a cathode ray tube are usually very high, of the order of 10 000 V or more. This creates a strong electric field between the electrodes inside the tube. If the electrodes are 10 cm apart and the voltage is 10 000 V, the electric field strength will be

$$
\begin{aligned}
E &= \frac{V}{d} \\
&= \frac{10\,000\,\text{V}}{0.10\,\text{m}} \\
&= 1\,00\,000\,\text{V}\,\text{m}^{-1}
\end{aligned}
$$

The field lines are directed from the positive electrode, or anode, to the negative electrode, or cathode. The gas inside the tube contains some free electrons and positive ions. The electrons accelerate against the field, towards the anode, and collide with other gas molecules, ionising them and thereby creating more free electrons and positive ions. The positive ions accelerate towards the cathode, strike it and knock many more electrons free of the metal surface and into the electric field. The result is a flood of electrons towards the anode, which was originally called cathode rays. The tube should have a gap, or collimator, that will form the electron stream into a beam or ray and some surface that will fluoresce when the ray strikes it.

If a stream of charged particles flows across a magnetic field, each of the particles will experience a centripetal force $\boldsymbol{F} = q\boldsymbol{vB}$. This will result in the particle stream following a circular path.

Method

1. Set up the cathode ray tube and power supply so that the cathode ray can be seen on the fluorescent surface.
2. Place the north pole of a magnet directly in front of the tube with the ray going past it, and note the curve of the cathode ray. Can you confirm this direction with the hand rule? Remember that the conventional current direction is opposite to the direction of the ray … you can compensate for this by using your left hand instead. Note the circular arc formed by the path of the ray.
3. Flip the magnet over so that the south pole is in front of the tube. The ray should now curve in the opposite direction. Can you confirm this direction as well?
4. Try several other positions for the magnet to note the effects of these changes.

Questions

1. Were you able to confirm the direction of the circular arcs followed by the cathode ray when you imposed a magnetic field?
2. The radius of the circular arc is given by the formula $r = \frac{mV}{Bq}$. The mass and charge of the cathode rays cannot be changed, but it should be possible to affect the radius using the other two variables.
   (a) Increasing the voltage V should increase the radius of the path. Can you confirm this?
   (b) Increasing the magnetic field strength B should decrease the radius of the path. Try using stronger (or weaker) magnets in place of the one you used before. Can you confirm this effect?

TOPIC 5 The motor effect

5.1 Overview

5.1.1 Module 6: Electromagnetism

The motor effect

Inquiry question: Under what circumstances is a force produced on a current-carrying conductor in a magnetic field?

Students:

- investigate qualitatively and quantitatively the interaction between a current-carrying conductor and a uniform magnetic field, $\boldsymbol{F} = l\boldsymbol{I}_{\perp}\boldsymbol{B} = l\boldsymbol{IB}\sin\theta$, to establish: (ACSPH080, ACSPH081)
  - conditions under which the maximum force is produced
  - the relationship between the directions of the force, magnetic field strength and current
  - conditions under which no force is produced on the conductor
- conduct a quantitative investigation to demonstrate the interaction between two parallel current-carrying wires
- analyse the interaction between two parallel current-carrying wires, $\frac{\boldsymbol{F}}{l} = \frac{\mu_0}{2\pi} \times \frac{\boldsymbol{I}_1\boldsymbol{I}_2}{r}$, and determine the relationship between the International System of Units (SI) definition of an ampere and Newton's Third Law of Motion (ACSPH081, ACSPH106)

FIGURE 5.1 (a) A conductor is arranged as a swing, in a magnetic field (red is north). A current flows in the conductor and creates a magnetic field that interacts with the permanent magnetic field. (b) What change could have been made to the experiment to make the conductor swing in the opposite direction?

(a)

(b)

5.2 Magnetic fields

5.2.1 Current-carrying conductors and magnetic fields

What would your life be like without electricity? Modern industrialised nations are dependent on electricity. Electricity is easy to produce and distribute, and is easily transformed into other forms of energy. Electric motors are used to transform electricity into useful mechanical energy. They are used in homes, for example in refrigerators, vacuum cleaners and many kitchen appliances, and in industry and transport.

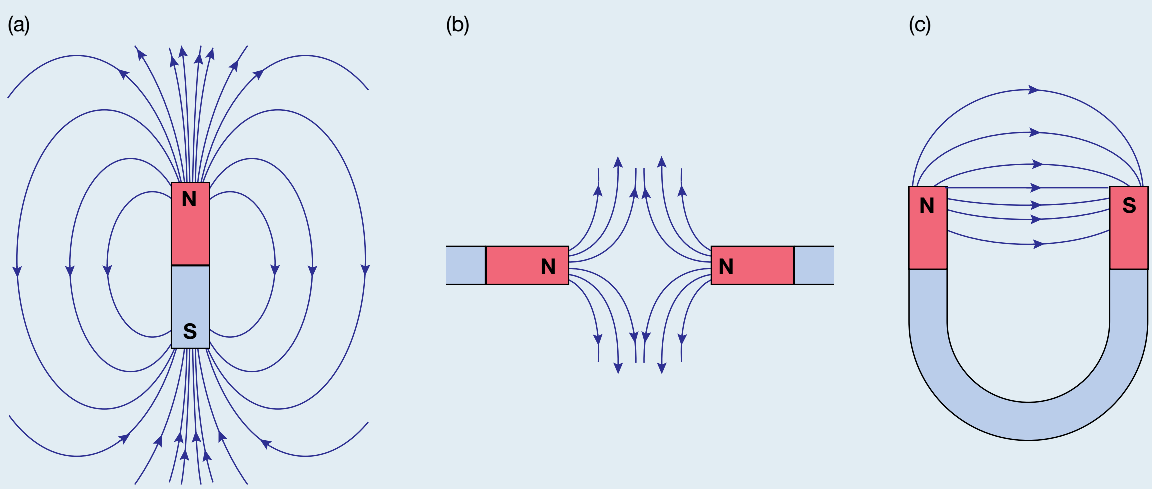
Use the box below to revise your work on magnetic fields. This material is fundamental to the understanding of how DC electric motors operate.

PHYSICS IN FOCUS

Review of magnetic fields

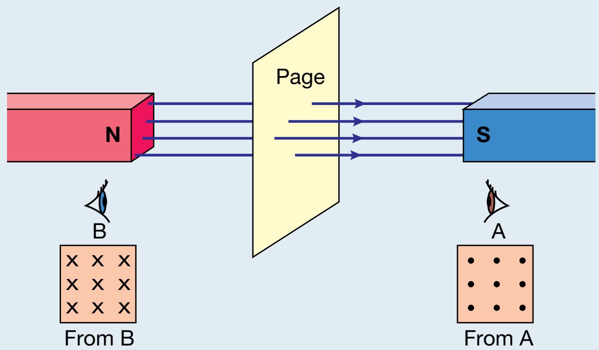
- The law of magnetic poles states that opposite poles of magnets attract each other and like poles of magnets repel each other.
- Magnetic fields are represented in diagrams using lines. These show the direction and strength of the field. The density of the field lines represents the strength of the magnetic field. The closer the lines are together, the stronger the field.
- The direction of the magnetic field at a particular point is given by the direction of the force on the N pole of a magnet placed within the magnetic field. It is shown by arrows on the magnetic field lines.
- Magnetic field lines never cross. When a region is influenced by the magnetic fields of two or more magnets or devices, the magnetic field lines show the strength and direction of the resultant magnetic field acting in the region. They show the combined effect of the individual magnetic fields.
- The spacing of the magnetic field lines represents the strength of the magnetic field. It follows that field lines that are an equal distance apart represent a uniform magnetic field.
- Magnetic field lines leave the N pole of a magnet and enter the S pole.
- The following diagrams in Figure 5.2 represent the magnetic fields around (a) a single bar magnet, (b) two N poles close to each other, and (c) a horseshoe magnet.

FIGURE 5.2 Magnetic field lines for (a) a single bar magnet, (b) two N poles close to each other and (c) a horseshoe magnet.



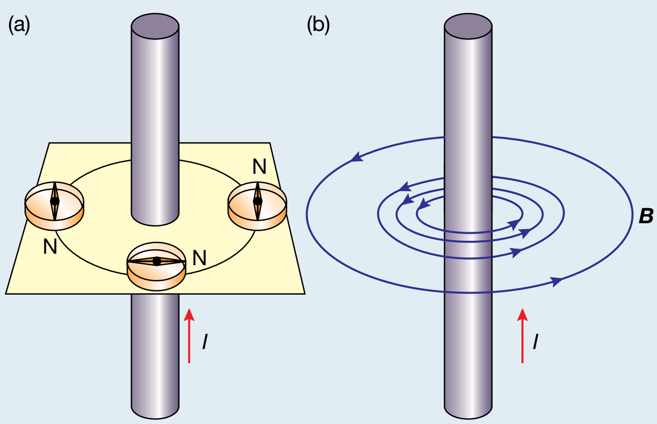
- In a diagram, as seen in Figure 5.3, magnetic field lines going out of the page are represented using dot points $(\cdot)$. This is like an observer seeing the pointy end of an arrow as it approaches.
- Magnetic field lines going into the page, also seen in Figure 5.3, are represented using crosses $(\times)$. This is as an archer would see the rear end of an arrow as it leaves the bow.

FIGURE 5.3 Magnetic field lines coming out of the page as observed by **A** and going into the page, as observed by **B**.



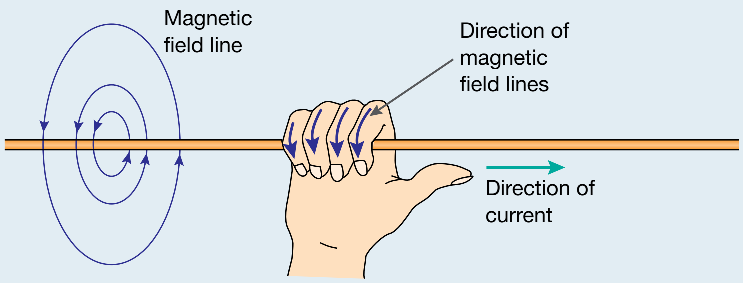
- The movement of charged particles, as occurs in an electric current, produces a magnetic field. The magnetic field is circular in nature around the current-carrying conductor, as shown in Figure 5.4, and can be represented using concentric field lines. The field gets weaker as the distance from the current increases.

FIGURE 5.4 (a) Compasses can be used to show the circular nature of the magnetic field around a straight current-carrying conductor. (b) The magnetic field is circular and stronger closer to the wire.



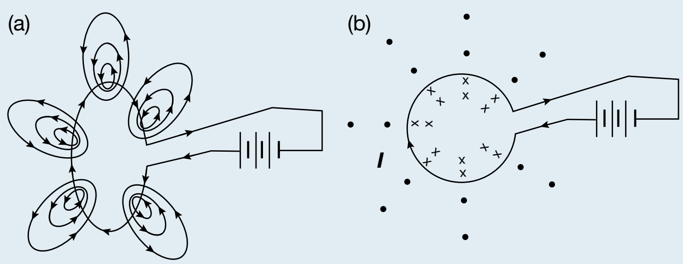
- The direction of the magnetic field around a straight current-carrying conductor is found using the right-hand grip rule, as shown in Figure 5.5. When the right hand grips the conductor with the thumb pointing in the direction of conventional current, the curl of the fingers gives the direction of the magnetic field around the conductor.

FIGURE 5.5 The right-hand grip rule.



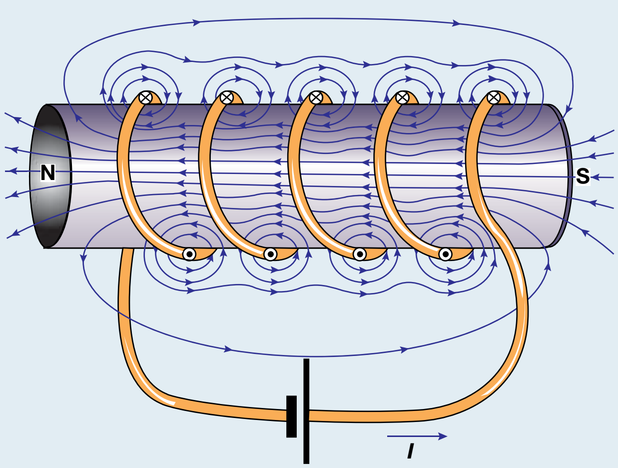
- When a current-carrying conductor is bent into a loop, the effect is to concentrate the magnetic field within the loop, as shown in Figure 5.6.

FIGURE 5.6 The magnetic field of a loop (a) 3-D representation (b) 2-D representation.



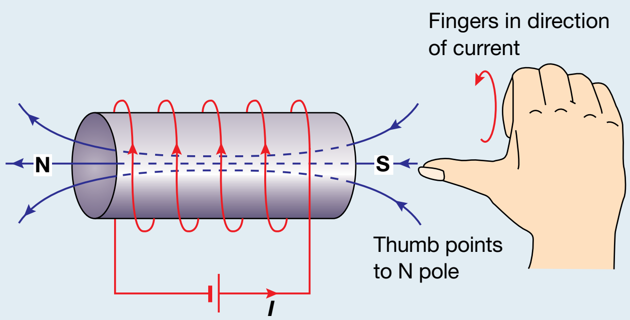
- A **solenoid** is a coil of insulated wire that can carry an electric current and is shown in Figure 5.7. The number of times that the wire has been wrapped around a tube to make the solenoid is known as the number of 'turns' or 'loops' of the solenoid. The magnetic fields around each loop of wire add together to produce a magnetic field similar to that of a bar magnet. Note that the magnetic field goes through the centre of the solenoid as well as outside it.

FIGURE 5.7 The magnetic field around a current-carrying solenoid.



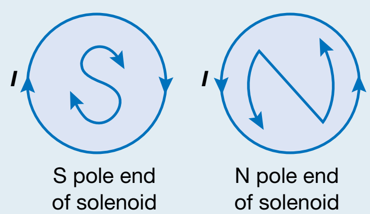
- The direction of the magnetic field produced by a solenoid can be determined using another right-hand grip rule; see Figure 5.8. In this case, the right hand grips the solenoid with the fingers pointing in the same direction as the conventional current flowing in the loops of wire and the thumb points to the end of the solenoid that acts like the N pole of a bar magnet, that is, the end of the solenoid from which the magnetic field lines emerge.

FIGURE 5.8 Determining the N pole of a solenoid.



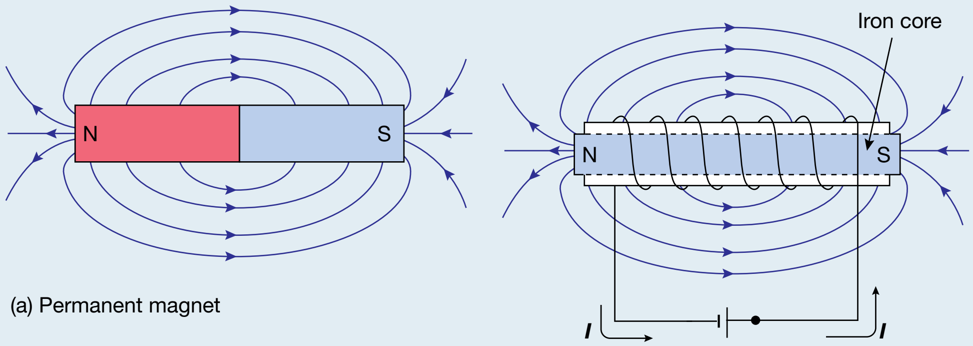
- Another method for determining the poles of a solenoid is to look at a diagram of the ends of the solenoid (see Figure 5.9), and mark in the direction of the conventional current around the solenoid. Then mark on the diagram the letter N or S that has the ends of the letter pointing in the same direction as the current. N is for an anticlockwise current, S is for a clockwise current.

FIGURE 5.9 Another method for determining the poles of a solenoid.



- An electromagnet is a solenoid that has a soft iron core. When a current flows through the solenoid, the iron core becomes a magnet. The polarity of the iron core is the same as the polarity of the solenoid. The core produces a much stronger magnetic field than is produced by the solenoid alone. In Figure 5.10 the magnetic field of a permanent magnet is compared to that of an electromagnet.
- The strength of an electromagnet can be increased by:
  - increasing the current through the solenoid
  - adding more turns of wire per unit length for a long solenoid
  - increasing the amount of soft iron in the core.

FIGURE 5.10 A permanent magnet and an electromagnet. Note the polarity of the iron core.



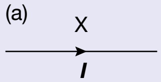
(a) Permanent magnet

(b) Electromagnet

5.2 Exercise 1

Use the right-hand-grip rule to determine the direction of the magnetic field at point X in the following diagrams.

1 (a)



(b)

current into page

2 Six compasses surround a current-carrying wire, as shown. The dark end of the compasses indicates the north pole. Which direction is a current flowing in the wire?

FIGURE 5.11

eBook*plus* RESOURCES

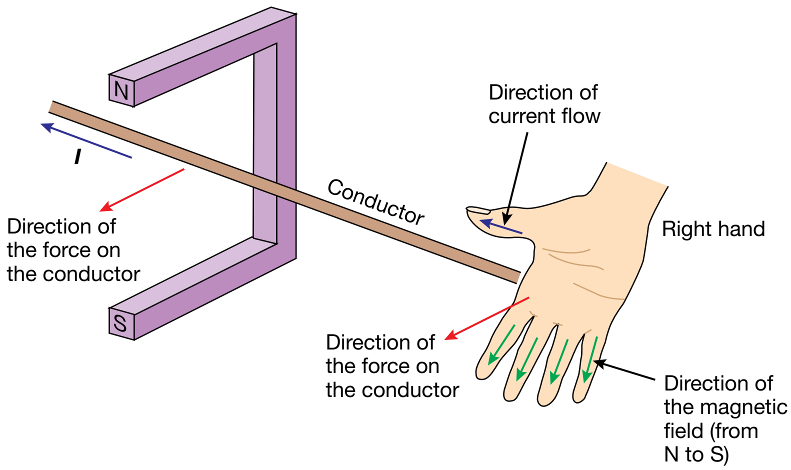
Watch this eLesson: Magnetic fields
Searchlight ID: med-0195

Explore more with this weblink: Magnetic field around a wire applet

5.3 The motor effect

5.3.1 The right-hand push rule

FIGURE 5.12 The right-hand push rule for a current-carrying conductor — the direction of a line at right angles to the palm gives the direction of the force.



A current-carrying conductor produces a magnetic field. When the current-carrying conductor passes through an external magnetic field, the magnetic field of the conductor interacts with the external magnetic field and the conductor experiences a force. This effect was discovered in 1821 by Michael Faraday (1791–1867) and is known as the **motor effect**. The motor effect is the action of a force experienced by a current-carrying conductor in an external magnetic field. The direction of the force on the current-carrying conductor in an external magnetic field can be determined using the **right-hand push rule** and can be seen in Figure 5.12. The right-hand push rule (also called the right-hand palm rule) is used to find the direction of the force acting on a current-carrying conductor in an external magnetic field.

5.3.2 Factors affecting the magnitude of the force

The magnitude of the force on a straight conductor in a magnetic field depends on the following factors:

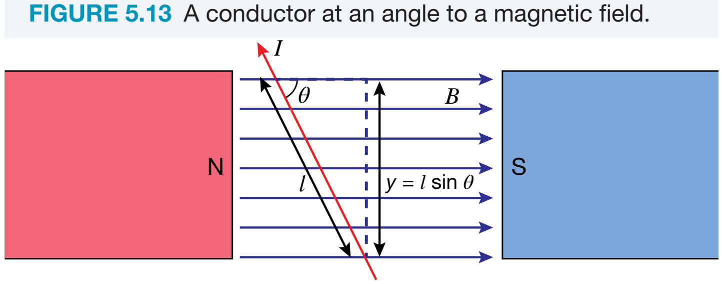
- the strength of the external magnetic field. The force is proportional to the magnetic field strength, B
- the magnitude of the current in the conductor. The force is proportional to current, I
- the length of the conductor in the field. The force is proportional to the length, l
- the angle between the conductor and the external magnetic field. The force is at a maximum when the conductor is at right angles to the field, and it is zero when the conductor is parallel to the field. The magnitude of the force is proportional to the component of the field that is at right angles to the conductor. If θ is the angle between the field and the conductor, then the force is the maximum value multiplied by the sine of θ.

  These factors are shown in Figure 5.13 and can be expressed mathematically as

  $\boldsymbol{F} = \boldsymbol{BI}l \sin\theta$.

  $\boldsymbol{F}$ is measured in newtons (N).
  $\boldsymbol{B}$ is the strength of the magnetic field, measured in tesla (T).
  $\boldsymbol{I}$ is the current flowing in the conductor, measured in ampere (A).
  l is the length of the conductor, measured in metres (m).

FIGURE 5.13 A conductor at an angle to a magnetic field.



5.3 SAMPLE PROBLEM 1

If a conductor of length 8.0 cm carries a current of 30 mA, calculate the magnitude of the force acting on it when in a magnetic field of strength 0.25 T if:

(a) the conductor is at right angles to the field
(b) the conductor makes an angle of 30° with the field
(c) the conductor is parallel with the field.

SOLUTION:

Use the equation:

$$F = BIl \sin\ \theta$$

where

$$l = 8.0 \times 10^{-2}\ \text{m}$$
$$I = 3.0 \times 10^{-2}\ \text{A}$$
$$B = 0.25\ \text{T}.$$

(a) $F = BIl \sin 90°$

$$= 3.0 \times 10^{-2} \times 8.0 \times 10^{-2} \times 0.25 \times 1$$
$$= 6.0 \times 10^{-4}\ \text{N}$$

(b) $F = BIl \sin 30°$

$$= 3.0 \times 10^{-2} \times 8.0 \times 10^{-2} \times 0.25 \times 0.5$$
$$= 3.0 \times 10^{-4}\ \text{N}$$

(c) $F = BIl \sin 0°$

$$= 3.0 \times 10^{-2} \times 8.0 \times 10^{-2} \times 0.25 \times 0$$
$$= 0$$

5.3 Exercise 1

1 Calculate the force on a 100 m length of wire carrying a current of 250 A when the strength of Earth's magnetic field at right angles to the wire is 5.00×10^{-5} T.

2 The force on a 10 cm wire carrying a current of 15 A when placed in a magnetic field perpendicular to *B* has a maximum value of 3.5 N. What is the strength of the magnetic field?

Working scientifically 5.1

Measuring the strength of a magnetic field

The relationship between ***F*** and ***BIl*** can be investigated by suspending several loops of wire from a spring balance in a magnetic field and putting a current through the loop. How should the field and the loops be arranged so that the force pulls the loops down and stretches the spring balance?

The loops initially stretch the spring balance, giving a reading of their weight. When there is a current through the loop, the magnetic force stretches the balance further. The extra weight is the magnetic force. If the current and the length of the loops are measured, and the number of loops is determined, the strength of the magnetic field can be calculated. Try this out with a horseshoe magnet.

What would happen if the current is reversed?

eBook*plus* RESOURCES

Try out this interactivity: Changing magnetic force
Searchlight ID: int-0115

5.4 The interaction between two parallel current-carrying wires

5.4.1 Forces between two parallel conductors

If a finite distance separates two parallel current-carrying conductors, then each conductor will experience a force due to the interaction of the magnetic fields that exist around each.

Figure 5.14, shows the situation where two long parallel conductors carry currents $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ in the same direction.

Figure 5.14a shows the magnetic field of conductor 1 in the region of conductor 2. Conductor 2 is cutting through the magnetic field due to conductor 1. The right-hand push rule shows that conductor 2 experiences a force directed towards conductor 1.

Similarly, Figure 5.14b shows the magnetic field of conductor 2 in the region of conductor 1. The right-hand push rule shows that conductor 1 experiences a force directed towards conductor 2. This means that the conductors are forced towards each other.

FIGURE 5.14 The forces acting on two long parallel conductors carrying currents in the same direction.

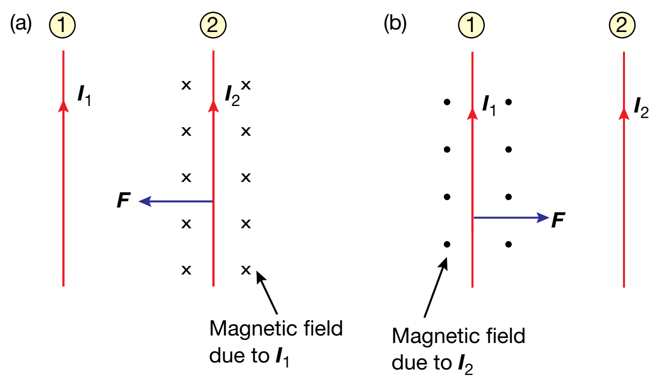


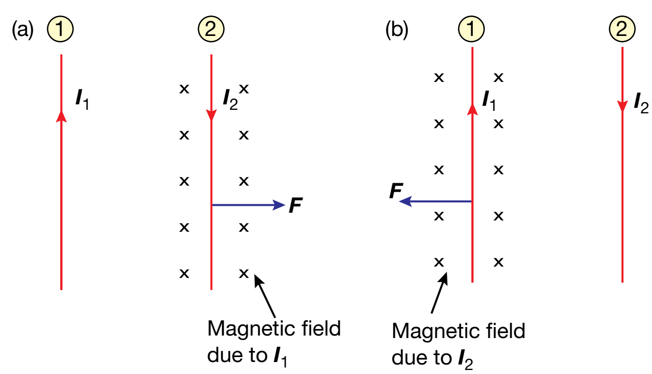
Figure 5.15 shows the situation where two long parallel conductors carry currents $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ in opposite directions.

Figure 5.15a shows the magnetic field of conductor 1 in the region of conductor 2. The right-hand push rule shows that conductor 2 experiences a force directed away from conductor 1.

Similarly, Figure 5.15b shows the magnetic field of conductor 2 in the region of conductor 1. The right-hand push rule shows that conductor 1 experiences a force directed away from conductor 2. This means that the conductors are forced apart.

Note that the magnitude of the forces acting on each pair of wires is equal, but the directions are opposite. This is true even if the conductors carry currents of different magnitudes.

FIGURE 5.15 The forces acting on two long parallel conductors carrying currents in opposite directions.



5.4.2 Determining the magnitude of the force between two parallel conductors

The magnetic field strength at a distance, d, from a long straight conductor carrying a current, I, can be found using the formula:

$$\boldsymbol{B} = \frac{\mathrm{k}\boldsymbol{I}}{d}$$

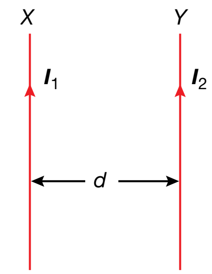
where

$\mathrm{k} = 2.0 \times 10^{-7}\,\mathrm{N\,A^{-2}}$.

Note that, in this equation, k is a constant derived from careful experimentation and that d is the perpendicular distance from the wire to the point at which B is to be calculated.

Figure 5.16 shows two parallel conductors, X and Y, that are carrying currents $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ respectively. X and Y are separated by a distance of d.

FIGURE 5.16 Two parallel current-carrying conductors.



The magnetic field strength in the region of Y due to the current flowing through X is:

$$B_X = \frac{kI_1}{d}.$$

The magnitude of the force experienced by a length, l, of conductor Y due to the external magnetic field provided by conductor X is:

$F = I_2 l B_X$, or

$$F = I_2 l \left(\frac{k\boldsymbol{I}_1}{d}\right)$$

This can be rearranged to give the formula:

$$\frac{F}{l} = k\frac{\boldsymbol{I}_1\boldsymbol{I}_2}{d}$$

A similar process can be used to show that the same formula will give the force experienced by a length, l, of conductor X due to the magnetic field created by the current flowing in conductor Y.

5.4 SAMPLE PROBLEM 1

FORCE BETWEEN PARALLEL CONDUCTORS

What is the magnitude and direction of the force acting on a 5.0 cm length of conductor X in Figure 5.16 if I_1 is 3.2 A, I_2 is 1.2 A, and the separation of X and Y is 25 cm?

SOLUTION:

Quantity	Value
F	?
k	2.0×10^{-7} N A^{-2}
l	5.0×10^{-2} m
I_1	3.2 A
I_2	1.2 A
d	0.25 m

Use the equation:

$$\frac{F}{l} = \frac{kI_1 I_2}{d}.$$

This transposes to give:

$$\begin{aligned} F &= \frac{k\,l\,I_1 I_2}{d} \\ &= \frac{2.0 \times 10^{-7} \times 5.0 \times 10^{-2} \times 3.2 \times 1.2}{0.25} \\ &= 1.5 \times 10^{-7}\text{ N.} \end{aligned}$$

To determine the direction of the force, first find the direction of the magnetic field at X, due to the current in Y, by using the right-hand grip rule. The field is out of the page. Next determine the direction of the force on X using the right-hand push rule. This shows that the force is to the right.

5.4.3 Determining the magnitude of the force between two parallel conductors

Determining the value of k

The force between two parallel conductors in free space is often stated

$$\frac{\boldsymbol{F}}{l} = \frac{\mu_o}{2\pi} \times \frac{\boldsymbol{I}_1\boldsymbol{I}_2}{r}$$

μ_o is the **magnetic permeability** of free space (or a vacuum). The magnetic permeability of a substance is the ability of a material to support the formation of a magnetic field. It is the degree of magnetisation that a material obtains in response to an applied magnetic field.

In general, the permeability of a substance is the ability of that substance to allow something to pass through it. You may have heard of semipermeable membranes in osmosis.

μ_o has a precise value of $4\pi \times 10^{-7}$ N A^{-2}.

The magnetic permeability of air has approximately the same value as the magnetic permeability of free space.

Hence, in air or a vacuum, $\frac{\mu}{4\pi} = 2.0 \times 10^{-7}$ N A^{-2} and the equation can be written as

$$\boldsymbol{F} = \frac{k\boldsymbol{I}_1\boldsymbol{I}_2 l}{r}$$

Where k = 2.0×10^{-7} N A^{-2}.

The magnetic permeability of iron is 2.5×10^{-1} N A^{-2}. Iron can become much more magnetised than air.

5.5 SI definition for electrical current; the ampere and Newton's Third Law of Motion

5.5.1 The ampere

The ampere is the unit for electrical current in the System International (SI) of Units. The formal definition of the ampere states: One ampere is the constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed one metre apart in vacuum, would produce between those conductors a force equal to 2×10^{-7} newtons per metre of length.

This is an application of Newton's Third Law of Motion, which states: In a two-body system, if body A exerts a force on body B, then body B exerts a force on Body A that is equal in magnitude, but opposite in direction.

Newton's Third Law of Motion is sometimes stated: For every action there is an equal and opposite reaction.

If one wire applies a force to a second wire, the second wire will apply a force that is equal in magnitude and opposite in direction on the first wire.

5.6 Review

5.6.1 Summary

- According to the motor effect, a current-carrying conductor in a magnetic field will experience a force that is perpendicular to the direction of the magnetic field. The direction of the force is determined using the right-hand push rule.
- The right-hand push rule is applied by:
  - extending the fingers in the direction of the magnetic field
  - pointing the thumb in the direction of the current in the conductor.
  - The palm of the hand indicates the direction of the force.

- The magnitude of the force, $\boldsymbol{F}$, on a current-carrying conductor is proportional to the strength of the magnetic field, $\boldsymbol{B}$, the magnitude of the current, $\boldsymbol{I}$, the length, l, of the conductor in the external field and the sine of the angle between the conductor and the field: $\boldsymbol{F} = \boldsymbol{BI}l \sin\theta$
- If the conductor is at right angles to the magnetic field, the force has the maximum value.
- If the conductor is parallel to the magnetic field, there is no force.
- Two long parallel current-carrying conductors will exert a force on each other. The magnitude of this force is determined using the following formula:

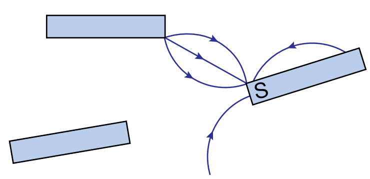
$$\frac{\boldsymbol{F}}{l} = \frac{\mu_o}{2\pi} \times \frac{\boldsymbol{I}_1\boldsymbol{I}_2}{r}$$

- If the currents in two parallel conductors are running in the same direction, the conductors attract each other. If the currents are in opposite directions, the conductors repel each other.
- One ampere is the constant current which, if maintained in two straight parallel conductors of infinite length and of negligible circular cross-section and placed one metre apart in a vacuum, would produce between those conductors a force equal to 2×10^{-7} newtons per metre of length.

5.6.2 Questions

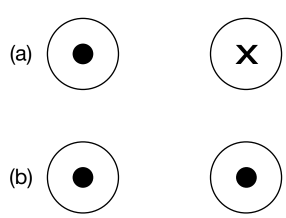
1. State the law of magnetic poles.
2. Draw a bar magnet and the magnetic field around it. Label the diagram to show that you understand the characteristics of magnetic field lines.
3. Are the north magnetic pole of the Earth and the north pole of a bar magnet of the same polarity? Explain your reasoning.
4. Figure 5.17 shows three bar magnets and some of the field lines of the resulting magnetic field.
   (a) Copy and complete the diagram to show the remaining field lines.
   (b) Label the polarities of the magnets.

FIGURE 5.17



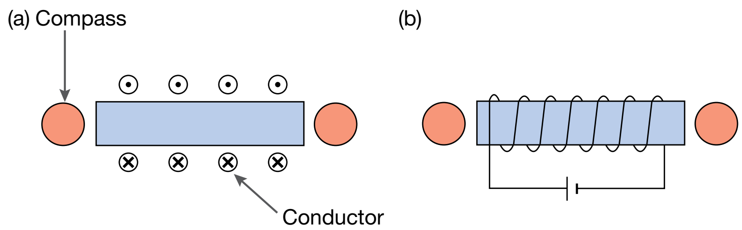
5. Draw a diagram to show the direction of the magnetic field lines around a conductor when the current is (a) travelling towards you and (b) away from you.
6. Each diagram in Figure 5.18 represents two parallel current-carrying conductors. In each case, determine whether the conductors attract or repel each other. Explain your reasoning.

FIGURE 5.18



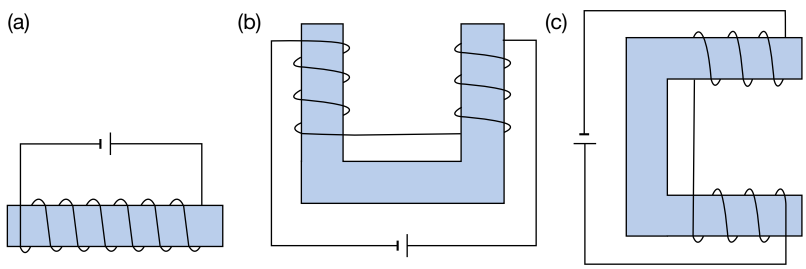
7. Each empty circle in Figure 5.19 represents a plotting compass near a coiled conductor. Copy the diagram and label the N and S poles of each coil, and indicate the direction of the needle of each compass.

FIGURE 5.19



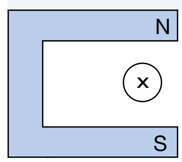
8. The diagrams in Figure 5.20 show electromagnets. Identify which poles are N and which are S.

FIGURE 5.20



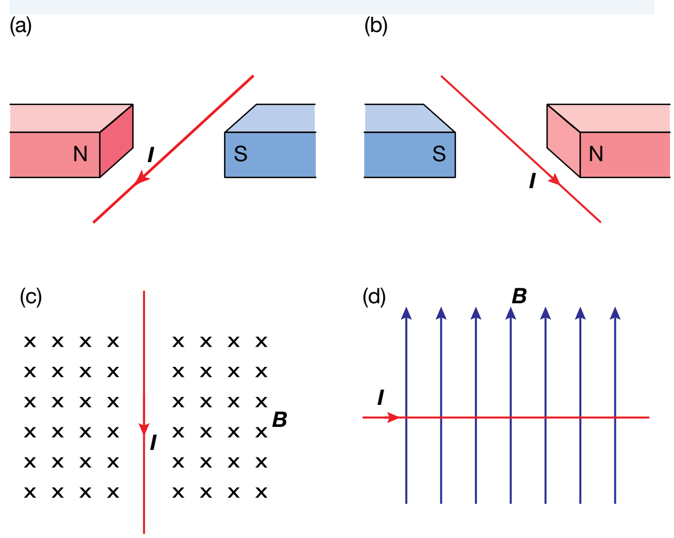
9. In Figure 5.21 a current-carrying conductor is in the field of a U-shaped magnet. Identify the direction in which the conductor is forced.

FIGURE 5.21



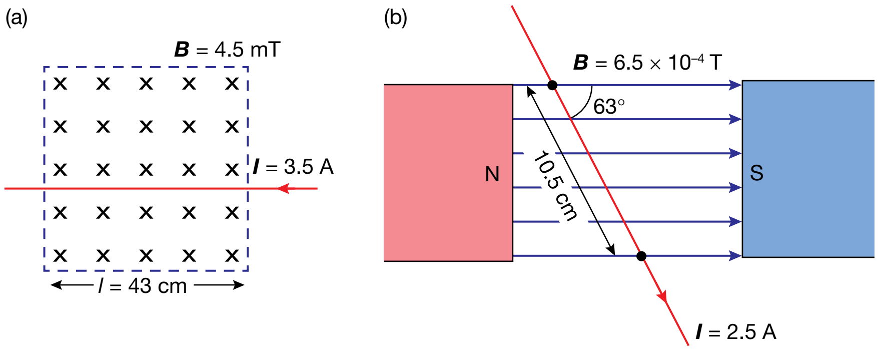
10. Identify the direction of the force acting on each of the current-carrying conductors shown in Figure 5.22. Use the terms 'up the page', 'down the page', 'into the page', 'out of the page', 'left' and 'right'.

FIGURE 5.22



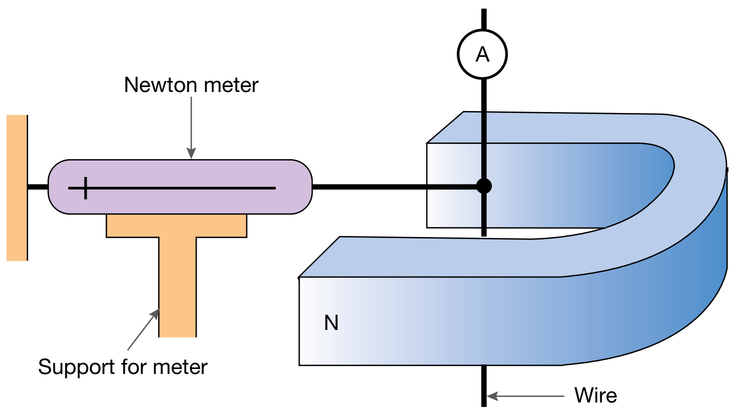
11. Deduce both the magnitude and direction of the forces acting on the lengths of conductors shown in Figure 5.23.

FIGURE 5.23



12. A student wishes to demonstrate the strength of a magnetic field in the region between the poles of a horseshoe magnet. He sets up the apparatus shown in Figure 5.24.

FIGURE 5.24

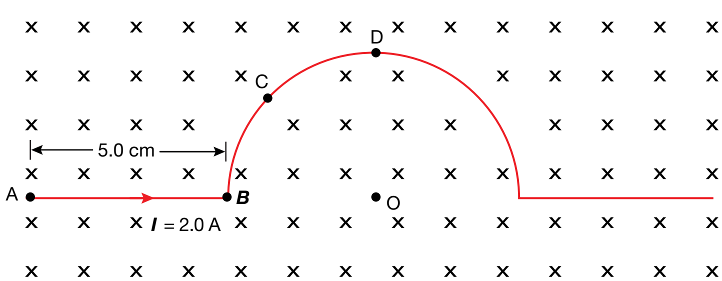


The length of wire in the magnetic field is 2.0 cm. When the ammeter reads 1.0 A, the force measured on the newton meter is 0.25 N.

(a) What is the strength of the magnetic field?
(b) In this experiment the wire moves to the right. In what direction is the current flowing, up or down the page?

13. A wire with the shape shown in Figure 5.25 carries a current of 2.0 A. It lies in a uniform magnetic field of strength 0.60 T.

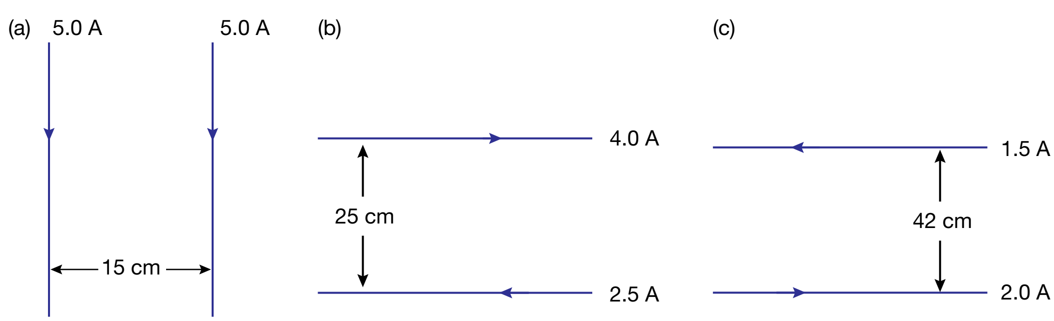
FIGURE 5.25



(a) Calculate the magnitude of the force acting on the section of wire, AB.
(b) Which of the following gives the direction of the force acting on the wire at the point, C?
  i. Into the page
  ii. Out of the page
  iii. In the direction OC
  iv. In the direction CO
  v. In the direction OD
  vi. In the direction DO
(c) Which of the following gives the direction of the net force acting on the semicircular section of wire?
  i. Into the page
  ii. Out of the page
  iii. In the direction OC
  iv. In the direction CO
  v. In the direction OD
  vi. In the direction DO

14. A wire of length 25 cm lies at right angles to a magnetic field of strength 4.0×10^{-2} T. A current of 1.8 A flows in the wire. Calculate the magnitude of the force that acts on the wire.
15. Two long straight parallel current-carrying wires are separated by 6.3 cm. One wire carries a current of 3.4 A upwards and the other carries a current of 2.5 A downwards.
    (a) Evaluate the magnitude of the force acting on a 45 cm length of one of the wires.
    (b) Is the force between the wires attraction or repulsion?
16. Evaluate the magnitude of the force acting on a 40 cm length of one of the two long wires shown in Figures 5.26 (a), (b) and (c).

FIGURE 5.26



eBook*plus* RESOURCES

Complete this digital doc: Investigation: The motor effect
Searchlight ID: doc-26587

Complete this digital doc: Investigation: The force between two parallel current-carrying conductors
Searchlight ID: doc-26588

PRACTICAL INVESTIGATIONS

Investigation 5.1: The motor effect

Aim

To observe the direction of the force on a current-carrying conductor in an external magnetic field.

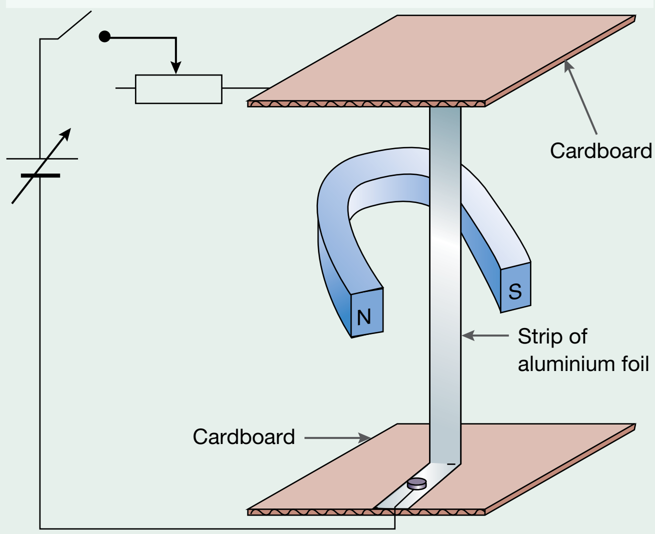
Apparatus

variable DC power supply
variable resistor (15 Ω rheostat)
connecting wires
retort stand
clamp
two pieces of thick card or balsa wood 10 cm × 10 cm
strip of aluminium foil 1 cm × 30 cm (approximately)
two drawing pins
switch
horseshoe magnet

Method

1. Pin the foil strip between the pieces of card. Rest one card on the bench-top and support the other with the clamp and retort stand.
2. Connect the wires to the power pack's DC terminals, switch, variable resistor and strips as shown in Figure 5.27. This will produce a current in the strip.
3. Position the horseshoe magnet so that the strip is between the poles. Note the position of the poles of the magnet and the direction of the current through the strip when the switch is closed.
4. Set the power pack to its lowest value and turn it on.
5. Briefly close the switch and record the movement of the foil strip.
6. Turn the magnet over so that the magnetic field is in the opposite direction across the strip.
7. Briefly close the switch and record the movement of the foil strip.

FIGURE 5.27 The set-up for the motor effect activity.



Analysis

1. Did the strip experience a force when a current flowed?
2. Verify that the movement of the aluminium strip is in accordance with the right-hand push rule.

Investigation 5.2: The force between two parallel current-carrying conductors

Aim

To observe the direction of the forces between two parallel current-carrying conductors.

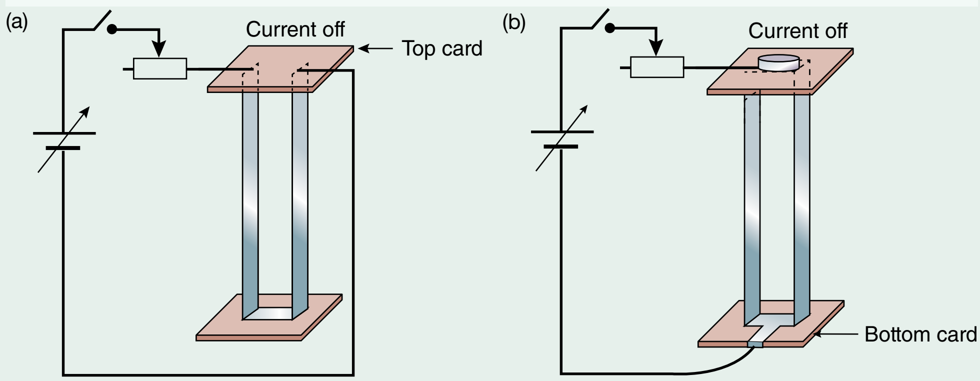
Apparatus

variable DC power supply
variable resistor (15Ω rheostat)
connecting wires
retort stand
clamp
two pieces of thick card or balsa wood 10 cm × 10 cm
two strips of aluminium foil 1 cm × 30 cm (approximately)
four drawing pins
push switch

Method

1. Pin each foil strip between the pieces of card so that they are parallel when the top card is supported by the clamp and retort stand.
2. Connect the wires to the power pack's DC terminals, switch, variable resistor and strips as shown in Figure 5.28a. This will produce currents in the strips that are flowing in opposite directions.
3. Set the power pack to its lowest value and turn it on.
4. Briefly close the switch and record the movement of the foil strips.
5. Connect the wires to the power pack's DC terminals, switch, variable resistor and strips as shown in Figure 5.28b. This will produce currents in the strips that are flowing in the same direction.
6. Briefly close the switch and record the movement of the foil strips.

FIGURE 5.28 (a) The set-up for currents flowing in opposite directions (b) The set-up for currents flowing in the same direction.



Analysis

Account for your observations.

TOPIC 6
Electromagnetic induction

6.1 Overview

6.1.1 Module 6: Electromagnetism

Electromagnetic Induction

Inquiry question: How are electric and magnetic fields related?

Students:

- describe how magnetic flux can change, with reference to the relationship $\phi = B_{\parallel}A = BA\cos\theta$ (ACSPH083, ACSPH107, ACSPH109)
- analyse qualitatively and quantitatively, with reference to energy transfers and transformations, examples of Faraday's Law and Lenz's Law, $\varepsilon = -N\dfrac{\Delta\phi}{\Delta t}$, including but not limited to: (ACSPH081, ACSPH110)
  - the generation of an electromotive force (emf) and evidence for Lenz's Law produced by the relative movement between a magnet, straight conductors, metal plates and solenoids
  - the generation of an emf produced by the relative movement or changes in current in one solenoid in the vicinity of another solenoid
- analyse quantitatively the operation of ideal transformers through the application of: (ACSPH110)
  - $\dfrac{V_p}{V_s} = \dfrac{N_p}{N_s}$
  - $V_pI_p = V_sI_s$
- evaluate qualitatively the limitations of the ideal transformer model and the strategies used to improve transformer efficiency, including but not limited to:
  - incomplete flux linkage
  - resistive heat production and eddy currents
- analyse applications of step-up and step-down transformers, including but not limited to:
  - the distribution of energy using high-voltage transmission lines

FIGURE 6.1 A large 90MVA 330/33KV power transformer for a windfarm and powerlines, located in NSW.

6.2 Electromagnetic induction

6.2.1 The discoveries of Michael Faraday

FIGURE 6.2 Portrait of Michael Faraday.

Michael Faraday (1791–1867) was the son of an English blacksmith. He started his working life at the age of twelve as an errand boy at a bookseller's store and later became a bookbinder's assistant. At the age of nineteen he attended a series of lectures at the Royal Institution in London that were given by Sir Humphrey Davey. This led to Faraday studying chemistry by himself. In 1813 he applied to Davey for a job at the Royal Institution and was hired as a research assistant. He soon showed his abilities as an experimenter and made important contributions to the understanding of chemistry, electricity and magnetism. He later became the superintendent of the Royal Institution.

In September 1821, following the 1820 discovery by Hans Christian Oersted (1777–1851) that an electric current produces a magnetic field, Michael Faraday discovered that a current-carrying conductor in a magnetic field experiences a force. This became known as the motor effect.

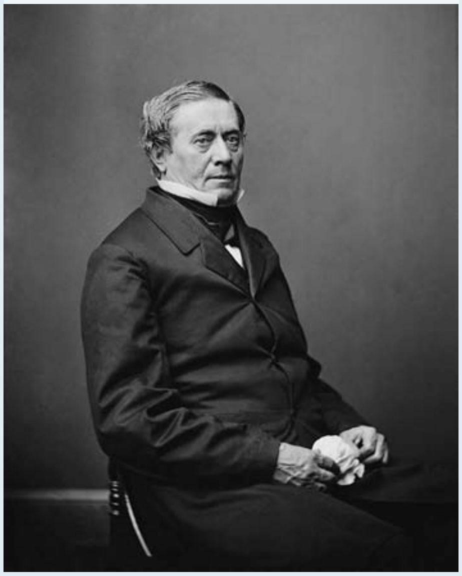
Almost 10 years later, in August 1831, Faraday discovered **electromagnetic induction**. Electromagnetic induction is the generation of an emf and/or electric current through the use of a magnetic field. Faraday's discovery was not accidental. He and other scientists spent many years searching for ways to produce an electric current using a magnetic field. Faraday's breakthrough eventually led to the development of the means of generating electrical energy in the vast quantities that we use in our society today.

PHYSICS FACT

Joseph Henry

American Joseph Henry (1797–1878) seems to have observed an induced current before Faraday, but Faraday published his results first and investigated the subject in more detail. Henry discovered self-inductance, while building electromagnets. Self-inductance occurs when the magnetic field created by a changing current in a circuit induces a voltage in that circuit. Henry developed the electromagnet to have many practical applications, such as the electric doorbell and the electric relay. His work was the basis for the electrical telegraph, which was invented, separately, by Samuel Morse and Sir Charles Wheatstone. In his honour, the SI unit of inductance is named the henry.

FIGURE 6.3 Portrait of Joseph Henry taken in the 1860s.

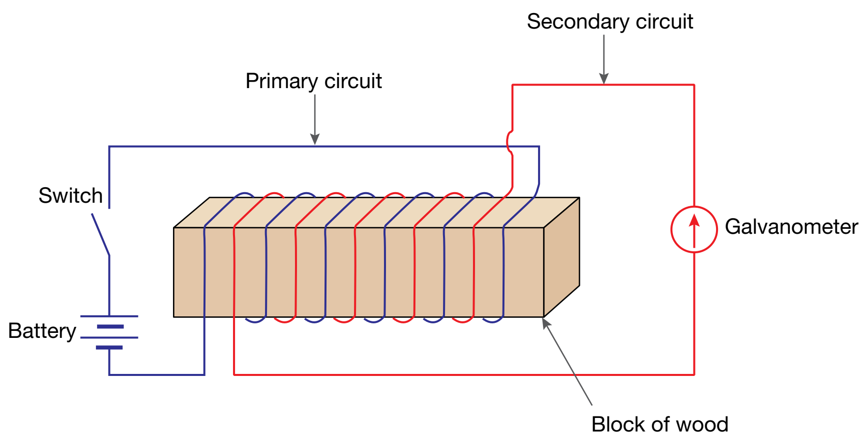


First experiments

In his first successful experiment, Faraday set out to produce and detect a current in a coil of wire by the presence of a magnetic field set up by another coil. He appears to have coiled about 70 m of copper wire around a block of wood. A second length of copper wire was then coiled around the block in the spaces between the first coil. The coils were separated with twine. One coil was connected to a **galvanometer** and the other to a battery. (A galvanometer is an instrument for detecting small electric currents. Faraday's early efforts to detect an induced current failed because of the lack of sensitivity of his galvanometers.) A simplified diagram of this experiment is shown in Figure 6.4.

When the battery circuit (or primary circuit) was closed, Faraday observed 'there was a sudden and very slight effect [deflection] at the galvanometer.' This means that Faraday had observed a small brief current that was created in the galvanometer circuit (or secondary circuit). A similar effect was also produced when the current in the battery circuit was stopped, but the momentary deflection of the galvanometer needle was in the opposite direction.

FIGURE 6.4 A simplified diagram of Faraday's first experiment.



Faraday was careful to emphasise that the current in the galvanometer circuit was a temporary one and that no current existed when the current in the battery circuit was at a constant value.

Faraday modified this experiment by winding the secondary coil around a glass tube. He placed a steel needle in the tube and closed the primary circuit. He then removed the needle and found that it had been magnetised. This also showed that a current had been produced (induced) in the secondary circuit. It was the magnetic field of the induced current in the secondary circuit that had magnetised the needle.

The next experiment was to place a steel needle in the secondary coil when a current was flowing in the primary coil. The primary current was stopped and the needle was again found to be magnetised, but with the poles reversed to the direction of the first experiment.

Iron ring experiment

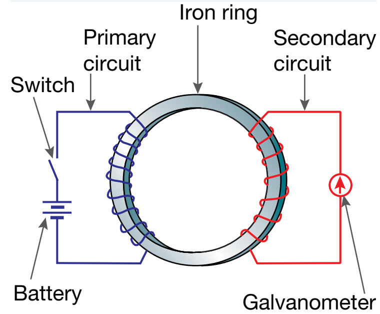
In a further experiment, Faraday used a ring made of soft iron (see Figure 6.5). He wound a primary coil on one side and connected it to a battery and switch. He wound a secondary coil on the other side and connected it to a galvanometer. A simplified diagram of this experiment is shown in Figure 6.6.

When the current was set up in the primary coil, the galvanometer needle immediately responded, as Faraday stated, 'to a degree far beyond what has been described when the helices [coils] without an iron core were used, but although the current in the primary was continued, the effect was not permanent, for the needle soon came to rest in its natural position, as if quite indifferent to the attached electromagnet'. When the current in the primary coil was stopped, the galvanometer needle moved in the opposite direction.

He concluded that when the magnetic field of the primary coil was changing, a current was induced in the secondary coil.

FIGURE 6.5 Photograph of the apparatus (two coils of insulated copper wire wound around an iron ring) that Faraday used to induce an electric current on 29 August 1831.

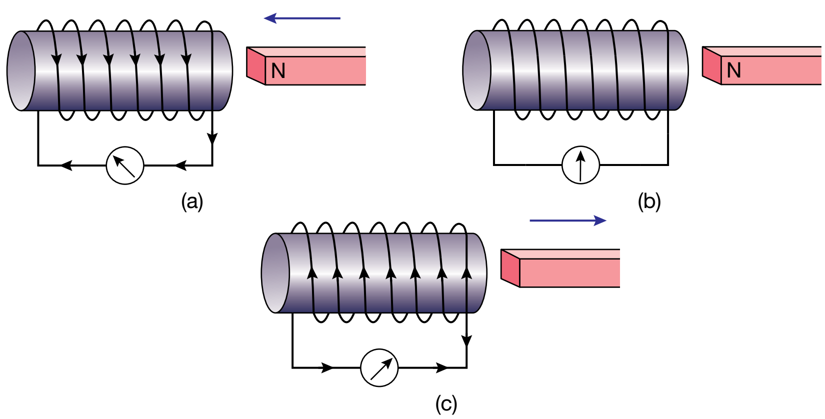
FIGURE 6.6 A simplified diagram of Faraday's iron ring experiment. Faraday's iron ring apparatus, an iron ring with a primary and secondary coil wrapped around it, is the basis of modern electrical transformers.



Using a moving magnet

Faraday was also able to show that moving a magnet near a coil could generate an electric current in the coil. The diagrams in Figure 6.7 show the effect when the N pole of a magnet is brought near a coil, held stationary, and then taken away from the coil.

FIGURE 6.7 (a) When the N pole of a bar magnet is brought near one end of the coil, the galvanometer needle momentarily deflects in one direction, indicating that a current has been induced in the coil circuit. (b) When the magnet is held without moving near the end of the coil, the needle stays at the central point of the scale (no deflection), indicating no induced current. (c) When the N pole of the magnet is taken away from the coil, the galvanometer needle momentarily deflects in the opposite direction to the first situation, indicating that an induced current exists and that it is flowing in the opposite direction.

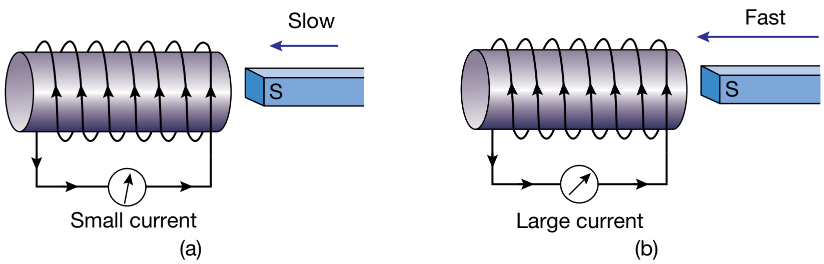


Similar results occur when the S pole is moved near the same end of the coil, except that the deflection of the galvanometer needle is in the opposite direction to when the N pole moves in the same direction.

Another observation from Faraday's experiments with a coil and a moving magnet is that the magnitude of the induced current depends on the speed at which the magnet is moving towards or away from the coil.

If the magnet moves slowly, a small current is induced. If the magnet moves quickly, the induced current has a greater magnitude. This observation is illustrated in Figure 6.8. Note that in the case that has been illustrated, the S pole is approaching the coil and the current is in the opposite direction to when the N pole approaches the coil from the same side.

FIGURE 6.8 (a) A slow-moving magnet induces a small current. (b) A fast-moving magnet induces a larger current.



You can repeat some of Faraday's experiments by doing practical investigations 6.1 and 6.2, included at the end of this topic.

6.3 Inducing a current

6.3.1 Magnetic flux

Electromagnetic **induction** is the creation of an electromotive force, emf, in a conductor when it is in relative motion to a magnetic field, or it is situated in a changing magnetic field. Such an emf is known as an induced emf. In a closed conducting circuit, the emf gives rise to a current known as an induced current.

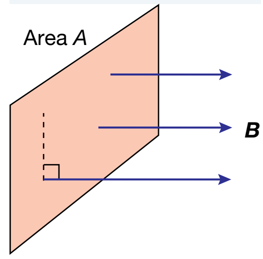
Faraday demonstrated that it was possible to produce (or induce) a current in a coil by using a changing magnetic field. For there to be a current in the coil, there must have been an emf induced in the coil. We will now examine how this is achieved.

The magnetic field in a region can be represented diagrammatically using field or **flux** lines. You can imagine the magnetic field 'flowing' out from the N pole of a magnet, spreading out around the magnet and then 'flowing' back into the magnet through the S pole. The field lines on a diagram show the direction of magnetic force experienced by the N pole of a test compass if it were placed at that point. The closeness (or density) of the lines represents the strength of the magnetic field. The closer together the lines, the stronger the field.

Magnetic flux is the name given to the amount of magnetic field passing through a given area. It is given the symbol ϕ_B. In the SI system, ϕ_B is measured in weber (Wb). If the particular area, A, is perpendicular to a uniform magnetic field of strength B (as shown in Figure 6.9) then the magnetic flux ϕ_B is the product of B and A.

$$\phi_B = BA$$

FIGURE 6.9 The magnetic field passing through an area at right angles.



The strength of a magnetic field, B, is also known as the **magnetic flux density**. It is the 'amount' of magnetic flux passing through a unit area. In the SI system, B is measured in tesla (T) or weber per square metre $\left(\text{Wb m}^{-2}\right)$.

The magnetic flux, ϕ_B, passing through an area is reduced if the magnetic field is not perpendicular to the area, and ϕ_B is zero if the magnetic field is parallel to the area. The above relationship between magnetic flux, magnetic flux density and area is often written as:

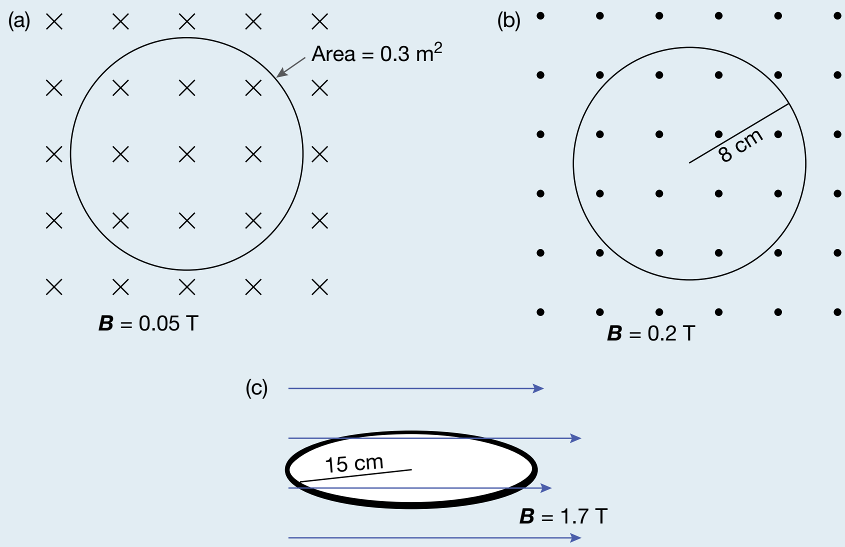
$$\phi_B = B_{\perp}A$$

where $B_{\perp}$ is the component of the magnetic flux density that is perpendicular to the area, A.

6.3 SAMPLE PROBLEM 1

Calculate the magnetic flux in each of the following situations.

FIGURE 6.10



SOLUTION:

(a) $\phi_B = B_\perp \times A$
$= 0.05\ \text{T} \times 0.3\ \text{m}^2$
$= 0.015\ \text{Wb}$

(b) First calculate area A. (Don't forget to convert the radius to metres.)

$A = \pi r^2$
$= \pi \times (0.08\ \text{m})^2$
$= 0.020\ 106\ \text{m}^2$ (Don't round off the final answer.)

Now calculate the flux:

$\phi_B = B_\perp \times A$
$= 0.2\ \text{T} \times 0.020\ 106\ \text{m}^2$
$= 0.004\ \text{Wb}$

(c) Note that the plane of the loop is parallel to the magnetic field,

$B_\perp = 0.$

$\phi_B = B_\perp \times A$
$= 0 \times A$
$= 0\ \text{Wb}$

6.3 Exercise 1

1 Estimate the maximum amount of magnetic flux passing through an earring when placed near a typical school magnet.

6.4 Generating a potential difference

6.4.1 Faraday's Law of Induction

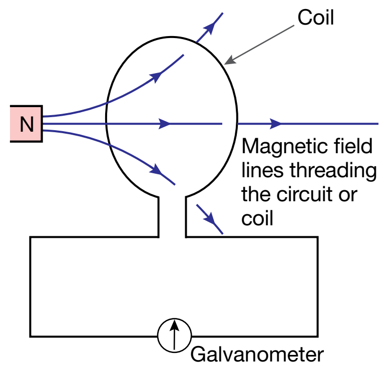
For a current to flow through the galvanometer in Faraday's experiments there must be an electromotive force (emf, symbol E or ε). The magnitude of the current through the galvanometer depends on the resistance of the circuit and the magnitude of the emf generated in the circuit.

Faraday noted that there had to be change occurring in the apparatus for an emf to be created. The quantity that was changing in each case was the amount of magnetic flux threading (or passing through) the coil in the galvanometer circuit (see Figure 6.11). The rate at which the magnetic flux changes determines the magnitude of the generated emf.

This gives Faraday's Law of Induction, which can be stated as follows:

The induced emf in a circuit is equal in magnitude to the rate at which the magnetic flux through the circuit is changing with time.

FIGURE 6.11 A galvanometer circuit showing magnetic flux threading the coil.



Faraday's law can be written in equation form as:

$$\varepsilon = -\frac{\Delta\phi_B}{\Delta t}.$$

The negative sign in the above equation indicates the direction of the induced emf. This is explained in Section 6.5, Lenz's Law.

PHYSICS FACT

The symbol for the Greek letter delta is Δ. It is used in mathematics and physics to represent a change in a quantity.

The change in a quantity is calculated by subtracting the initial value from the final value. For example, the change in your bank balance over a month is the final balance minus your initial balance.

When calculating quantities using Faraday's Law of Induction,

$$\Delta\phi_B = \phi_{B\text{final}} - \phi_{B\text{initial}}.$$

Since $\phi_B = \boldsymbol{B}_\perp A$, a change in ϕ_B can be caused by a change in the magnetic field strength, B, or in the area of the coil that is perpendicular to the magnetic field, or both.

If a coil has n turns of wire on it, the emf induced by a change in the magnetic flux threading the coil would be n times greater than that produced if the coil had only one turn of wire, that is,

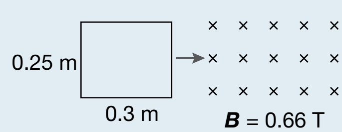
$$\varepsilon = \frac{-n\phi_B}{\Delta t}$$

6.4 SAMPLE PROBLEM 1

The rectangular loop shown takes 2.0 s to fully enter a perpendicular magnetic field of 0.66 T strength.

(a) What is the magnitude of the emf induced in the loop?
(b) In which direction does the current flow around the loop?

FIGURE 6.12



SOLUTION:

(a) First calculate the area of the loop.

$$A = 0.25\text{ m} \times 0.3\text{ m}$$
$$= 0.075\text{ m}^2$$

Now find the change in flux.

$$\begin{aligned}\Delta\phi_B &= \phi_{B\text{ final}} - \phi_{B\text{ initial}}\\ &= (BA)_{\text{final}} - (BA)_{\text{initial}}\\ &= \left(0.66\text{ T} \times 0.075\text{ m}^2\right) - \left(0\text{ T} \times 0.075\text{ m}^2\right)\ \text{(The initial field strength through the coil is zero.)}\\ &= \left(0.05\text{ T m}^2\right) - \left(0\text{ T m}^2\right)\\ &= 0.05\text{ Wb into the page}\end{aligned}$$

Finally, using Faraday's Law:

$$\begin{aligned}\text{emf, } \varepsilon &= \frac{-N\Delta\phi_{\text{B}}}{\Delta t}\\ &= -1 \times \frac{0.05\text{Wb}}{2.0\text{ s}}\\ &= -0.025\text{ V}\end{aligned}$$

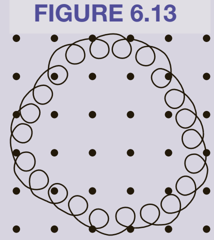
So the magnitude of the induced voltage is 0.025 V.
The minus sign is there to indicate that the induced emf opposes the change in magnetic flux.

(b) Change in flux = final – initial = flux into the page
Direction of induced magnetic field = out of the page (Len'z law, see Section 6.5)
Direction of induced current = anticlockwise (right-hand-grip rule)

6.4 Exercise 1

1 A spring is bent into a circle and stretched out to a radius of 5.0 cm. It is then placed in a magnetic field of strength 0.55 T. The spring is released and contracts down to a circle of radius 3.0 cm. This happens in 0.15 seconds.
(a) What is the magnitude of the induced emf?
(b) In what direction does the current move?

FIGURE 6.13



6.4.2 Rotating coils in uniform magnetic fields

When a coil rotates in a magnetic field, as occurs in generators and motors, the flux threading the coil is a maximum when the plane of the coil is perpendicular to the direction of the magnetic field. If the plane of the coil is parallel to the direction of the magnetic field, the flux threading the coil is zero, so rotating the coil changes the magnetic flux.

eBook*plus* RESOURCES

Try this interactivity: Magnetic flux and Lenz's Law
Searchlight ID: int-0050

Watch this eLesson: Magnetic flux and Lenz's Law
Searchlight ID: eles-0026

Watch this eLesson: Magnetic flux
Searchlight ID: med-0196

6.5 Lenz's law

H.F. Lenz (1804 –1864) was a German scientist who, without knowledge of the work of Faraday and Henry, duplicated many of their experiments. Lenz discovered a way to predict the direction of an induced current. This method is given the name Lenz's Law. It can be stated in the following way:

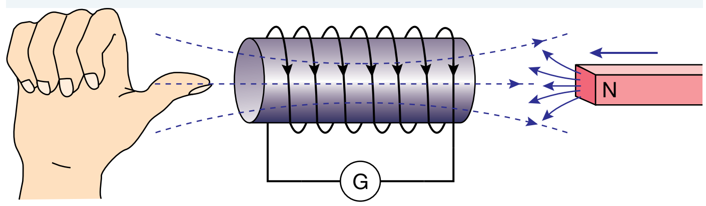
An induced emf always gives rise to a current that creates a magnetic field that opposes the original change in flux through the circuit.

This is a consequence of the Principle of Conservation of Energy. The minus sign in Faraday's Law of Induction is placed there to remind us of the direction of the induced emf.

6.5.1 Using Lenz's Law

When determining the direction of the induced emf, it is useful to use the field line method for representing magnetic fields. Figure 6.14 shows the effect of a magnet moving closer to a coil connected to a galvanometer. The coil is wound on a cardboard tube. As the magnet approaches the coil, the magnetic flux density within the coil increases. The induced current sets up a magnetic field (shown in dotted lines) that opposes this change. The approaching magnet increases the number of field lines pointing to the left that pass through the coil. The induced current in the coil produces field lines that point to the right to counter this increase.

FIGURE 6.14 The N pole of a magnet approaches a coil. Note that the induced magnetic field of the coil repels the approaching N pole.



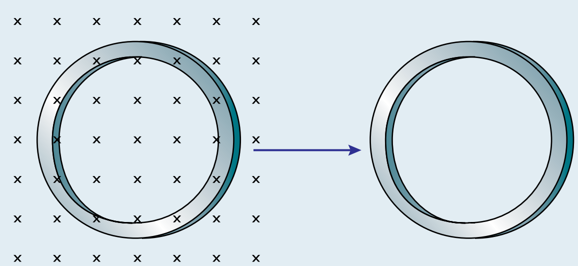
The direction of the induced current in the coil can be deduced using the right-hand rule for coils. The thumb points in the direction of the induced magnetic field within the coil, the curl of the fingers holding the coil shows the direction of the induced current in the coil. Note that magnetic field lines do not cross. Dotted lines have been used to show the general direction of the induced field lines, not the resultant field.

6.5 SAMPLE PROBLEM 1

INDUCED CURRENT IN A COIL

A metal ring initially lies in a uniform magnetic field, as shown in Figure 6.15. The ring is then removed from the magnetic field. In which direction does the induced current flow in the coil?

FIGURE 6.15



SOLUTION:

Initially the magnetic field lines of the external field are passing into the page through the coil. As the coil is removed from the field, these field lines reduce in number. The induced current flows in such a way as to create a magnetic field to replace the missing lines. Therefore, the current in the ring must flow in a clockwise direction in the ring, as indicated by using the right-hand rule for coils. The current stops flowing when the entire ring has been removed from the external magnetic field.

6.6 Eddy currents

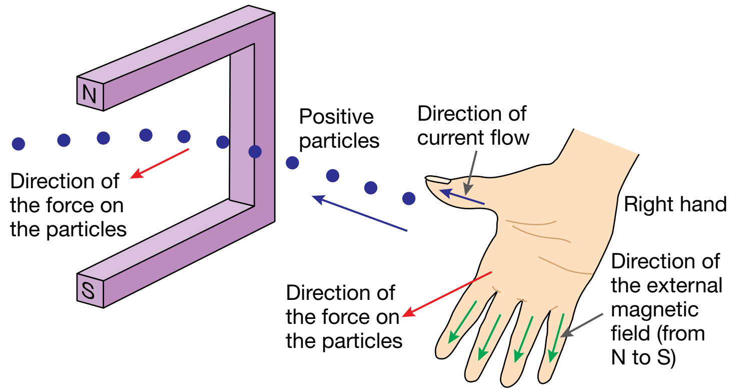
6.6.1 Charged particles moving in magnetic fields

Moving charged particles, for example electrons or alpha particles, produce magnetic fields. The direction of the magnetic field is found using the **right-hand grip rule**.

When moving charged particles enter an external magnetic field, the magnetic field created by the moving charged particles interacts with the external magnetic field. (An external magnetic field is one that already exists or that is caused by another source.)

When the moving charged particles enter the magnetic field at right angles to the field, they experience a force that is at right angles to the velocity and to the direction of the external field. The direction of the force is determined by using the right-hand push rule (also known as the right-hand palm rule) and is demonstrated in Figure 6.16.

FIGURE 6.16 The right-hand push rule for moving charged particles.



To use the right-hand push rule, position your right hand so that:

- the fingers point in the direction of the external field
- the thumb points in the direction of conventional current flow (this means in the direction of the velocity of positive charges or in the opposite direction to the velocity of negative charges)
- the direction of the force on the particles is directly away from the palm of the hand.

6.6.2 Magnetic fields and eddy currents

Induced currents do not occur in only coils and wires. They can also occur:

- when there is a magnetic field acting on part of a metal object and there is relative movement between the magnetic field and the object
- when a conductor is moving in an external magnetic field
- when a metal object is subjected to a changing magnetic field.

Such currents are known as **eddy currents**. They resemble the eddies or swirls left in the water after a boat has gone by. Eddy currents are an application of Lenz's Law. The magnetic fields set up by the eddy currents oppose the changes in the magnetic field acting in the regions of the metal objects.

FIGURE 6.17 The production of eddy currents in a sheet of metal.

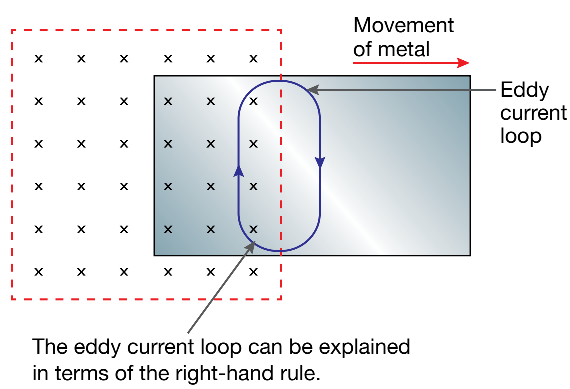


Figure 6.17 shows one method of producing an eddy current. A rectangular sheet of metal is being removed from an external magnetic field that is directed into the page. On the left of the edge of the magnetic field, charged particles in the metal sheet experience a force because they are moving relative to the magnetic field. By applying the right-hand push rule, it can be seen that positive charges experience a force up the page in

this region. To the right of the edge of the magnetic field, charged particles experience no force. Therefore the charged particles that are free to move at the edge of the field contribute to an upward current that is able to flow downwards in the metal that is outside the field. This forms a current loop that is known as an eddy current.

The side of the eddy current loop that is inside the magnetic field experiences a force due to the magnetic field. The direction of the force on the eddy current can be determined using the right-hand push rule and it is always opposite to the direction of motion of the sheet. (This means, referring to Figure 6.17, it is harder to move the metal to the right when the magnetic field is present than when the field is not present.)

6.6.3 Eddy currents in switching devices

Induction switches are electronic devices that detect the presence of metals and switch on another part of a circuit. Walk-through metal detectors at airports use induction switching devices.

Induction switching devices consist of a high-frequency oscillator, an analysing circuit and a relay. The oscillator produces an alternating current in a coil. This produces an electromagnetic field with a frequency of up to 22 MHz. When a metal object comes near the coil, eddy currents are created in the object. The eddy currents place a load on the coil and the frequency of the oscillator is reduced. The analysing circuit monitors the frequency of the oscillator and, when it falls below a certain threshold value, switches on an alarm circuit using the relay. The threshold frequency can be adjusted so that small loads such as a few coins or metal buttons and zippers will not trigger the alarm, but larger loads such as guns and knives will.

PHYSICS IN FOCUS

Induction heating

Another effect of eddy currents is that they cause an increase in the temperature of the metal. This is due to the collisions between moving charges and the atoms of the metal, as well as the direct agitation of atoms by a magnetic field changing direction at a high frequency.

Induction heating is the heating of an electrically conducting material by the production of eddy currents within the material. This is caused by a changing magnetic field that passes through the material. Induction heating is undesirable in electrical equipment such as motors, generators and transformers, but it has been put to good use with induction cookers and induction furnaces.

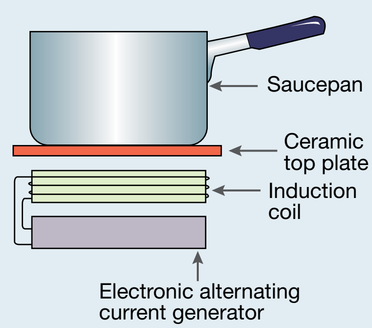
Applying the principle of induction to cooktops in electric ranges

A gas stovetop cooks food by burning gas to produce hot gases. The gases then flow across the bottom of a saucepan and transfer heat into it by conduction. However, a large amount of the thermal energy in the gases is carried away into the environment of the kitchen. The heat transferred to the saucepan is used to cook the food.

Some electric cooktops contain induction cookers instead of heating coils. An induction cooker sets up a rapidly changing magnetic field that induces eddy currents in the metal of the saucepan placed on the cooktop. The eddy currents cause the metal to heat up directly without the loss of thermal energy that occurs with gas cooking. The heat produced in the metal saucepan is used to cook the food. The induction coils of the cooker are separated from the saucepan by a ceramic top plate. Induction cookers have an efficiency of about 80% while gas cookers have an efficiency rating of about 43%. A diagram of an induction cooker is shown in Figure 6.18.

As the current is alternating in the coil, there is a changing magnetic field that cuts through the metallic saucepan, causing eddy currents in the saucepan.

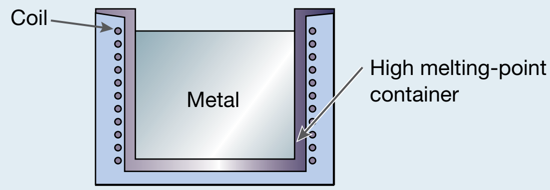
FIGURE 6.18 An induction cooker.



Induction furnaces

An induction furnace makes use of the heating effect of eddy currents to melt metals. This type of furnace consists of a container made from a nonmetal material that has a high melting point and that is surrounded by a coil. The metal is placed in the container. The coil is supplied with an alternating current that can have a range of frequencies and this produces a changing magnetic field through the metal. Eddy currents in the metal raise its temperature until it melts. The eddy currents also produce a stirring effect in the molten metal, making the production of alloys easier. Induction furnaces take less time to melt the metal than flame furnaces. They are also cleaner and more efficient. A diagram of an induction furnace is shown in Figure 6.19.

FIGURE 6.19 An induction furnace.



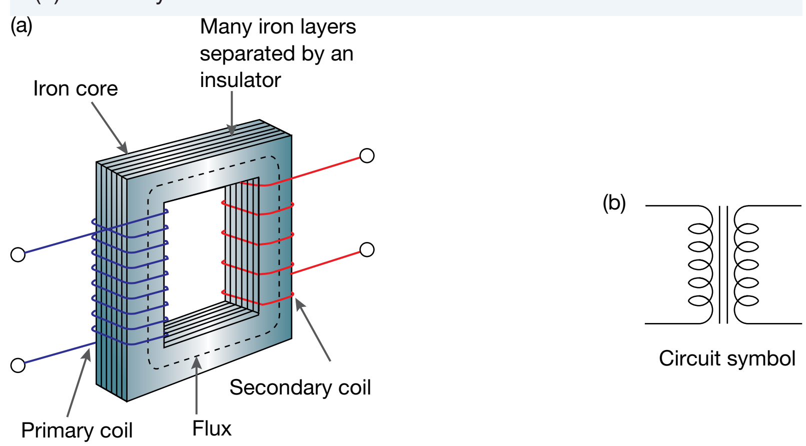
6.7 Transformers

6.7.1 The operation of ideal transformers

Transformers are devices that increase or decrease AC voltages. They are used in television sets and computer monitors. They are used in electronic appliances such as radios to provide lower voltages for amplifier circuits. They are also found in digital cameras, battery chargers, digital clocks, computers, phones, printers, electronic keyboards, the electric power distribution system and many other devices.

Transformers consist of two coils of insulated wire called the *primary* and *secondary* coils. These coils can be wound together onto the same soft iron core, or linked by a soft iron core. The structure of the most common type of transformer and its circuit symbol are shown in Figure 6.20.

FIGURE 6.20 (a) A transformer with coils linked by a soft iron core (b) Circuit symbol.



Transformers are designed so that almost all the magnetic flux produced in the primary coil threads the secondary coil. When an alternating current flows through the primary coil, a constantly changing magnetic flux threads (or passes through) the secondary coil. This constantly changing flux passing through the secondary coil produces an AC voltage at the terminals of the secondary coil with the same frequency as the AC voltage supplied to the terminals of the primary coil.

The difference between the primary voltage, $V_{p,}$ and the secondary voltage, $V_{s,}$ is in their magnitudes. The secondary voltage can be greater than or less than the primary voltage, depending on the design of the transformer. The magnitude of the secondary voltage depends on the number of turns of wire on the primary coil, $n_{p,}$ and secondary coil, n_s.

If the transformer is ideal, it is 100% efficient and the energy input at the primary coil is equal to the energy output of the secondary coil. The rate of change of flux $\left(\frac{\Delta\phi}{\Delta t}\right)$ through both coils is the same. Faraday's Law can be used to show that the secondary voltage is found using the formula:

$$V_s = n_s \frac{\Delta\phi}{\Delta t}.$$

Similarly, the input primary voltage, V_p, is related to the change in flux by the equation:

$$V_p = n_p \frac{\Delta\phi}{\Delta t}.$$

Dividing these equations produces the transformer equation:

$$\frac{V_p}{V_s} = \frac{n_p}{n_s}.$$

If n_s is greater than n_p, the output voltage, V_s, will be greater than the input voltage, V_p. Such a transformer is known as a **step-up transformer**. If n_s is less than n_p, the output voltage, V_s, will be less than the input voltage, V_p. Such a transformer is known as a **step-down transformer**.

6.7.2 Transformers and the Principle of Conservation of Energy

The Principle of Conservation of Energy states that energy cannot be created or destroyed but that it can be transformed from one form to another. This means that if a step-up transformer gives a greater voltage at the output, there must be some kind of a trade-off. The rate of supply of energy to the primary coil must be greater than or equal to the rate of supply of energy from the secondary coil. For example, if 100 J of energy is supplied each second to the primary coil, then the maximum amount of energy that can be obtained each second from the secondary coil is 100 J.

You cannot get more energy out of a transformer than you put into it. (Some energy is usually transformed into thermal energy in the transformer due to the occurrence of eddy currents in the iron core. In other words, eddy currents in the iron core cause the transformer to heat up.) There is a decrease in useable energy whenever energy is transformed from one form to another. The 'lost' energy is said to be dissipated, usually as thermal energy.

The rate of supply of energy is known as *power* and is found using the equation:

$$P = VI.$$

In ideal transformers, there is assumed to be no power loss and the primary power is equal to the secondary power. In this case:

$$P_p = P_s.$$

Substituting the power formula stated earlier, this equation becomes:

$$V_p I_p = V_s I_s.$$

Combining this equation with the transformer equation, we get another very important relationship for transformers:

$$\frac{I_s}{I_p} = \frac{n_p}{n_s}.$$

6.7 SAMPLE PROBLEM 1

The transformer in an electric piano reduces a 240 V AC voltage to a 12.0 V AC voltage. If the secondary coil has 30 turns and the piano draws a current of 500 mA, calculate the following quantities:

(a) the number of turns in the primary coil
(b) the current in the primary coil
(c) the power output of the transformer.

SOLUTION:

(a)

QUANTITY	VALUE
V_p	240 V
V_s	12.0 V
n_s	30
n_p	?

$$\frac{V_p}{V_s} = \frac{n_p}{n_s}$$

$$\Rightarrow n_p = \frac{n_s V_p}{V_s} = \frac{30 \times 240}{12.0} = 600$$

Therefore the primary coil has 600 turns.

(b)

QUANTITY	VALUE
I_p	?
I_s	500 mA
n_s	30
n_s	600

$$\frac{I_s}{I_p} = \frac{n_p}{n_s}$$

$$\Rightarrow I_p = \frac{n_s I_s}{n_p} = \frac{30 \times 500}{600} = 25\text{ mA}$$

(c)

QUANTITY	VALUE
V_s	12.0 V
I_s	500 mA
P_s	?

$$P_s = V_s I_s = 12.0 \times 500 = 6000\ \text{mW} = 6.0\ \text{W}$$

6.7.3 Limitations of the ideal transformer model

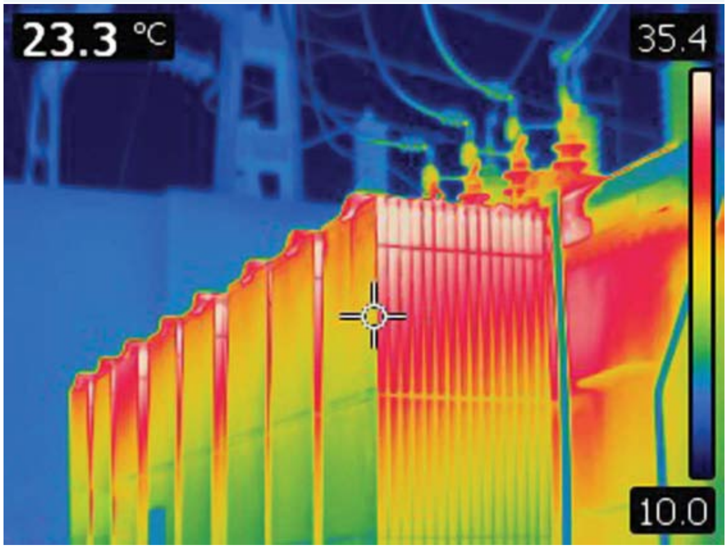
So far you have been looking at an ideal transformer: it is 100% efficient. For such a transformer, the output power equals the input power.

But real transformers are not 100% efficient. The secondary voltage and current will both be less than that predicted by using the formula $V_p I_p = V_s I_s$. This means that $V_p I_p > V_s I_s$.

There are three main reasons for this less-than-ideal behaviour:

- Incomplete flux linkage (or flux leakage): the magnetic field generated by the primary coil does not entirely pass through the secondary coil. A magnetic field can always be detected near a transformer.
- Resistive heat production: the primary and secondary coils of transformers are made from thin copper wires. When a current passes through the wires they heat up, so energy (and power) are lost in the primary and secondary coils. Thermal images of transformers show that they have higher temperatures than their surroundings.

FIGURE 6.21 This infrared photo shows the heat generated from a transformer. The red areas are the hottest.



- Eddy currents in the iron core: the changing magnetic field passing through the iron core causes a force on the loosely bound electrons, creating eddy currents.

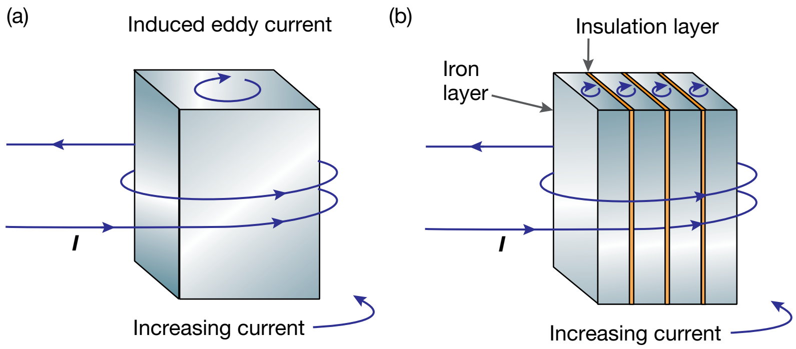
FIGURE 6.22 Workers assemble the laminated core at Wilson Transformers.

6.7.4 Reducing heat losses due to eddy currents

As we saw in Section 6.5, eddy currents within a metal are circular movements of electrons due to a changing flux passing through the metal. These circular movements are at right angles to the direction of the changing flux.

By constructing the iron core from many layers of iron that are coated with an insulator, the size of the eddy currents is reduced and the losses due to heating effects are reduced. Such a core is called a laminated iron core. The cross-sections of the thin layers are perpendicular to the direction of the magnetic flux so the size of the eddy currents is greatly reduced, as illustrated in Figure 6.23.

FIGURE 6.23 Eddy currents in (a) an ordinary iron core, (b) a laminated iron core.



Another method for reducing eddy current losses in transformers is to use materials called *ferrites*, which are complex oxides of iron and other metals. These materials are good transmitters of magnetic flux but are poor conductors of electricity, so the magnitudes of eddy currents are significantly reduced.

6.7.5 Applications of step-up and step-down transformers

PHYSICS IN FOCUS

Household use of transformers

Australian houses are provided with AC electricity that has a value of 240 V_{RMS}. Most electronic circuits are designed to operate at low DC voltages of between 3 V and 12 V. Therefore, household appliances that have electronic circuits in them will either have a 'power-cube' transformer that plugs directly into the power outlet socket, or have transformers built into them.

Power-cube transformers can be found in rechargeable appliances such as mobile phone chargers, electric keyboards, answering machines, cordless telephones and laptop computers. You can probably find more in your own home. These transformers also have a rectifier circuit built into them that converts AC to DC.

The RMS value of an AC voltage is a way of describing a voltage that is continuously changing. The voltage actually swings between −339 V and +339 V at a frequency of 50 Hz. This voltage has the same heating effect on a metal conductor as a DC voltage of 240 V; hence, we usually describe it as 240 V.

6.7 Exercise 1

1 A step-down transformer is designed to convert 230 V AC to 12 V AC. If there are 190 turns in the primary coil, how many turns are in the secondary coil?
2 A generator supplies 10 kW of power to a transformer at 1.0 kV. The current in the secondary coil is 0.50 A. What is the turns ratio of the transformer? Is it a step-up or a step-down transformer?

eBook*plus* RESOURCES

Explore more with this weblink: Transformer applet

6.8 Energy distribution

6.8.1 High-voltage transmission lines

Power stations are usually situated large distances from cities where most of the consumers are located. This presents problems with power losses in the transmission lines. Transmission lines are essentially long metallic conductors that have significant resistance. This means that they have a significant voltage drop across them when they carry a large current. This could result in greatly decreased voltages available to the consumer, as illustrated in 6.8 sample problem 1.

The resistance of a metallic conductor is proportional to its resistivity, ρ, its length, l, and is inversely proportional to its cross-sectional area, A:

$$R = \frac{\rho l}{A}.$$

The voltage drop, V, across a conductor equals the current, I, multiplied by the resistance, R. That is:

$$V = IR.$$

The rate of energy transfer in a conductor is called power, P, where:

$$P = VI.$$

If you know the current through a conductor and its resistance, the previous equation becomes:

$$P = I^2R.$$

Therefore, the power lost in a transmission line is given by the formula:

$$P_{\text{loss}} = I^2R$$

where

I = current flowing through the transmission line
R = the resistance of the transmission line.

6.8 SAMPLE PROBLEM 1

TRANSMISSION LINE CALCULATIONS

A power station generates electric power at 120 kW. It sends this power to a town 10 km away through transmission lines that have a total resistance of 0.40 Ω. If the power is transmitted at 240 V, calculate:

(a) the current in the transmission lines
(b) the voltage drop across the transmission lines
(c) the voltage available in the town
(d) the power loss in the transmission lines.

SOLUTION:

(a) For this calculation, use the station's power and the voltage across the transmission lines.

$$\begin{aligned} P &= VI \\ \Rightarrow I &= \frac{P}{V} \\ &= \frac{120\,000}{240} \\ &= 500\text{ A} \end{aligned}$$

(b) $V = IR$
$= 500 \times 0.40$
$= 200\text{ V}$

(c) $V_{\text{town}} = V_{\text{station}} - V_{\text{lines}}$
$= 240 - 200$
$= 40\text{V}$

(d) $P_{\text{loss}} = I^2R$
$= (500)^2 \times 0.40$
$= 100\,000\text{ W}$
$= 100\text{ kW}$

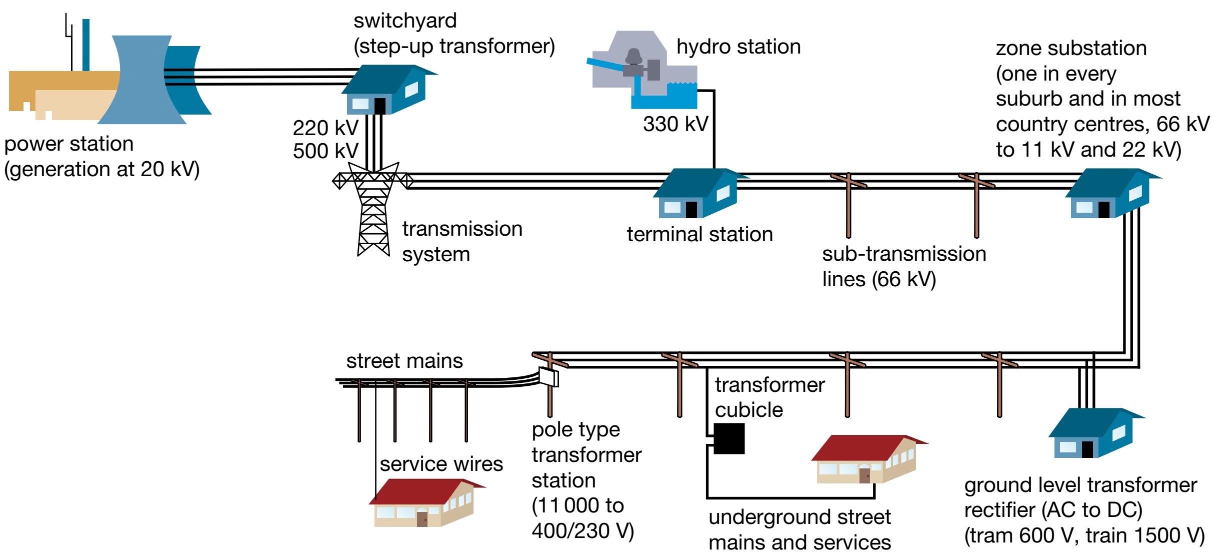
6.8.2 Using transformers to reduce power loss

Sample problem 1, above, demonstrates the difficulties involved in transmitting electrical energy at low voltages over large distances. The solution is to use transformers to step up the voltage before transmission. If the voltage is increased, the current is reduced. Recall that the power lost in transmission lines is given by the formula:

$$P_{\text{loss}} = I^2R.$$

If the transmission voltage is doubled, the current is halved and the power loss is reduced by a factor of four. If the current is reduced by a factor of 10, the power loss is reduced by a factor of 100, and so on.

FIGURE 6.24 Representation of a power distribution system.

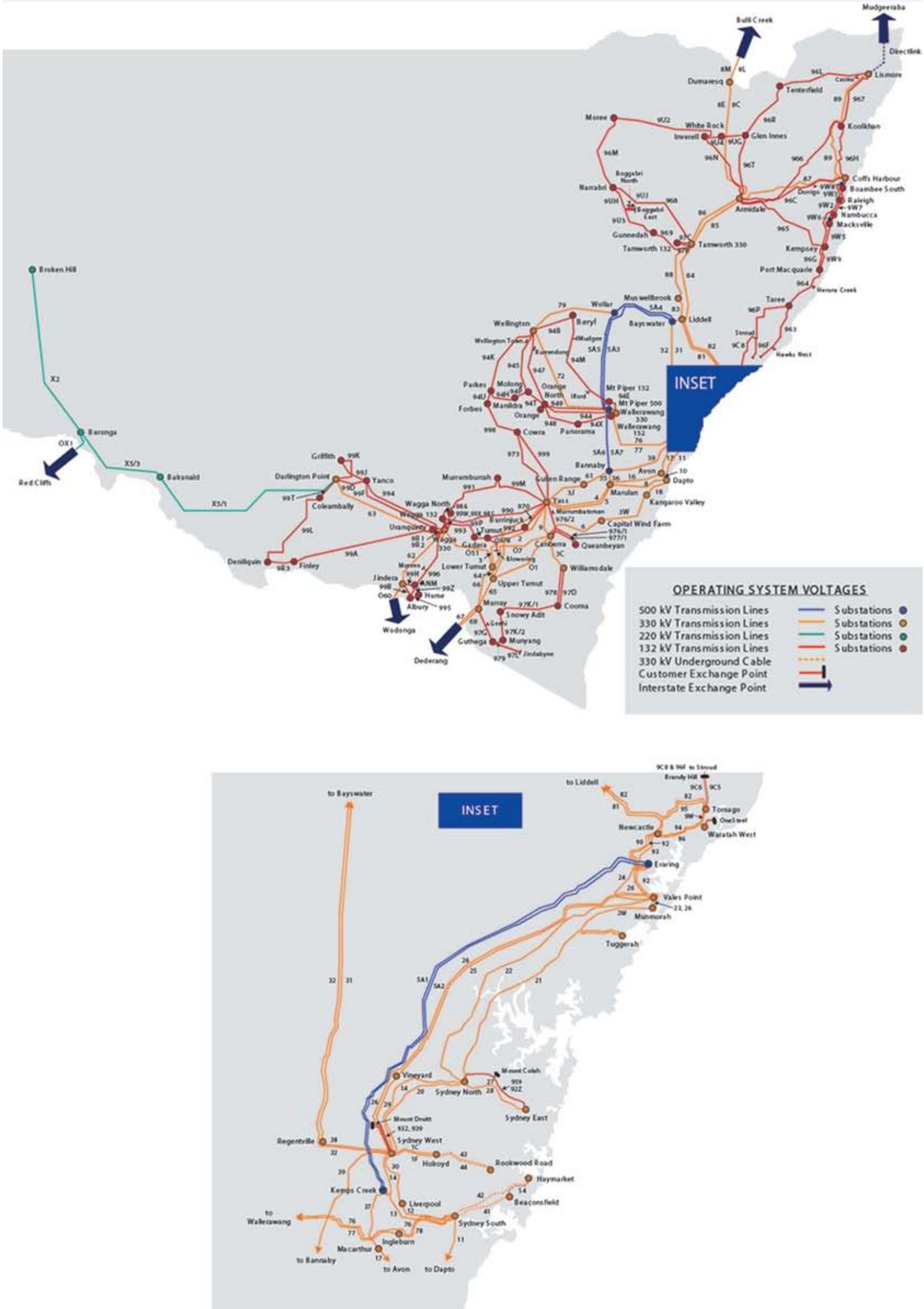


Using transformers enables electricity to be supplied over large distances without wasting too much electrical energy. This has had a significant effect on society. If transformers were not used in the power distribution system, either power stations would have to be built in the cities and towns or the users of electricity would have to be located near the power stations. The latter would mean that industries and population centres would have to be located near the energy sources such as hydro-electric dams and coal mines. The former would mean that fossil fuel stations would dump their pollution on the near-by population centres.

6.8.3 NSW electrical distribution system

The New South Wales electrical distribution system is shown in Figure 6.25. The Bayswater power station in the Hunter Valley has four 660-megawatt generators that each have an output voltage of 23 kV. (Each set of coils in the generators has an output of 220 MW.) The three-phase power then enters a transmission substation where transformers step up the voltage to 330 kV.

FIGURE 6.25 The New South Wales electrical supply network.



The transmission lines end at a terminal station where the voltage is stepped down to 66 kV for transmission to zone power substations like the one shown in Figure 6.26. There is usually a zone substation in each regional centre and in each municipality of a city. Here the voltage is stepped down to values of 11 kV and 22 kV. Power substations can perform three tasks:

1. step down the voltage using transformers
2. split the distribution voltage to go in different directions
3. enable, using circuit breakers and switches, the disconnection of the substation from the transmission grid or sections of the distribution grid to be switched on and off.

FIGURE 6.26 A transformer at a substation.

Finally, pole transformers, as shown in Figure 6.27, step the voltage down to 415 V for industry and 240 V for domestic consumption.

FIGURE 6.27 A pole transformer.

PHYSICS IN FOCUS

Protecting power transmission lines from lightning

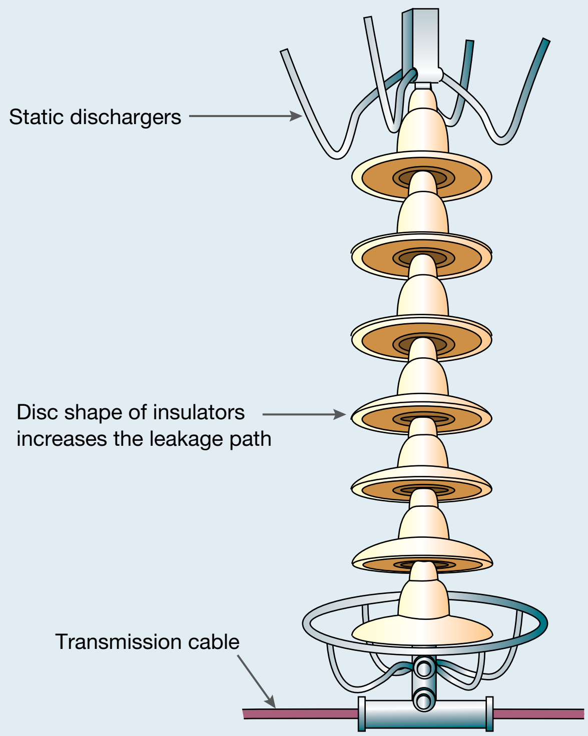
When lightning strikes, it will usually pass between the bottom of a thundercloud and the highest point on the Earth below. This means that it will strike tall trees, the tops of buildings such as church spires and the metal power towers used to support high voltage power transmission lines. Many such power towers have a cable running between them known as the continuous earth line. This cable normally carries no current, but it may carry a current if a fault develops in the system. A second function of this cable is that it acts as a continuous lightning conductor. If this cable or a tower is struck by lightning, the electricity of the lightning will be conducted to the Earth by the metal towers, and the transmission lines will not suffer from a sudden surge of voltage that could damage substations.

PHYSICS IN FOCUS

Insulating transmission lines

In dry air, sparks can jump a distance of 1 cm for every 10 000 V of potential difference. Therefore, a 330 kV line will spark to a metal tower if it comes within a distance of 33 cm. In high humidity conditions, the distance is larger. To prevent sparks jumping from transmission lines to the metal support towers, large insulators separate them from each other. It is important that these insulators are strong and have high insulating properties. Suspension insulators, illustrated in Figure 6.28, are used for all high voltage power lines operating at voltages above 33 kV, where the towers or poles are in a straight line. Note that the individual sections of the insulators are disc shaped. This is because dust and grime collect on the insulators and can become a conductor when wet. Many wooden poles catch fire after the first rain following a prolonged dry period because a current flows across wet dirty insulators. The disc shape of the insulator sections increases the distance that a current has to pass over the surface of the insulator and so decreases the risk. There is also less chance that dirt and grime will collect on the undersides of the sections, and these are also less likely to get wet.

FIGURE 6.28 Suspension insulator used for high voltage transmission lines.



eBook*plus* RESOURCES

Model energy loss over a distance using this spreadsheet: Modelling power transmission
Searchlight ID: doc-0040

6.9 Review

6.9.1 Summary

- Magnetic flux, ϕ_B, is the amount of magnetic field passing through a given area. It depends on the strength of the field, B, as well as the area, A.
- The magnetic flux through a coil is the product of the area, A, of the coil and the component of the magnetic field strength, B, that is perpendicular to the area: $\phi = BA$.
- Magnetic field strength is also known as magnetic flux density.
- Faraday's Law of Induction states that a changing magnetic flux through a circuit induces an emf in the circuit.
- The magnitude of an induced emf depends on the rate of change of magnetic flux through a circuit.
- Lenz's Law states that an induced emf always gives rise to a current that creates a magnetic field that, in turn, opposes the original change in flux through the circuit.
- Eddy currents are created when there is relative movement between a magnetic field and a metal object. The area of the magnetic field, however, does not cover the whole of the metal object. Eddy currents are also created when a conducting material is in the presence of a changing magnetic field.

- A changing magnetic flux through a coil can induce a voltage across the terminals of the coil.
- Transformers are devices that can convert an AC input voltage signal to a higher or lower AC output voltage.
- A transformer consists of a primary and secondary coil, usually linked by a soft iron core.
- The following equations apply to ideal transformers:
  - $\dfrac{V_p}{V_s} = \dfrac{N_p}{N_s}$
  - $V_p I_p = V_s I_s$
- Eddy currents increase the temperature of metal objects.
- When a metal object is moving relative to a region affected by a magnetic field, the region of magnetic field exerts a force on the eddy currents that opposes the relative motion of the object to the field.
- Losses in transmission lines can be calculated using the formula $P_{loss} = I^2R$.
- Power losses in transmission lines can be reduced by using step-up transformers to increase the transmission voltage, thereby reducing the transmission current.

6.9.2 Questions

1. Explain how Michael Faraday was able to produce an electrical current using a magnet.
2. Define the concept of magnetic flux in terms of magnetic flux density and surface area.
3. Calculate the magnetic flux threading (or passing through) the areas in the following cases.
   (a) An area of 1.5 m^2 is perpendicular to a magnetic field of flux density 2.0 T.
   (b) An area of 0.75 m^2 is perpendicular to a magnetic field of strength 0.03 T.
   (c) A rectangle with a length of 4.0 cm and width 3.0 cm is perpendicular to a magnetic field of flux density 5.0×10^{-3} T.
   (d) A circle of radius 7.0 cm that is parallel with a magnetic field of flux density 5.0×10^{-3} T.
4. Evaluate the direction of the induced current through the galvanometer in each of the galvanometer circuit coils shown in Figure 6.29. The arrows represent the motion of the coil or the magnet.
5. Describe the effect that the speed of movement of a magnet has on the magnitude of a current induced in a coil.
6. A magnet moving near a conducting loop induces a current in the circuit as shown in Figure 6.30. The magnet is on the far side of the loop and is moving in the direction indicated by the dotted line. Describe two ways in which the magnet can be moving to induce the current as shown.

FIGURE 6.29

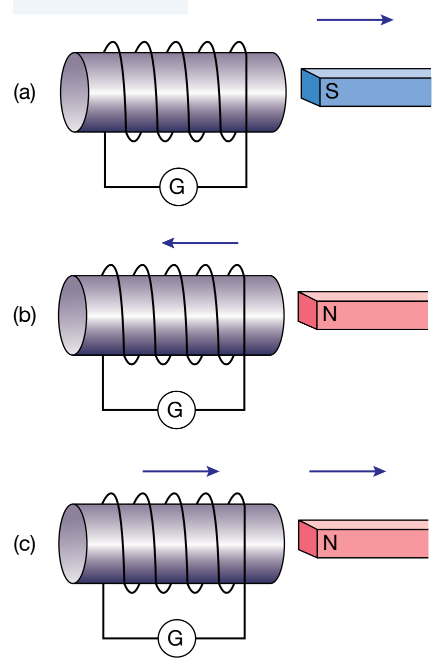
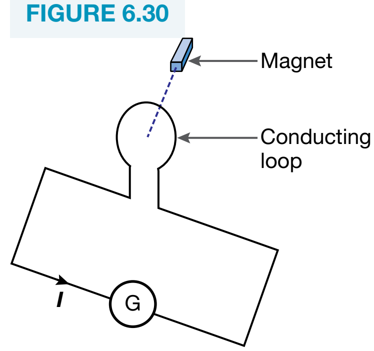
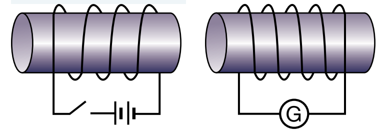


FIGURE 6.30



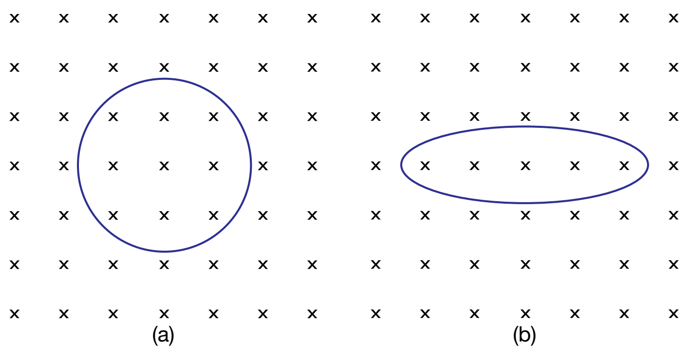
7. Deduce the direction of the induced current through the galvanometer in Figure 6.31 when:
   (a) the switch is closed
   (b) the switch remains closed and a steady current flows in the battery circuit
   (c) the switch is opened.

FIGURE 6.31



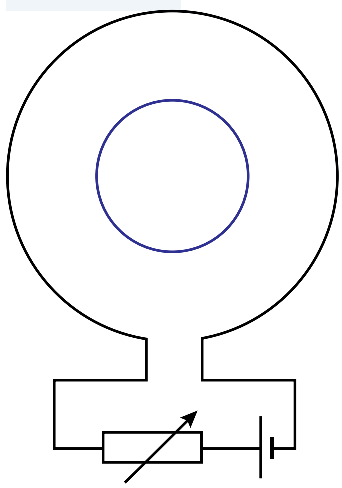
8. A flexible metal loop is perpendicular to a magnetic field as shown in Figure 6.32a. It is distorted to the shape shown in Figure 6.32b. Is the direction of the induced current in the loop clockwise or anticlockwise? Explain your answer.

FIGURE 6.32



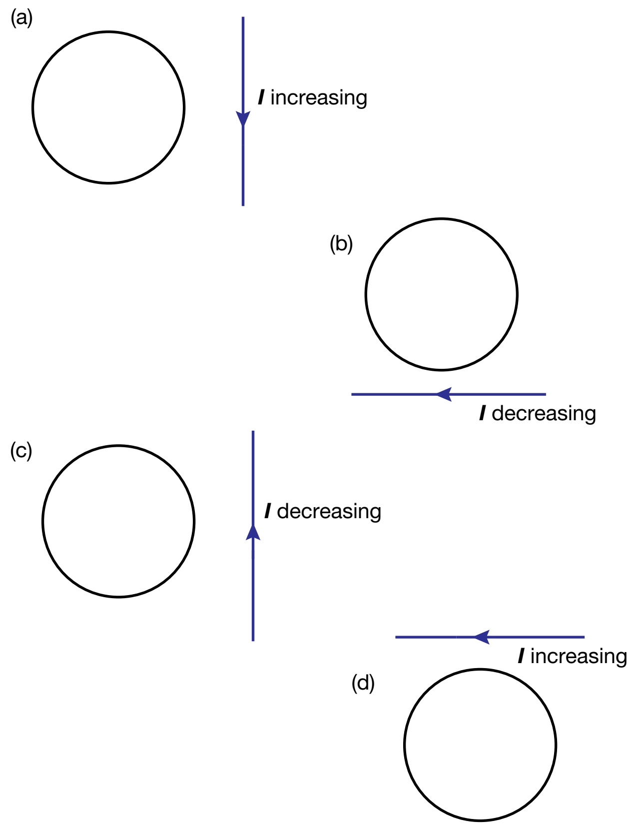
9. Figure 6.33 shows a loop of wire connected in series to a source of emf and a variable resistor. Describe the direction of the induced current in the central loop when the resistance of the outer loop circuit is increasing. Explain your reasoning.

FIGURE 6.33



10. In what direction, clockwise or anticlockwise, is the induced current in the loop of wire in each situation shown in Figure 6.34?

FIGURE 6.34

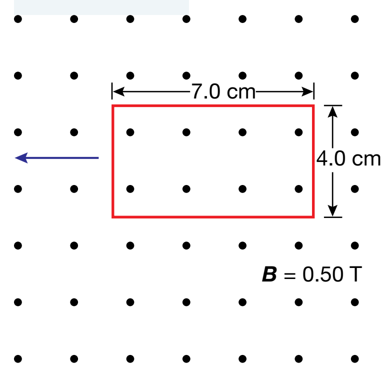


11. A metal rectangle has a length of 7.0 cm and a width of 4.0 cm. It is initially at rest in a uniform magnetic field of strength 0.50 T as shown in Figure 6.35.

    The rectangle is completely removed from the magnetic field in 0.28 s.

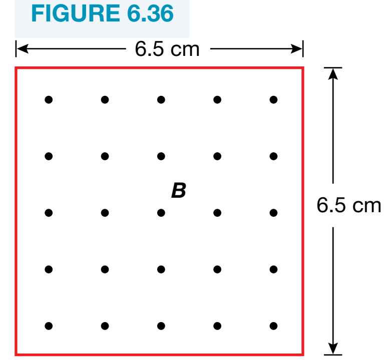
    (a) What is the initial magnetic flux through the rectangle?
    (b) In what direction, clockwise or anticlockwise, is the induced current in the rectangle when it is being removed from the magnetic field?

FIGURE 6.35



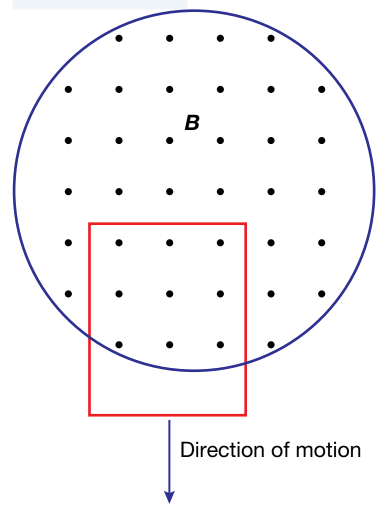
12. A square loop of wire has sides of length 6.5 cm. The loop is sitting in a magnetic field of strength 1.5×10^{-3} T as shown in Figure 6.36. The magnetic field is reduced to 0 T in a period of 5.0 ms.
    (a) What is the flux through the loop initially?
    (b) What will be the effect on the induced emf if a 25-loop coil is used, rather than a single loop?
    (c) In what direction will the current flow in the loop?

FIGURE 6.36



13. Use diagrams to show three ways to change the flux passing through a conductor loop.
14. (a) Explain the production of a back emf in an electric motor.
    (b) Describe how the back emf of an electric motor is produced.
    (c) Explain why the back emf in an electric motor opposes the supply emf.
    (d) How does the back emf determine the maximum rotating speed of an electric motor?
    (e) How can overloading an electric motor cause it to burn out?
15. The armature winding of an electric motor has a resistance of 5.0 Ω. The motor is connected to a 240 V supply. When the motor is operating with a normal load, the back emf is equal to 237 V.
    (a) Calculate the current that passes through the motor when it is first started.
    (b) Calculate the current that passes through the motor when it is operating normally.
16. The armature winding of an electric drill has a resistance of 10 Ω. The drill is connected to a 240 V supply. When the drill is operating normally, the current drawn is 2.0 A.
    (a) Calculate the current that passes through the drill when it is first started.
    (b) Calculate the back emf of the drill when it is operating normally.
17. (a) What is an eddy current?
    (b) Discuss how eddy currents are produced.
    (c) Describe how eddy currents raise the temperature of metals.
18. A rectangular sheet of aluminium is pulled from a magnetic field as shown in Figure 6.37.
    (a) Copy the diagram and indicate the position and direction of an eddy current loop.
    (b) Which way does the force due to the external magnetic field and the eddy current act on the aluminium sheet?

FIGURE 6.37



19. Discuss how eddy currents are utilised in the following situations:
    (a) induction electric cooktops
    (b) electromagnetic braking of trains or trams
    (c) induction furnaces.
20. Draw a labelled cross-section diagram of a simple transformer. Referring to your diagram, explain how it operates in terms of the principles of electromagnetic induction.
21. Explain why a steady DC current input will not operate a transformer.
22. A transformer changes 240 V to 15 000 V. There are 4000 turns on the secondary coil.
    (a) Identify what type of transformer this is.
    (b) Calculate how many turns there are on the primary coil.
23. A doorbell is connected to a transformer that has 720 turns in the primary coil and 48 turns in the secondary coil. If the input voltage is 240 V AC, calculate the voltage that is delivered to the doorbell.
24. A school power pack that operates from a 240 V mains supply consists of a transformer with 480 turns on the primary coil. It has two outputs, 2 V AC and 6 V AC.
    (a) Calculate the number of turns on the 2 V and 6 V secondary coils.
    (b) The power rating of the transformer is 15 W. Calculate the maximum current that can be drawn from the 6 V secondary coil.
25. An ideal transformer has 100 turns on the primary coil and 2000 turns on the secondary coil. The primary voltage is 20 V. The current in the secondary coil is 0.50 A. Calculate:
    (a) the secondary voltage
    (b) the output power
    (c) the input power
    (d) the current flowing through the primary coil.

26. A transformer has 110 turns on the primary coil and 330 turns on the secondary coil.
 (a) Identify this type of transformer.
 (b) By what factor does it change the voltage?
27. An ideal transformer is designed to provide a 9.0 V output from a 240 V input. The primary coil is fitted with a 1.0 A fuse.
 (a) Calculate the ratio

$$\frac{\text{number of turns on the primary coil}}{\text{number of turns on the secondary coil}}.$$

 (b) Calculate the maximum current that can be delivered from the output terminals.
28. A neon sign requires 12 kV to operate. Calculate the ratio

$$\frac{\text{number of turns on the primary coil}}{\text{number of turns on the secondary coil}}$$

of the sign's transformer if it is connected to a 240 V supply.
29. A 20.0 W transformer gives an output voltage of 25 V. The input current is 15 A.
 (a) Calculate the input voltage.
 (b) Is this a step-up or step-down transformer?
 (c) Calculate the output current.
30. Describe the advantages gained by transmitting AC electrical power at high voltages over large distances.
31. A power station generates electric power at 120 kW. It sends this power to a town 10 kilometres away through transmission lines that have a total resistance of 0.40 Ω. If the power is transmitted at 5.00×10^5 V, calculate:
 (a) the current in the transmission lines
 (b) the voltage drop across the transmission lines
 (c) the voltage available at the town
 (d) the power loss in the transmission lines.
32. A generator coil at the Bayswater power station produces 220 MW of power in a single phase at 23 kV. This voltage is stepped up to 330 kV by a transformer in the transmission substation. Calculate:
 (a) the ratio

$$\frac{\text{number of turns on the primary coil}}{\text{number of turns on the secondary coil}}$$

 for the transformer
 (b) the output power of the transformer
 (c) the current in the transmission line.
33. A generator has an output of 20 kW at 4.0 kV. It supplies a factory via two long cables with a total resistance of 16 Ω.
 (a) Calculate the current in the cables.
 (b) Calculate the power loss in the cables.
 (c) Calculate the voltage between the ends of the cables at the factory.
 (d) Describe how the power supplied to the workshop could be increased.

eBookplus RESOURCES

- **Complete this digital doc: Investigation:** Inducing current in a coiled conductor
  Searchlight ID: doc-26589
- **Complete this digital doc: Investigation:** Linking coils
  Searchlight ID: doc-26590
- **Complete this digital doc: Investigation:** The direction of induced currents
  Searchlight ID: doc-26591
- **Complete this digital doc: Investigation:** Making a simple transformer
  Searchlight ID: doc-26592
- **Complete this digital doc: Investigation:** Transformer ins and outs
  Searchlight ID: doc-26593
- **Complete this digital doc: Investigation:** Transmission line power losses
  Searchlight ID: doc-26594
- **Complete this digital doc: Investigation:** Making a simple transformer using a kit
  Searchlight ID: doc-26595

PRACTICAL INVESTIGATIONS

Investigation 6.1 Inducing current in a coiled conductor

Aim

(a) To study ways of inducing a current in a coiled conductor
(b) To study factors affecting the size of the induced current.

Apparatus

galvanometer
two coils having different numbers of turns of wire
two bar magnets, of different strengths, if possible
an iron core that fits into one of the coils. This could be made by taping large iron nails together.
connecting wires

Theory

You will be reproducing some of Faraday's experiments. Modern galvanometers are much more sensitive than those available to Faraday.

Method

1. Connect the coil with the fewest number of turns to the galvanometer. Push the N pole of a bar magnet into the coil. Describe what happens.
2. Hold the magnet stationary near the coil. Describe what happens.
3. Withdraw the N pole from the coil. Describe what happens.
4. Repeat steps 1 and 3 at different speeds. Describe what happens.
5. Hold the magnet stationary and move the coil in different directions. Rotate the coil so that first one end and then the other approaches the magnet. Describe what happens.
6. Place the iron core in the coil and touch it with the N pole. Remove the magnet. Describe what happens.
7. Design an experiment to examine the factors that affect the size of the induced current.

Analysis

1. Relate your results to Faraday's law of electromagnetic induction.
2. Describe how you would make a generator to create a relatively large current.

Investigation 6.2 Linking coils

Aim

To see if the magnetic field of a current-carrying coil can induce a current in another coil.

Apparatus

galvanometer
two coils having different numbers of turns of wire, preferably one of which fits into the other

an iron core that fits into the smaller of the coils. This could be made by taping large iron nails together.
a $10\,\Omega$ resistor
a power supply
connecting wires

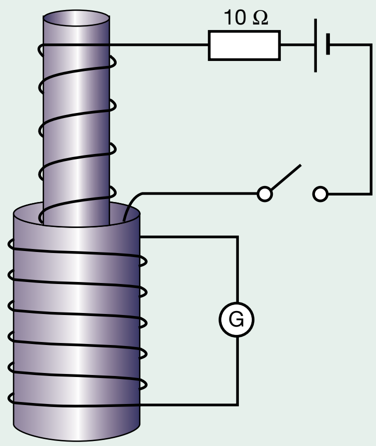
Theory

A current flowing in a coil will create a magnetic field that threads the second coil. When there is a changing magnetic field threading the second coil, an emf will be induced in the coil.

Method

1. Set up the apparatus as shown in Figure 6.38. Set the power supply to $2.0\,$V.

FIGURE 6.38



2. Close the switch and observe the effects on the galvanometer.
3. Open the switch and observe the effects on the galvanometer.
4. Put an iron core in the smaller coil and repeat steps 2 and 3.

Analysis

Relate your results to Faraday's law of electromagnetic induction.

Questions

1. When was a current induced in the secondary (galvanometer) coil?
2. What happened when a steady current was flowing in the primary (power supply) coil?
3. What effect did the iron core have on the induced current?

Investigation 6.3 The direction of induced currents

Aim

To learn how to predict the direction of an induced current.

Apparatus

galvanometer
coil
bar magnet
connecting wires
battery

Theory

Lenz's Law states that the direction of an induced current in a coil is such that the magnetic field that it establishes opposes the change of the original flux threading the coil.

Method

1. Carefully examine the coil to see which way the wire is coiled around the cylinder.
2. Use a battery to establish which way the galvanometer deflects when currents flow through it in different directions.

3. Connect the coil to the galvanometer.
4. Push the N pole of the magnet towards the coil. Note the deflection of the galvanometer. Answer the following questions.
   - In which direction did the current flow in the coil?
   - Was the end of the coil nearest the magnet a N or S pole?
   - Does the magnetic field of the coil assist or oppose the motion of the magnet?
   - Does the magnetic field of the coil assist or oppose the change in flux threading the coil?
5. Pull the N pole away from the coil. Answer the questions of step 4.
6. Push the S pole of the magnet towards the coil. Answer the questions of step 4.
7. Pull the N pole away from the coil. Answer the questions of step 4.

Analysis
Do your results verify Lenz's Law? Explain.

Investigation 6.4 Making a simple transformer

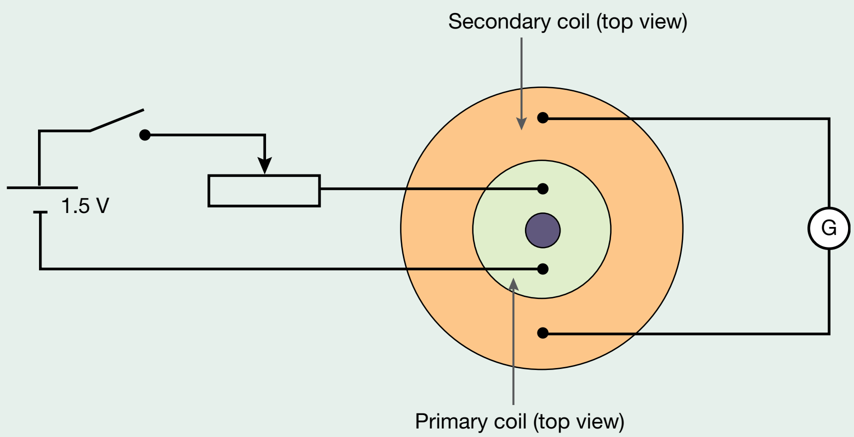
Aim
(a) To set up a simple transformer
(b) To observe that a steady DC current does not produce an output current from a transformer.

Apparatus
galvanometer
two coils having different numbers of turns, one coil fitting inside the other
1.5 V cell
switch
variable resistor
connecting wires

Theory
A transformer consists of a primary and secondary coil. In this experiment the primary coil fits inside the secondary coil. The changing flux produced in the primary coil induces a current in the secondary coil.

Method
1. Place the smaller (primary) coil in the larger (secondary) coil. Connect the secondary coil to the galvanometer. Connect the primary coil in series with the switch, variable resistor and cell as shown in Figure 6.39.

FIGURE 6.39



2. Set the variable resistor to its lowest value.
3. Observe the effects on the galvanometer as you close the switch, keep it closed for five seconds and then open the switch.
4. Describe what happens.
5. Close the switch and change the value of the variable resistor slowly and rapidly. Open the switch.
6. Record your observations.

Analysis

1. Relate your observations to Faraday's Law of Electromagnetic Induction.
2. Describe the conditions necessary for the operation of a transformer.

Investigation 6.5 Transformer ins and outs

Aim

To use an AC input voltage to produce an AC output voltage and to compare their values.

Apparatus

two coils having a different number of turns of wire, preferably with the number of turns on each being known, one coil fitting inside the other
two 2 V light globes in holders
AC power source
switch
dual trace cathode ray oscilloscope
connecting wires
long iron nails that fit inside the smaller coil to produce a soft iron core

Theory

An AC input voltage will induce an AC output voltage.

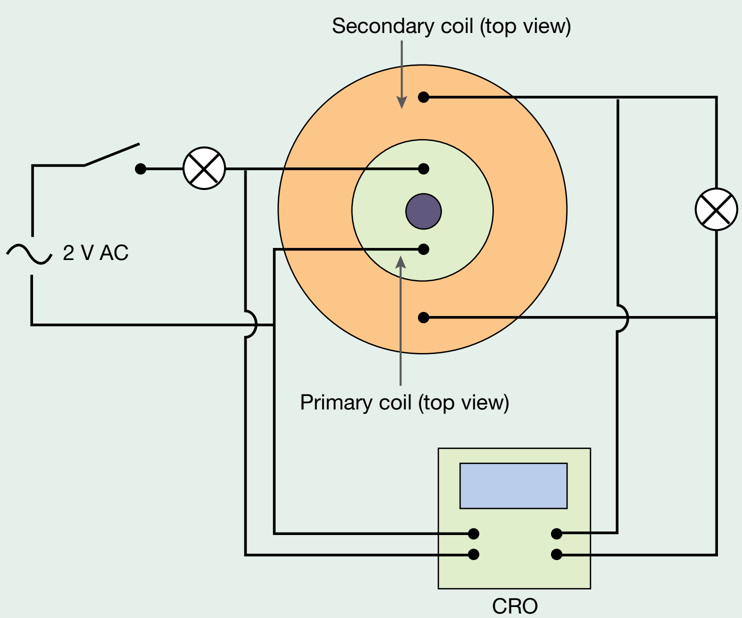
If the secondary coil has more turns than the primary, the device is a step-up transformer and the secondary voltage will be greater than the primary voltage.

The inclusion of an iron core increases the effectiveness of the transformer.

Method

1. Place the smaller (primary) coil in the other (secondary) coil.
2. Connect the secondary coil to a light globe.
3. Connect the primary coil in series with the switch, a light globe and the AC source.
4. Adjust the AC source to its lowest value (less than 2 V).
5. Connect one set of input leads of the CRO across the terminals of the primary coil, and the other across the terminals of the secondary coil. The set-up is illustrated in Figure 6.40.

FIGURE 6.40



6. Close the switch and observe the traces on the CRO.
7. Compare the primary and secondary peak voltages and frequencies.
8. Insert iron nails in the primary coil and repeat steps 6 and 7.

Analysis

- Explain your observations in terms of the theory you have studied.
- What effect did the insertion of the iron nails have on the effectiveness of the transformer?

Investigation 6.6 Transmission line power losses

Aim

(a) To investigate the effects of resistance in transmission lines
(b) To investigate the use of transformers in power distribution systems.

Apparatus

Transmission line experiment (Available from Haines Educational P/L, www.haines.com.au)

Theory

Power losses in transmission lines can be reduced by using transformers to step up the voltage for transmission and step it down again for use. This kit models transmission lines by using resistance wire.

Method

1. Set up the equipment as described in the instruction brochure.
2. Transmit power from an AC supply to the load globe using the transmission lines alone. Note the brightness of the globe and the current in the transmission lines.
3. Measure the voltage output of the supply and the voltage at the globe.
4. Repeat steps 1 and 2, this time using the transformers.

Analysis

Comment on your results.

Investigation 6.7 Making a simple transformer using a kit

Aim

(a) To set up a simple transformer
(b) To observe that a steady DC current does not produce an output current from a transformer.

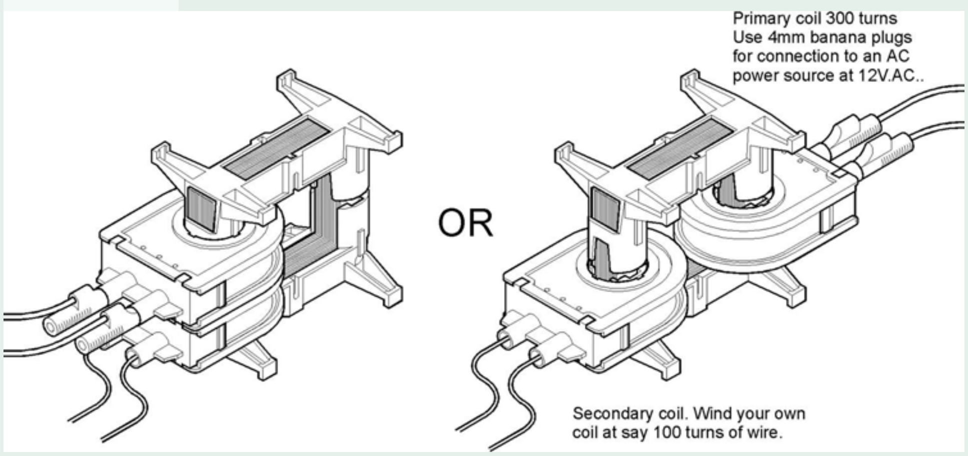
Apparatus

- ‘Hodson’ Induction Kit Cat: EM1973-001 and instruction booklet
  - 300 turn coil
  - 150 turn coil
  - Rubber bands
  - Connecting cables
- School laboratory AC Power Supply
- AC voltmeter

Theory

A transformer consists of a primary and secondary coil. In this experiment you will construct a transformer in two configurations.

FIGURE 6.41



Method

Follow the instructions from page 15 of the kit's instruction manual (see 'The transformer' below).
Measure and record the input voltage and output voltage.

Analysis

Describe how a transformer operates.
Does your transformer produce results consistent with the rule for ideal transformers?

$$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$

The Transformer

Take the 'U' core in the plastic holder and 'I' core in the plastic holder. Place the 300 turn primary coil over one leg of the 'U' core and place a secondary coil that you have wound yourself over the other leg of the 'U' core.

Invert the 'I' core and carefully place it over the 'U' core as shown below.

TAKE CARE NOT TO CATCH AND BEND THE THIN LAMINATIONS OF THE 'U' CORE AS YOU PLACE THE 'I' CORE ON TO IT.

The 4x rubber bands can be stretched between the legs, top to bottom, to hold the two halves of the transformer firmly together.

Apply say 12 volts AC to the 300 turn primary coil and, using an AC voltmeter, measure the output voltage from the 150 turn secondary coil that you wound yourself. It should be close to 6 V.AC.

Two coils can be fitted to either side of the transformer core and they can face from the ends of the iron core or across the iron core. It does not matter which are the primary and which are the secondary coils because the iron circuit passes through them on either side of the transformer, but the 300 turn coil supplied is normally used as the primary coil and is connected to 12 Volt AC power source with the cables supplied with the banana plugs fitted.

For transformer study, normally the iron core is fitted and the iron is closed tightly so there is no air space between the iron laminations and therefore minimum magnetic leakage. The rubber bands supplied pull the iron tightly together. If the iron core is not fully closed tightly, the voltages measured on the secondary coils will not accurately follow transformer theory.

TOPIC 7
Applications of the motor effect

7.1 Overview

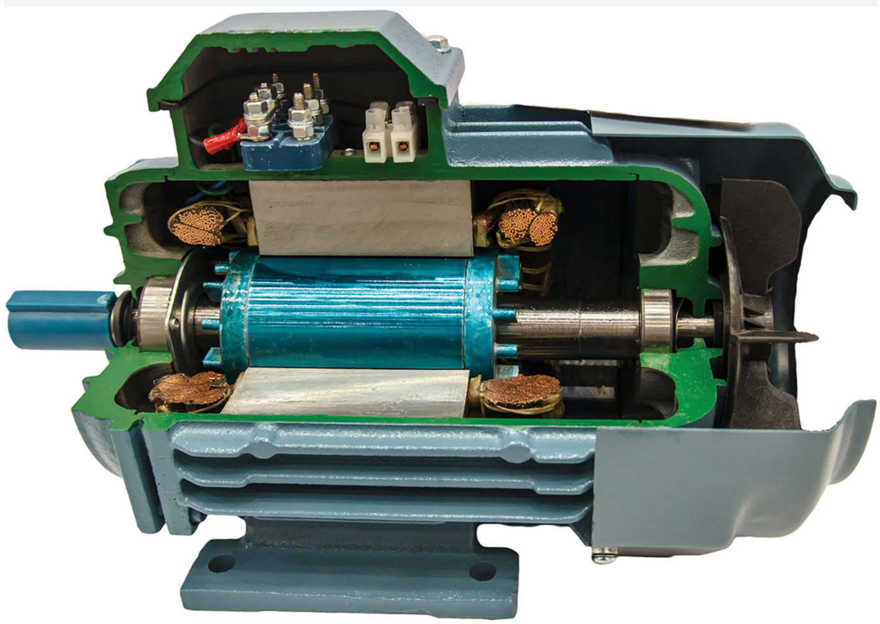
7.1.1 Module 6: Electromagnetism

Applications of the motor effect

Inquiry question: How has knowledge about the motor effect been applied to technological advances?
Students:

- investigate the operation of a simple DC motor to analyse:
  - the functions of its components
  - production of a torque: $\tau = nIA_{\perp}B = nIAB\sin\theta$
  - effects of back emf (ACSPH108)
- analyse the operation of simple DC and AC generators and AC induction motors (ACSPH110)
- relate Lenz's Law to the law of conservation of energy and apply the law of conservation of energy to:
  - DC motors and
  - magnetic braking

FIGURE 7.1 A cutaway view of the components of an electric motor.



7.2 Simple DC electric motors

7.2.1 Function of the components

An electric motor is a device that transforms electrical potential energy into rotational kinetic energy. Electric motors produce rotational motion by passing a current through a coil in a magnetic field. Electric motors that operate using direct current (DC) are discussed in this section. The operation of electric motors that use alternating current (AC) is discussed in Section 7.7.

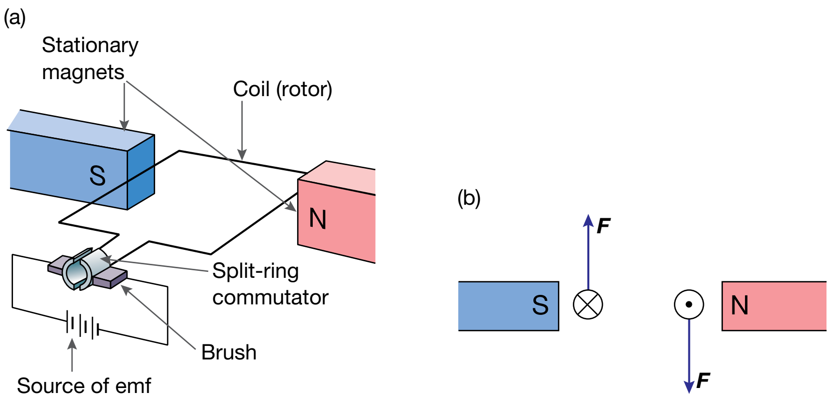
Anatomy of a motor

A simplified diagram of a single-turn DC motor is shown in Figure 7.2 (which shows only the parts of the DC motor that produce rotational motion).

The magnets provide an external magnetic field in which the coil rotates. As the magnets are fixed to the casing of the motor and are stationary, they are known as the **stator**. The stator sometimes consists of a pair of electromagnets.

The coil carries a direct current. In Figure 7.2 the coil has only one loop of wire and this is shown with straight sides. This makes it easier to visualise how forces on the sides come about and to calculate the magnitudes of forces. The coil is wound onto a frame known as an **armature**. This is usually made of ferromagnetic material and it is free to rotate on an axle. The armature and coil together are known as the rotor. The armature axle protrudes from the casing, enabling the movement of the coil to be used to do work.

FIGURE 7.2 (a) The functional parts of a simplified electric motor (b) The direction of current flow in the coil and the direction of the forces acting on the sides.



The force acting on the sides of the coil that are perpendicular to the magnetic field can be calculated using the previously discussed formula for calculating the force on a current-carrying conductor in a magnetic field:

$$F = BIl \sin \theta.$$

Real motor rotors have many loops or turns of wire on them. If the coil has n turns of wire on it, then these sides experience a force that is n times greater. In this case:

$$F = nBIl \sin \theta.$$

This extra force increases the torque acting on the sides of the coil.

The split-ring **commutator** and the brushes form a mechanical switch that changes the direction of the current through the coil every half turn so that the coil continues rotating in the same direction. The operation of the commutator is discussed in a later section of this topic.

The source of emf (electromotive force), for example a battery, drives the current through the coil.

How a DC motor operates

Figure 7.3 shows the simplified DC motor at five positions throughout a single rotation. The coil has been labelled with the letters K, L, M and N so that it is possible to observe the motion of the coil as it completes one rotation.

In Figure 7.3a, the side LK has a force acting on it that is vertically upwards. Side MN has a force of equal magnitude acting on it that is vertically downwards. In this position, the forces acting on the sides are perpendicular to the line joining the axle (the pivot line) to the place of application of the force. This means that the torque (see Section 7.3) acting on the coil is at its maximum value. Note that the current is flowing in the direction of K to L.

In Figure 7.3b, the side LK still has a force acting on it that is vertically upwards. Similarly, side MN still has a force of equal magnitude acting on it that is vertically downwards. In this position the forces acting on the sides are almost parallel to the line joining the axle (the pivot line) to the place of application of the force. This means that the torque acting on the coil is almost zero. It is just after this position that the commutator changes the direction of the current through the coil. The momentum of the coil keeps the coil rotating even though the torque is very small.

FIGURE 7.3 Forces acting on the sides of a current-carrying loop. The lower part of each diagram shows cross-sections of the coil.

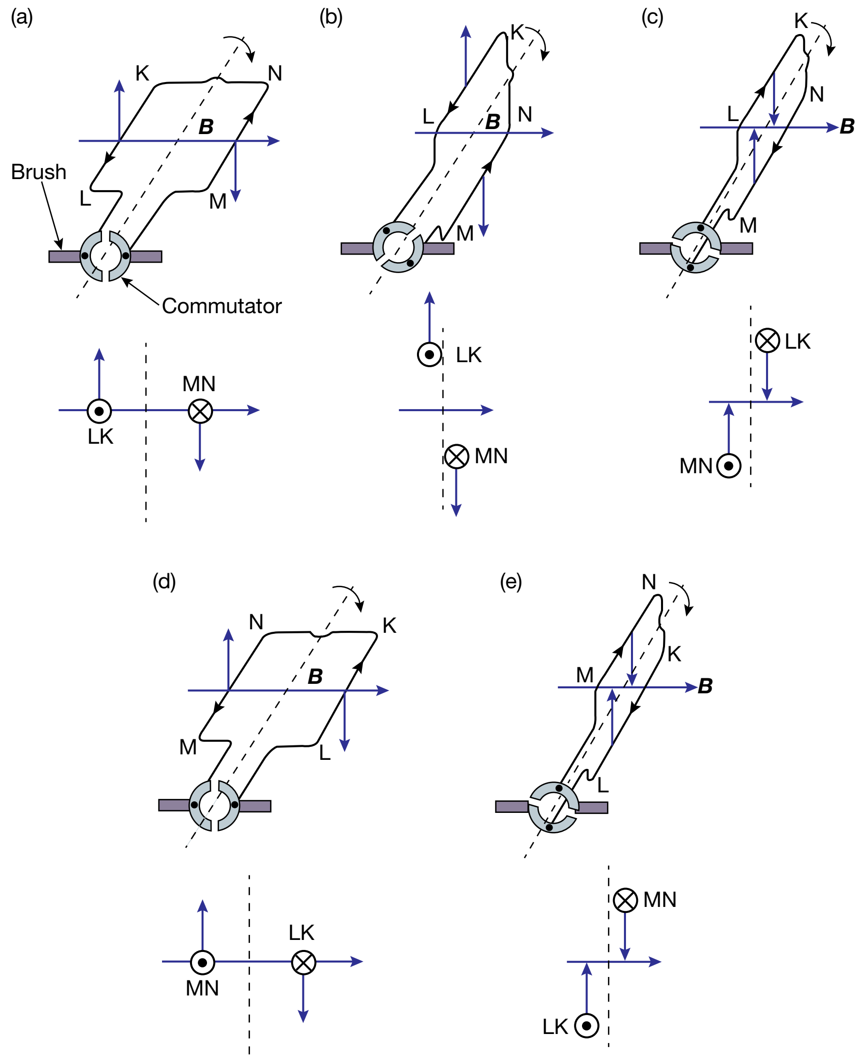


Figure 7.3c shows the situation when the coil has moved a little further than in the previous diagram, and the current direction through the coil has been reversed. The force acting on side LK is now downwards and the force acting on side MN is now upwards. This changing of direction of the forces and the momentum of the coil enable the coil to keep rotating in the same direction. If the current through the coil did not change its direction of flow through the coil, the coil would rock back and forth about this position. Note that the current is now flowing in the direction of L to K and the torque acting on the coil is still clockwise.

Figure 7.3d shows the position of the coil when the torque is again at a maximum value. In this case side MN has the upward force acting on it.

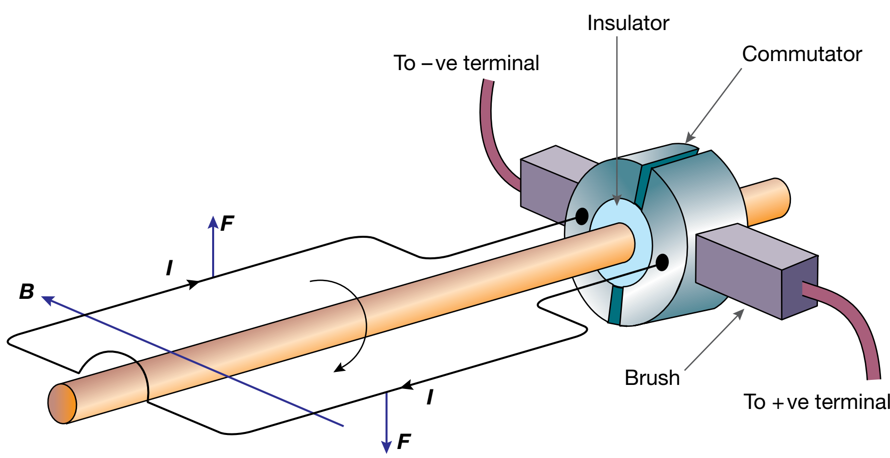
Figure 7.3e shows the position of the coil when the torque is again virtually zero and the current has again been reversed. Note that the current is again flowing in the direction of K to L and that there is still a clockwise torque acting on the coil.

The magnitude of the forces acting on sides LK and MN remained constant throughout the rotation just described. However, the torque acting on the coil changed in magnitude.

Commutators

The commutator is a mechanical switch that automatically changes the direction of the current flowing through the coil when the torque falls to zero. Figure 7.4 provides a close-up look at a commutator. It consists of a **split metal ring**, each part of which is connected to either end of the coil. As the coil rotates, first one ring and then the other make contact with a brush. This reverses the direction of the current through the coil. Conducting contacts called **brushes** connect the commutator to the DC source of emf. Graphite, which is used in the brushes, is a form of carbon that conducts electricity and is also used as a lubricant. They are called brushes because they brush against the commutator as it turns. The brushes are necessary to stop the connecting wires from becoming tangled.

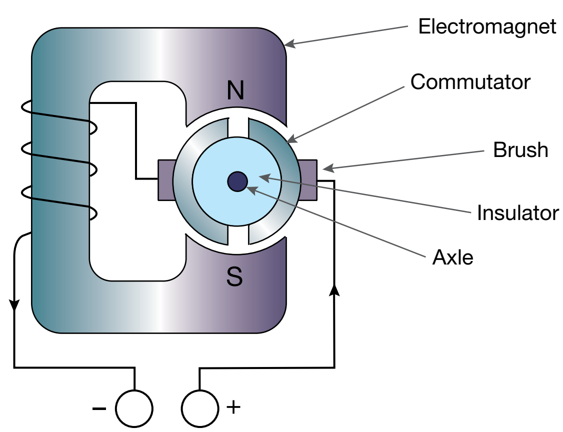
FIGURE 7.4 A close-up look at a split-ring commutator.



The magnetic field in a DC motor

The magnetic field of a DC motor can be provided either by permanent magnets (see Figure 7.6) or by electromagnets. The permanent magnets are fixed to the body of the motor. Electromagnets can be created using a soft iron shape that has coils of wire around it. The current that flows through the armature coil can be used in the electromagnet coils. One arrangement for achieving this effect is shown in Figure 7.5.

FIGURE 7.5 Using an electromagnet to provide the magnetic field. Note that the coil is not shown in this diagram!



7.2.2 Changing the speed of a DC motor

Increasing the maximum torque acting on the sides can increase the speed of a DC motor. This can be achieved by:

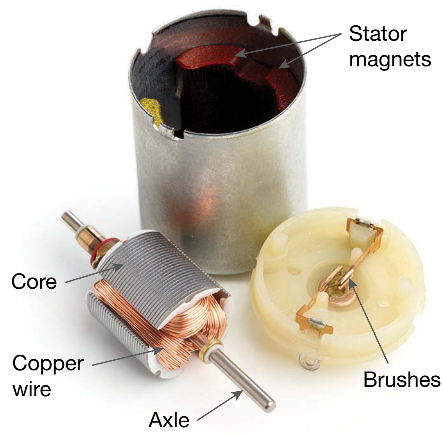
- increasing the force acting on the sides
- increasing the width of the coil
- using more than one coil mounted on the armature.

The force can be increased by

- increasing the current in the coil (this is achieved by increasing the emf across the ends of the coil)
- increasing the number of loops of wire in the coil
- producing a stronger magnetic field with the stator
- using a soft iron core in the centre of the loop. (The core then acts like an electromagnet that changes the direction of its poles when the current changes direction through the coil.) The soft iron core is a part of the armature.

Another method used to increase the average torque acting on the coil and armature is to have two or more coils that are wound onto the armature. This arrangement also means that the motor runs more smoothly than a single-coil motor. Having more than one coil requires a commutator that has two opposite segments for each coil. A stator with curved magnetic poles keeps the force at right angles to the line joining the position of application and the axle for longer. This keeps the torque at its maximum value for a longer period of time. Figure 7.6 shows many of these features in a small battery-operated DC motor, which is shown disassembled. Note that the stator magnets are visible in the casing. The poles of this magnet are on the inside and outside surfaces. The armature has three iron lobes that form the cores of the coils. The coils are made from enamelled copper wire wound in series on the lobes of the armature. The enamel insulates the wire and prevents short circuits.

FIGURE 7.6 A disassembled battery-operated DC motor.



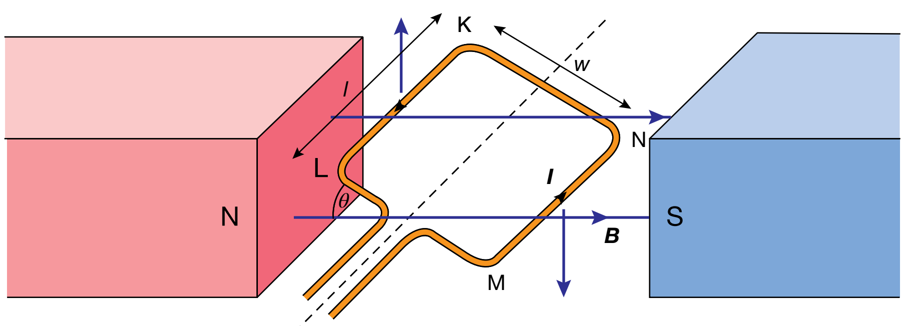
PHYSICS FACT

- Michael Faraday came up with the idea of an electric motor in 1821.
- The first electric motor was created by accident when two generators were connected together by a worker at the Vienna Exhibition in 1873.
- The French engineer and inventor Zénobe–Théophile Gramme produced the first commercial motors in 1873.
- Direct current (DC) motors were installed in trains in Germany and Ireland in the 1880s.
- Nikola Tesla patented the first significant alternating current (AC) motor in 1888.

7.2.3 Calculating the torque of a coil in a DC motor

Consider a single coil of length, l , and width, w, lying in a magnetic field, $\boldsymbol{B}$. The plane of the coil makes an angle, θ, with the magnetic field. The coil carries a current, $\boldsymbol{I}$, and is free to rotate about a central axis. This situation is shown in Figure 7.7.

FIGURE 7.7 The plane of the coil at an angle, θ, to the magnetic field.



Side KL experiences a vertically upward force of $\boldsymbol{I}l\boldsymbol{B}$. Side MN experiences a vertically downward force of $\boldsymbol{I}l\boldsymbol{B}$.

Both forces exert a clockwise torque on the coil. The magnitude of the torque on each side of the coil is given by:

$$\tau = Fd \sin \phi$$

where ϕ is the angle between side KL or NM and the magnetic field. Note that as ϕ decreases from 90° to 0°, θ, which is now the angle between the plane of the coil and the magnetic field, increases from 0° to 90°.

Also,

$F = IlB$, $d = \frac{w}{2}$ and $\phi = (90 - \theta)°$.

Therefore the total torque acting on the coil is given by:

$$\tau = 2 \times IlB \times \frac{w}{2} \times \sin(90 - \theta)°.$$

Since $l \times w = A$, the area of the coil and $\sin(90 - \theta)° = \cos\theta$, the total torque acting on a coil can be expressed as:

$$\tau = BIA \cos\theta.$$

If the coil has n loops of wire on it, the above formula becomes:

$$\tau = nBIA \cos\theta.$$

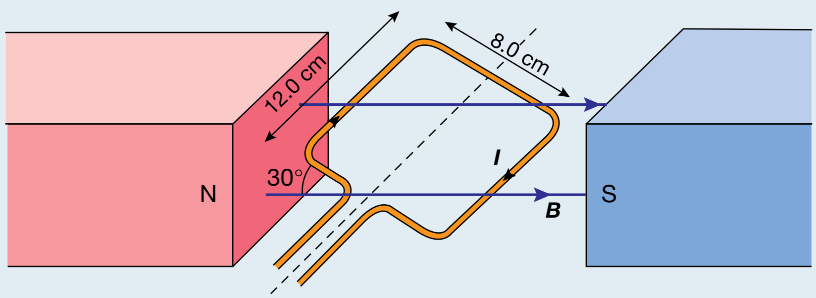
(Remember that θ is the angle between the *plane of the coil* and the magnetic field.)

7.2 SAMPLE PROBLEM 1

CALCULATING TORQUE ON A COIL

A coil contains 15 loops and its plane is sitting at an angle of 30° to the direction of a magnetic field of 7.6 mT. The coil has dimensions as shown in Figure 7.8 and a 15 mA current passes through the coil. Determine the magnitude of the torque acting on the coil and the direction (clockwise or anticlockwise) of the coil's rotation.

FIGURE 7.8



Use the relationship $\tau = nBIA \cos\theta$.

SOLUTION:

Quantity	Value
n	15 loops
A	9.6×10^{-3} m^2
I	1.5×10^{-2} A
θ	30°
τ	?

$$\tau = 15 \times 7.6 \times 10^{-3} \times 1.5 \times 10^{-2} \times 9.6 \times 10^{-3} \times \cos 30°$$
$$= 1.4 \times 10^{-5}\,\text{N m}$$

To determine the direction of rotation of the coil, apply the right-hand push rule to the left-hand side of the coil. This shows that the direction in this case is anticlockwise.

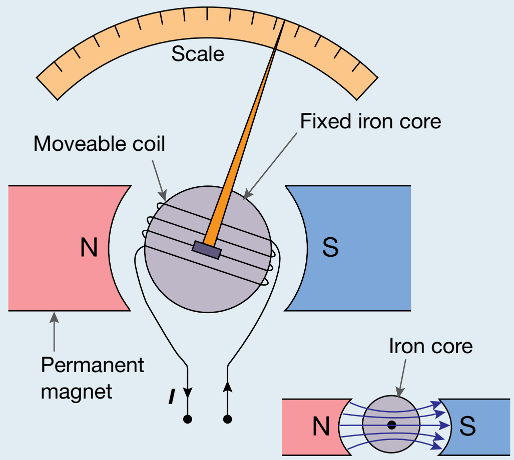
PHYSICS IN FOCUS

The galvanometer

A galvanometer is a device used to measure the magnitude and direction of small direct current (DC) currents. A schematic diagram of a galvanometer is shown in Figure 7.9.

The coil consists of many loops of wire and it is connected in series with the rest of the circuit so that the current in the circuit flows through the coil. When the current flows, the coil experiences a force due to the presence of the external magnetic field (the motor effect). The iron core of the coil increases the magnitude of this force. The needle is rotated until the magnetic force acting on the coil is equalled by a counter-balancing spring. Note that the magnets around the core are curved. This results in a radial magnetic field; the plane of the coil will always be parallel to the magnetic field and the torque will be constant no matter how far the coil is deflected. This also means that the scale of the galvanometer is linear, with the amount of deflection being proportional to the current flowing through the coil.

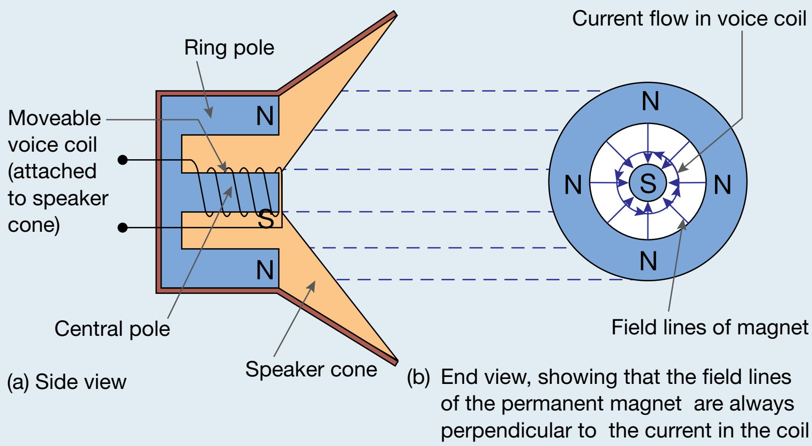
FIGURE 7.9 The galvanometer.



PHYSICS IN FOCUS

Loudspeakers

FIGURE 7.10 A schematic diagram of a loudspeaker.



Loudspeakers are devices that transform electrical energy into sound energy. A loudspeaker consists of a circular magnet that has one pole on the outside and the other on the inside. This is shown in Figure 7.10.

A coil of wire (known as the voice coil) sits in the space between the poles. The voice coil is connected to the output of an amplifier. The amplifier provides a current that changes direction at the same frequency as the sound that is to be produced. The current also changes magnitude in proportion to the amplitude of the sound. The voice coil is caused to vibrate or move in and out of the magnet by the motor effect. The direction of movement of the voice coil can be determined using the right-hand push rule. This can be shown by examining Figure 7.10b. When the current in the coil is anticlockwise, the force on the coil is out of the page. When the current is clockwise, the force on the coil is into the page. The voice coil is connected to a paper speaker cone that creates sound waves in the air as it vibrates. When the magnitude of the current increases, so too does the force on the coil. When the force on the coil increases, it moves more and the produced sound is louder.

7.3 Torque

7.3.1 Production of torque

A **torque** can be thought of as the turning effect of a force acting on an object. Examples of this turning effect occur when you turn on a tap, turn the steering wheel of a car, turn the handlebars of a bicycle or loosen a nut using a spanner, as shown in Figure 7.11. It is easier to rotate an object if the force, $\boldsymbol{F}$, is applied at a greater distance, d, from the pivot axis. It is also easier to rotate the object if the force is at right angles to a line joining the pivot axis to its point of application.

The torque, $\boldsymbol{\tau}$, increases when the force, $\boldsymbol{F}$, is applied at a greater distance, d, from the pivot axis. It is greatest when the force is applied at right angles to a line joining the point of application of the force and the pivot axis.

If the force is perpendicular to the line joining the point of application of the force and the pivot point, the following formula can be used:

$$\tau = Fd.$$

The SI unit for torque is the newton metre (N m).

If the force is not perpendicular to the line joining the point of application of the force and the pivot point, the component of the force that is perpendicular to the line (see Figure 7.12) can be used. The magnitude of the torque can then be calculated using the following formula:

$$\tau = Fd \sin\theta$$

where θ is the angle between the force and the line joining the point of application of the force and the pivot axis.

FIGURE 7.11 A force is applied to a spanner to produce torque on a nut and the spanner.

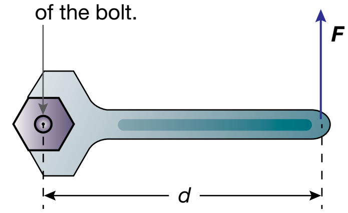
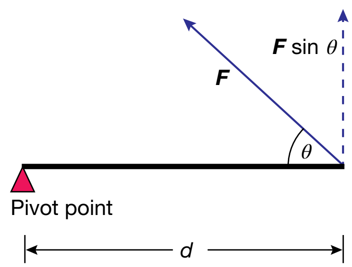


FIGURE 7.12 Calculating torque when $\boldsymbol{F}$ and d are not perpendicular.

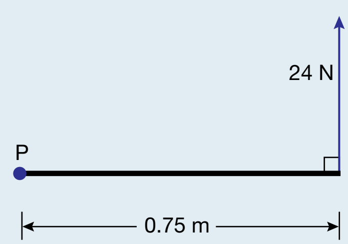


7.3 SAMPLE PROBLEM 1

CALCULATING TORQUE

A lever is free to rotate about a point, P. Calculate the magnitude of the torque acting on the lever if a force of 24 N acts at right angles to the lever at a distance of 0.75 m from P. The situation is shown in Figure 7.13.

FIGURE 7.13



SOLUTION:

Quantity	Value
F	24 N
d	0.75 m
τ	?

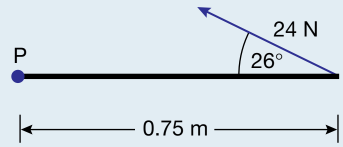
$$\begin{aligned}\tau &= Fd \\ &= 24 \times 0.75 \\ &= 18\ \text{N m}\end{aligned}$$

7.3 SAMPLE PROBLEM 2

CALCULATING TORQUE

What would be the magnitude of the torque in sample problem 1 if the force was applied at an angle of 26° to the lever, as shown in Figure 7.14?

FIGURE 7.14



SOLUTION:

Quantity	Value
F	24 N
d	0.75 m
θ	26°
τ	?

$$\begin{aligned}\tau &= Fd\sin\theta \\ &= 24 \times 0.75 \times \sin 26^{\circ} \\ &= 7.9\,\text{Nm}\end{aligned}$$

7.3 Exercise 1

1 A torque wrench is used to tighten nuts onto their bolts to a specific tightness or force. A torque wrench has a handle (black in the photo below) on one end and a socket that fits over a nut on the other end. In between is a scale that gives a reading in Newton metres.

FIGURE 7.15

The scale on a torque wrench has a reading of 30 Newton metres. If the hand applying the force is 30 cm from the end, what is the size of the force by the hand on the wrench?

2 The handle of a torque wrench is hollow so an extension rod can be inserted. If you can exert only 30 N of force, how far along the extension rod from the handle should you place your hand to achieve a torque of 30 Nm?

eBook*plus* RESOURCES

Try this interactivity: Torque
Searchlight ID: int-0049

Watch this eLesson: Torque
Searchlight ID: eles-0025

7.4 Lenz's Law and the production of back emf in motors

7.4.1 Lenz's Law

Electric motors use an input voltage to produce a current in a coil to make the coil rotate in an external magnetic field. It has been shown that an emf is induced in a coil that is rotating in an external magnetic field. The emf is produced because the amount of the magnetic flux that is threading the coil is constantly changing as the coil rotates. The emf induced in the motor's coil, as it rotates in the external magnetic field, is in the opposite direction to the input voltage or supply emf. If this was not the case, the current would increase and the motor coil would go faster and faster forever. The induced emf produced by the rotation of a motor coil is known as the **back emf** because it is in the opposite direction to the supply emf.

The net voltage across the coil equals the input voltage (or supply emf) minus the back emf. If there is nothing attached to an electric motor to slow it down (and if we ignore the minimal friction effects of an electric motor), the speed of the armature coil increases until the back emf is equal to the external emf. When this occurs, there is no voltage across the coil and therefore no current flowing through the coil. With no current through the coil there is no net force acting on it and the armature rotates at a constant rate.

When there is a load on the motor, the coil rotates at a slower rate and the back emf is reduced. There will be a voltage across the armature coil and a current flows through it, resulting in a force that is used to do the work. Since the armature coil of a motor has a fixed resistance, the net voltage across it determines the magnitude of the current that flows.

The smaller the back emf is, the greater the current flowing through the coil. If a motor is overloaded, it rotates too slowly. The back emf is reduced and the voltage across the coil remains high, resulting in a high current through the coil that could burn out the motor. Motors are usually protected from the initially high currents produced when they are switched on by a series resistor. This resistor is switched out of the circuit at higher speeds because the back emf results in a lower current in the coil.

7.4 SAMPLE PROBLEM 1

CURRENTS IN ELECTRIC MOTORS

The armature winding of an electric motor has a resistance of 10 Ω. The motor is connected to a 240 V supply. When the motor is operating with a normal load, the back emf is equal to 232 V.

(a) What is the current that passes through the motor when it is first started?

(b) What is the current that passes through the motor when it is operating normally?

SOLUTION:

(a) When the motor is first started, there is no back emf. The voltage drop across the motor is 240 V.

Quantity	Value
V	240 V
R	10 Ω
I	?

$$
\begin{aligned}
V &= IR \\
I &= \frac{V}{R} \\
&= \frac{240}{10} \\
&= 24\ \text{A}
\end{aligned}
$$

(b) When the motor is operating normally, the voltage drop across the motor equals the input voltage minus the back emf.

So $V = 240\text{ V} - 232\text{ V} = 8\text{ V}$.

Quantity	Value
V	8 V
R	10 Ω
I	?

$V = IR$

$$I = \frac{V}{R}$$
$$= \frac{8}{10}$$
$$= 0.8\text{ A}$$

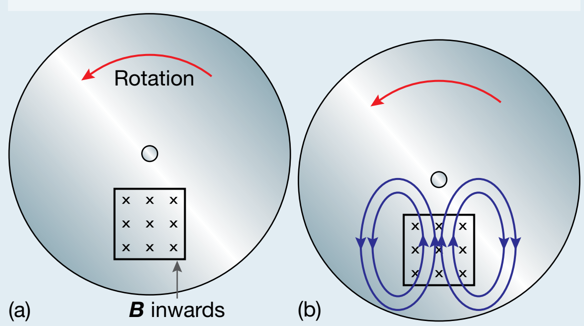
This example shows that electric motors require large currents when starting compared with when they are operating normally.

PHYSICS IN FOCUS

Electromagnetic braking

Consider a metal disc that has a part of it influenced by an external magnetic field, as illustrated in Figure 7.16a. As the disc is made of metal, the movement of the metal through the region of magnetic field causes eddy currents to flow. Using the right-hand push rule, it can be shown that the eddy current within the magnetic field in Figure 7.16 will be upwards. The current follows a downward return path through the metal outside the region of magnetic influence. This is shown in Figure 7.16b.

FIGURE 7.16 (a) A rotating metal disc acted upon by a magnetic field (b) The current that flows in the disc.



The magnetic field exerts a force on the induced eddy current. This can be shown to oppose the motion of the disc by applying the right-hand push rule. In this way eddy currents can be utilised in smooth braking devices in trams and trains. An electromagnet is switched on so that an external magnetic field affects part of a metal wheel or the steel rail below the vehicle. Eddy currents are established in the part of the metal that is influenced by the magnetic field. These currents inside the magnetic field experience a force that acts in the opposite direction to the relative motion of the train or tram, as explained below. In the case of the wheel, the wheel is slowed down. In the case of the rail, the force acts in a forward direction on the rail and there is an equal and opposite force that acts on the train or tram. *Note:* The right-hand push rule is used twice. The first time we use it, we show that an eddy current is produced. The thumb points in the direction of movement of the metal disc through the field because we imagine that the metal contains many positive charges moving through the field. We push in the direction of the force on these charges. This push gives us the direction of the eddy current.

The second time we use the right-hand push rule, we show that there is a force opposing the motion of the metal. Our thumb is put in the direction of the current in the field (the eddy current), then we push in the direction of the force on the moving charges (which are part of the metal disc). We then see that the force is always in the opposite direction to the movement of the metal.

7.5 Generators

7.5.1 Operation of a generator

A generator is a device that transforms mechanical kinetic energy into electrical energy. In its simplest form, a generator consists of a coil of wire that is forced to rotate about an axis in a magnetic field.

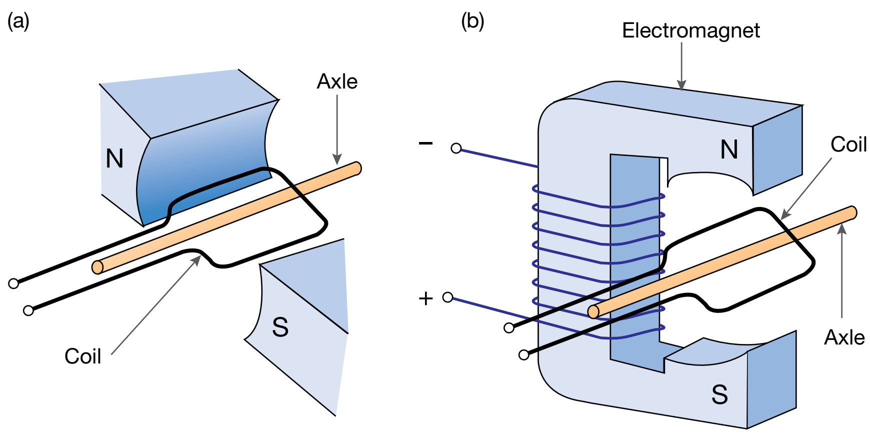
As the coil rotates, the magnitude of the magnetic flux threading (or passing through) the area of the coil changes. This changing magnetic flux produces a changing emf across the ends of the wire that makes up the coil. This is in accordance with Faraday's Law of Induction (see Topic 6), which can be stated as:

The induced emf in a coil is equal in magnitude to the rate at which the magnetic flux through the coil is changing with time.

The magnetic field of a generator can be provided either by using permanent magnets, as shown in Figure 7.17a or by using an electromagnet, as shown in Figure 7.17b.

The stationary functioning parts of a generator are called the **stator**, and the rotating parts are called the **rotor**. In figures 7.17a and 7.17b, the stators consist of the sections that produce the magnetic fields (permanent magnets or electromagnets). The rotors are the coils.

FIGURE 7.17 (a) Permanent magnets provide the magnetic field. (b) An electromagnet provides the magnetic field.



If the coil of a generator is forced to rotate at a constant rate, the flux threading the coil and the emf produced across the ends of the wire of the coil vary with time as shown in Figure 7.18.

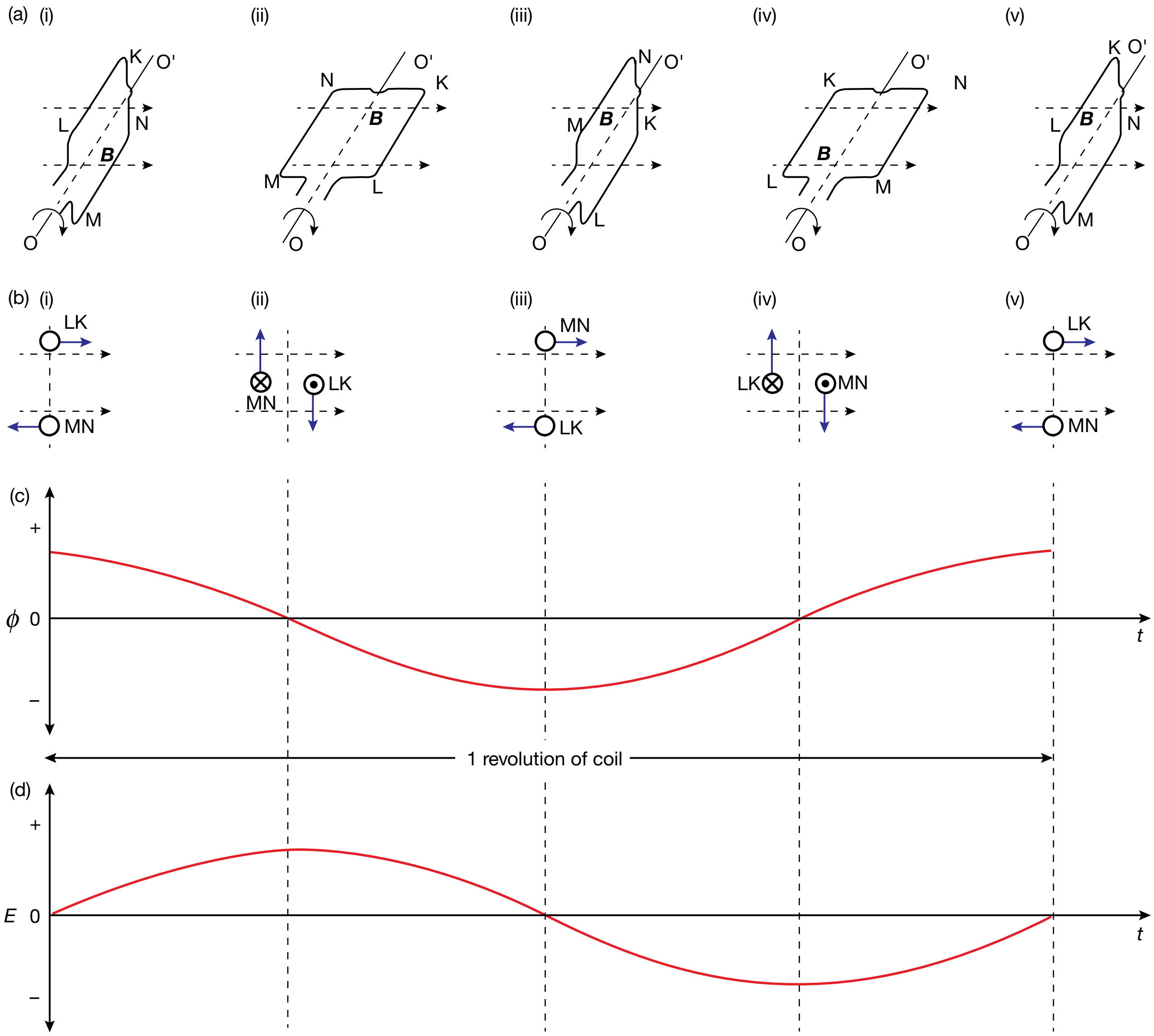
In Figure 7.18 the magnetic field is directed to the right. The corners of the coil have been labeled L, K, M and N so that you can see how the coil is rotating.

Beneath the diagrams of the coil is an end view of the sides LK and MN showing the direction of the induced current that would flow through the sides at that instant if the generator coil was connected to a load. The arrows on this part of the diagram show the direction of movement of the sides of the coil.

The next section down in the diagram is a graph showing the variation of magnetic flux through the coil as a function of time.

The last section of the diagram is a graph showing the variation of emf that would be induced in the coil (if there was a gap in the coil between the points L and M) as a function of time. The emf is given by the negative of the gradient of a graph of magnetic flux threading the coil versus time.

FIGURE 7.18 The variation of flux and emf of a generator coil as it completes a single revolution.



In Figure 7.18a (i) the flux threading the coil is at a maximum value. The emf is zero, as the gradient of the flux versus time graph is zero, which means that there is no change in flux through the coil at this instant.

In Figure 7.18a (ii) the flux threading the coil is zero. The emf is at a maximum positive value, as the flux versus time graph has a maximum negative gradient. At this instant the change in flux is happening at a maximum rate.

In Figure 7.18a (iii) the coil is again perpendicular to the magnetic field, but now the coil is reversed to its original orientation. The flux threading the coil is at a maximum negative value. The emf is zero, as the gradient of the flux versus time graph is again zero, meaning that at this instant there is again no change in flux.

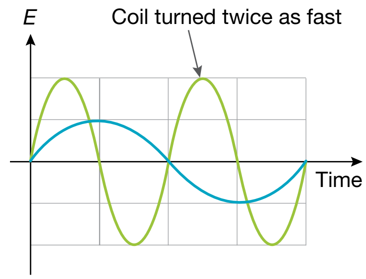
In Figure 7.18a (iv) the flux threading the coil is again zero. The emf now has its maximum positive value, as the gradient of the flux versus time graph has its maximum negative value. At this instant the change in flux is again happening at a maximum rate.

In Figure 7.18a (v) the flux threading the coil is again at a maximum value. The emf is zero, as the gradient of the flux versus time graph is zero and there is no change in flux at this instant. And so the cycle continues.

The frequency and amplitude of the voltage produced by a generator depend on the rate at which the rotor turns. If the rotor is turning at twice the original rate, then the period of the voltage signal halves, the frequency doubles and the amplitude doubles. This is shown in Figure 7.19.

The effectiveness of generators is increased by winding the coil onto an iron core armature. The iron core makes the coil behave like an electromagnet. This intensifies the changes in flux threading the coil as it is forced to rotate and increases the magnitude of the emf that is induced. This effect also occurs when the number of turns of wire on the armature is increased. The coil then behaves like a number of individual coils connected in series. If there are n turns of wire on the armature, the maximum emf will be n times that of a single coil rotating at the same rate.

FIGURE 7.19 Doubling the frequency of rotation doubles the maximum induced emf.



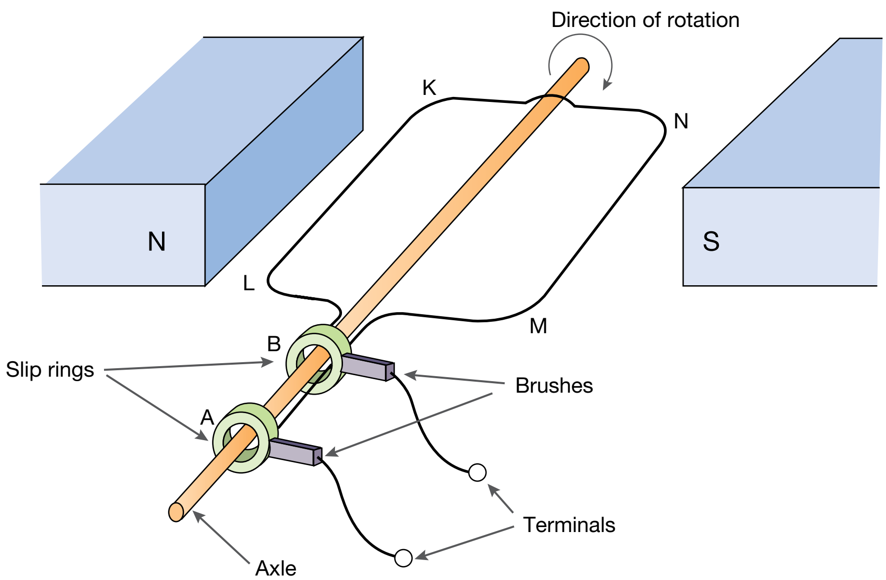
7.5.2 AC generators

Figure 7.18 shows how a coil forced to rotate smoothly in a magnetic field has a varying emf induced across the ends of the coil. The value of the emf varies sinusoidally with time. (This means that the graph of emf versus time has the same shape as a graph of $\sin x$ versus x.) If such an emf signal were placed across a resistor, the current flowing through the resistor would periodically alternate its direction. In other words, the emf across the ends of a coil rotating at a constant rate in a magnetic field produces an alternating current (AC). Alternating current electrical systems are used across the world for electrical power distribution.

This type of AC generator connects the coil to the external circuit or distribution system by the use of slip rings. Slip rings rotate with the coil. A slip ring system is shown in Figure 7.20.

In Figure 7.20, side LK of the coil is connected to slip ring B while side MN is connected to slip ring A. Brushes make contact with the slip rings and transfer the emf (or current) to the **terminals** of the generator. In this case, the terminals are the external points of the generator where it connects to the load.

FIGURE 7.20 The functional parts of an AC generator.

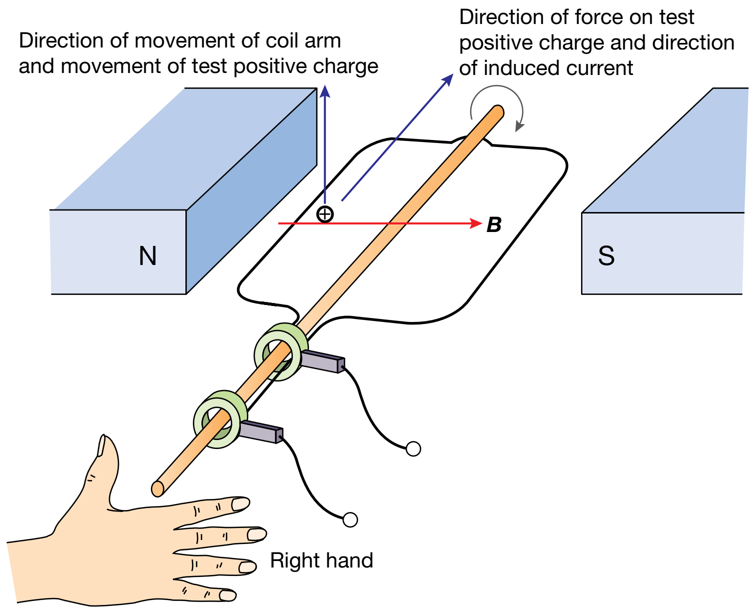


7.5.3 Which way will the current flow?

When asked to determine the direction of the current in the generator or some other part of a circuit connected to the generator, there are two methods that can be used.

The first method is to consider the magnetic force on a positive test charge in one side of the coil. The direction of the velocity of the charge in the magnetic field depends on the direction of the rotation of the coil. The direction of the magnetic force is determined using the right-hand push rule (see Figure 7.21). The direction of the force acting on the test charge is also the direction of the current on that side of the generator. It is then a matter of following that direction around the coil to the terminals. Note that the terminal from which the current emerges at a particular instant is acting as the positive terminal.

FIGURE 7.21 Using the right-hand push rule to determine the direction of current flow in a generator coil.



The other method is to apply Lenz's Law to the coil. First determine the way in which the flux threading the coil is changing at the instant in question. The current induced in the coil will produce a magnetic field that opposes the change in flux through the coil. Once you have established the direction of the flux produced by the induced current, apply the right-hand grip rule for coils to determine the direction of the current around the coil.

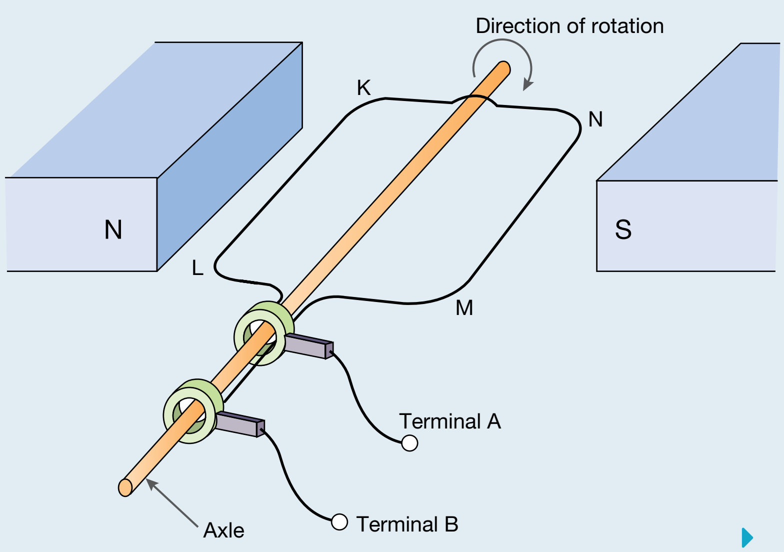
Both methods are illustrated in sample problem 1.

7.5 SAMPLE PROBLEM 1

DETERMINING THE POLARITY OF A GENERATOR'S TERMINALS

Figure 7.22 shows an AC generator at a particular instant. At this instant, which of the terminals, A or B, is positive?

FIGURE 7.22



SOLUTION:

Test charge method

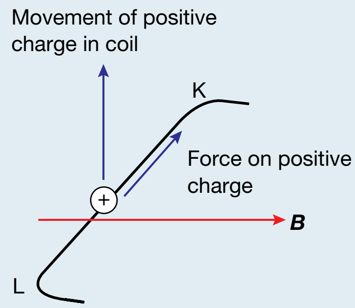
Consider a positive test charge in the side labelled LK. At the instant shown, this positive charge is moving upwards in a magnetic field directed to the right. Applying the right-hand rule, the positive charge is forced in the direction from L towards K. This situation is shown in Figure 7.23.

Side LK is connected to the slip ring leading to terminal A. Side MN is connected to the slip ring leading to terminal B. If a current were to flow, it would emerge from terminal B. Therefore, terminal B is positive at the instant shown.

Using Lenz's Law

At the instant shown in Figure 7.22, the flux is increasing to the right through the coil as it is forced to rotate in the indicated direction. The induced current in the coil will therefore produce a magnetic field that passes through the coil to the left to oppose the external change in magnetic flux through the coil. The right-hand grip rule for coils (thumb in the direction of the induced magnetic field through the coil, fingers grip the coil pointing in the direction of the current in the coil) shows that the induced current is clockwise around the coil as we view it. The current then emerges from the generator through terminal B. Therefore, terminal B is positive at the instant shown.

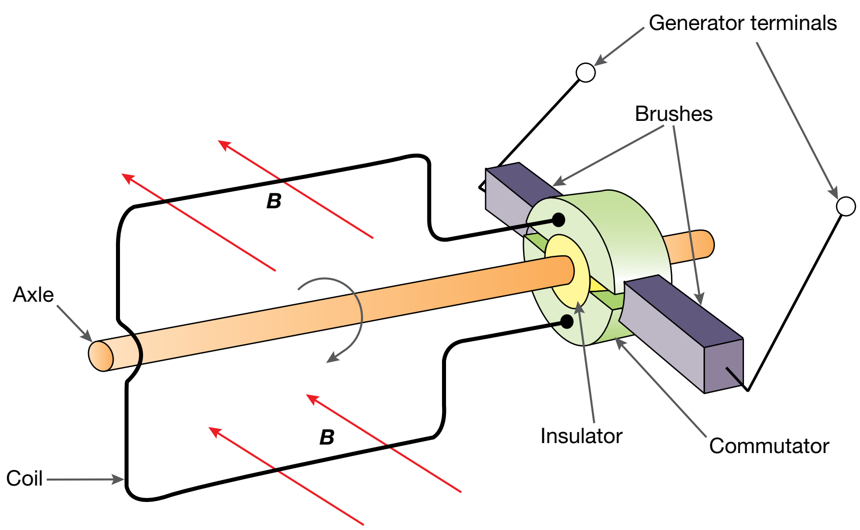
FIGURE 7.23 Positive charges are pushed in the direction from L towards K.



7.5.4 DC generators

A direct current (DC) is a current where the flow of charge is in one direction only. Direct currents provided by a battery or dry cell usually have a steady value. Direct currents may also vary with time, but keep flowing in the same direction. DC generators provide such currents.

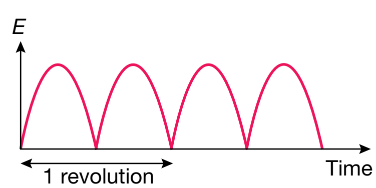
FIGURE 7.24 A simple DC generator.



A simple DC generator consists of a coil that rotates in a magnetic field. (This also occurs in an AC generator.) The difference between an AC and a DC generator is in the way that the current is provided to the external circuit. An AC generator uses slip rings. A DC generator uses a split-ring commutator to connect the rotating coil to the terminals. (Remember that a commutator is a switching device for reversing the direction of an electric current.) The functional parts of a simple DC generator are shown in Figure 7.24. The magnets have been omitted for clarity.

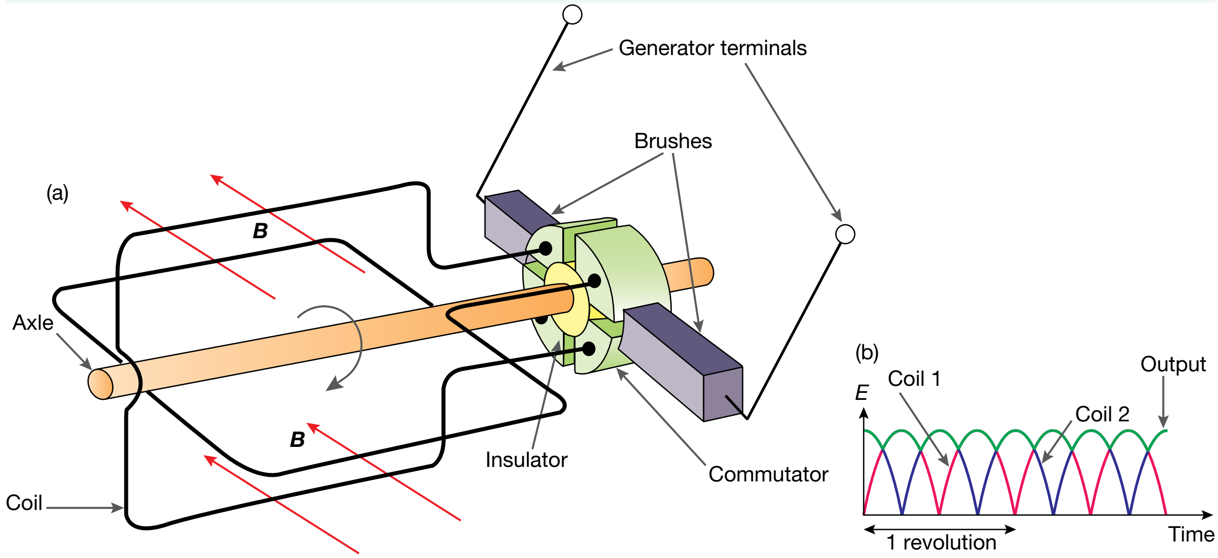
This diagram of a simple DC generator should remind you of a DC motor (see Section 7.2), as it has the same parts. In the generator, the coil is forced to rotate in the magnetic field. This induces an emf in the coil. The emf is transferred to the external circuit via the brushes that make contact with the commutator. When the emf of the coil changes direction, the brushes swap over the side of the coil they are connected to, thus causing the emf supplied to the external circuit to be in one direction only. The result of this process is shown in Figure 7.25.

FIGURE 7.25 The output from a simple DC generator.



The output from a DC generator can be made smoother by including more coils set at regular angles on the armature. Each coil is connected to two segments of a multi-part commutator and the brushes make contact only with the segments connected to the coil producing the greatest emf at a particular time. A two-coil DC generator is shown in Figure 7.26a and its output is shown in Figure 7.26b. Note that in this case the commutator has four segments.

FIGURE 7.26 (a) A two-coil DC generator (b) The output from a two-coil DC generator.

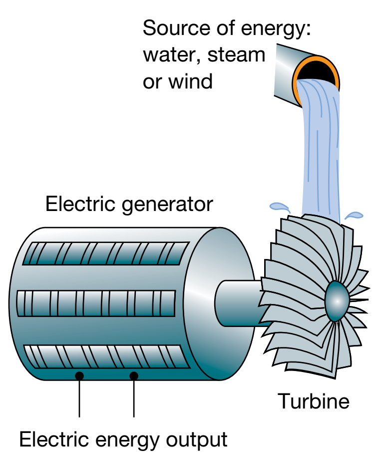


7.6 Electric power generating stations

7.6.1 Domestic and industrial generators

Electric power generating stations provide electrical power to domestic and industrial consumers. In a power station, mechanical or heat energy is transformed into electrical energy by means of a turbine connected to a generator. A turbine is a machine whose shaft is rotated by jets of steam or water directed onto blades attached to a wheel. Figure 7.27 shows a simple turbine and generator combination.

FIGURE 7.27 A turbine drives a generator.



The generators used in power stations have a different structure to those studied so far. A typical generator has an output of 22 kV. This requires the use of massive coils that would place huge forces on bearings if they were required to rotate. To eliminate this problem, a power station generator has stationary coils mounted on an iron core (making up the stator). The coils are linked in pairs on opposite sides of the rotor. The rotor is a DC supplied electromagnet that spins with a frequency of 50 Hz. A simplified diagram of a power station generator is shown in Figure 7.28. In this diagram only one set of linked coils is shown.

Power station generators have three sets of coils mounted at angles of 120° to each other on the stator. This means that each generator produces three sets of voltage signals that are out of phase with each other by 120°. This is known as three-phase power generation. Each generator is connected to four lines, one line for each phase and a return ground line. Figure 7.29a shows the arrangement of the coils on the stator and Figure 7.29b shows the voltage outputs of each set of coils.

There are two main types of power station used in Australia: fossil fuel steam stations and hydroelectric stations.

FIGURE 7.28 A single-coil generator.

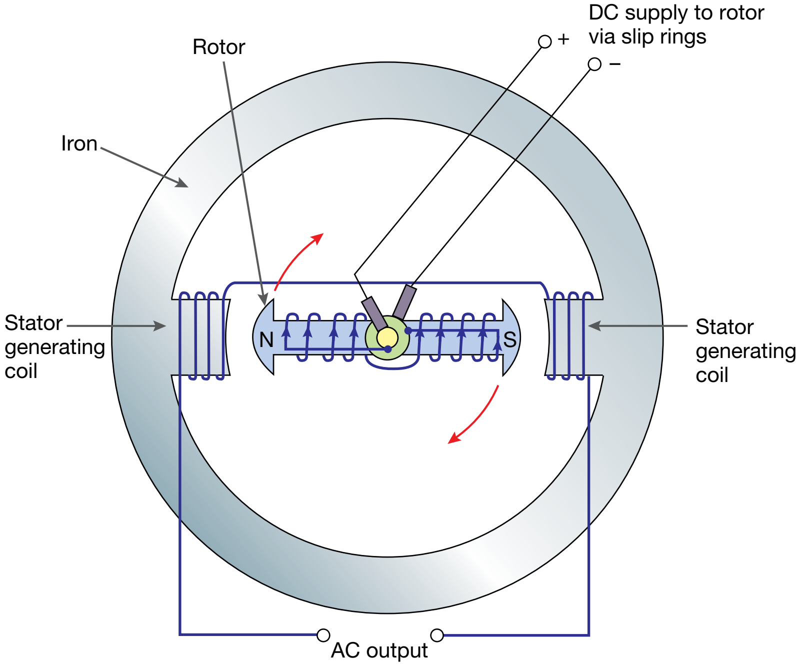
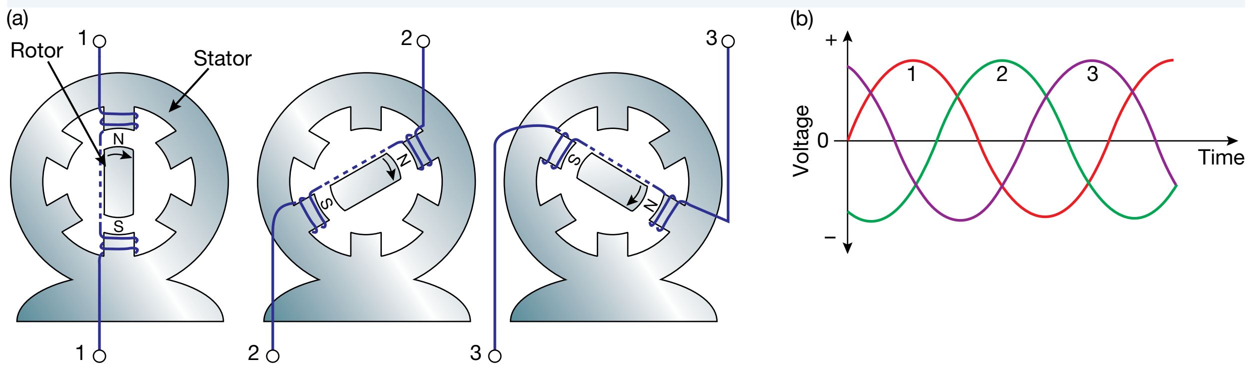


FIGURE 7.29 Three-phase power generation.



7.7 AC induction motors

7.7.1 Main featues of an AC motor

As we have previously seen, alternating current (AC) is widely used in today's world. It is easier to produce in power generating stations and easier to distribute over large distances with small energy losses due to the use of transformers. AC electricity is also produced at a very precise frequency. In Australia this frequency is 50 Hz.

AC motors are used when very precise speeds are required, for example in electric clocks. AC motors operate using an alternating current (AC) electrical supply. Electrical energy is usually transformed into rotational kinetic energy.

As with the DC motors, AC generators and DC generators that have been studied earlier, AC motors have two main parts. These are called the stator and the rotor.

The stator is the stationary part of the motor and it is usually connected to the frame of the machine. The stator of an AC motor provides the external magnetic field in which the rotor rotates. The magnetic field produces a torque on the rotor.

Most AC motors have a cylindrical rotor that rotates about the axis of the motor's shaft. This type of motor usually rotates at high speed, with the rotor completing about one revolution for each cycle of the AC electricity supply. This means that Australian AC motors rotate at about 50 revolutions per second or 3000 revolutions per minute. If slower speeds are required, they are achieved using a speed-reducing gearbox. This type of motor is found in electric clocks, electric drills, fans, pumps, compressors, conveyors, and other machines in factories. The rotor is mounted on bearings that are attached to the frame of the motor. In most AC motors the rotor is mounted horizontally and the axle is connected to a gearbox and fan. The fan cools the motor.

Both the rotor and the stator have a core of ferromagnetic material, usually steel. The core strengthens the magnetic field. The parts of the core that experience alternating magnetic flux are made up of thin steel laminations separated by insulation to reduce the flow of eddy currents that would greatly reduce the efficiency of the motor.

7.7.2 The universal motor

There are two main classifications of AC motors. Single-phase motors operate on one of the three phases produced at power generation plants. Single-phase AC motors can operate on domestic electricity. Polyphase motors operate on two or three of the phases produced at power generation plants. One type of single-phase AC electric motor is the **universal motor**.

Universal motors are designed to operate on DC and AC electricity. They are constructed on similar lines to the DC motor studied in Section 7.2. The rotor has several coils wound onto the rotor armature. The ends of these coils connect to opposite segments of a commutator. The external magnetic field is supplied by the stator electromagnets that are connected in series with the coils of the armature via brushes. The interaction between the current in a coil of the armature and the external magnetic field produces the torque that makes the rotor rotate. Even though the direction of the current is changing 100 times per second when the motor is connected to the mains, the universal motor will continue to rotate in the same direction because the magnetic field flux of the stator is also changing direction 100 times every second.

A variable resistor controls the speed of a universal motor by varying the current through the coils of the armature and the field coils of the stator. The universal motor is commonly used for small machines such as portable drills and food mixers. Figure 7.30 shows a

FIGURE 7.30 A universal motor.

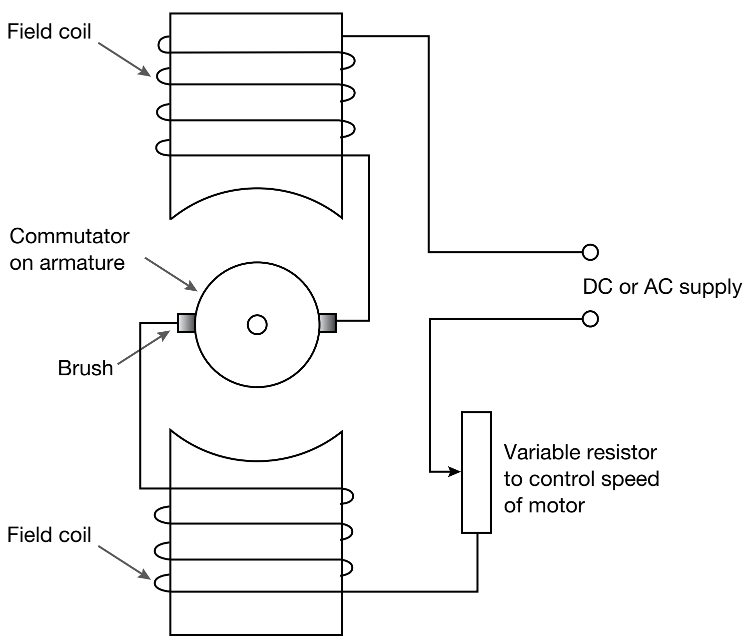
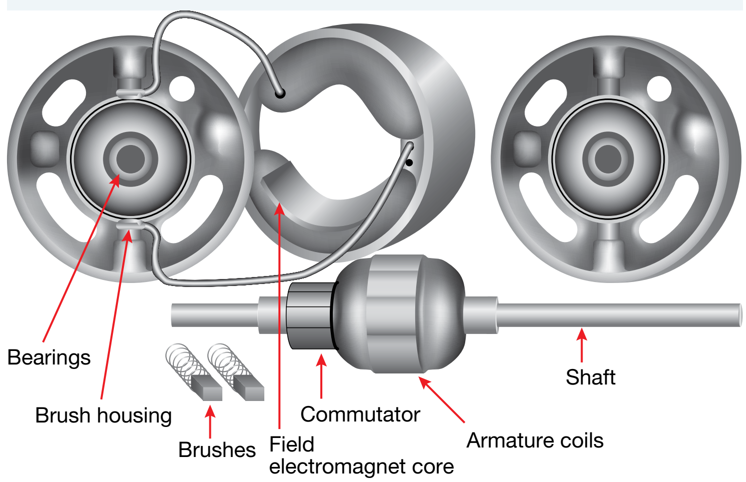


FIGURE 7.31 A dismantled universal motor.



schematic diagram of a universal motor, and Figure 7.31 shows a diagram of a universal motor that has been taken apart.

7.7.3 AC induction motors

Induction motors are so named because a changing magnetic field that is set up in the stator induces a current in the rotor. This is similar to what happens in a transformer, with the stator corresponding to the primary coil of the transformer and the rotor corresponding to the secondary. One difference is that, in an induction motor, the two parts are separated by a thin air gap. Another difference is that, in induction motors, the rotor (secondary coil) is free to move.

The simplest form of AC induction motor is known as the squirrel-cage motor. It is called a squirrel-cage motor because the rotor resembles the cage or wheel that people use to exercise their squirrels or pet mice. It is an induction motor because no current passes through the rotor directly from the mains supply. The current in the rotor is induced in the conductors that make up the cage of the rotor by a changing magnetic field, as explained later in this topic.

Squirrel-cage induction motors are by far the most common types of AC motor used domestically and in industry. Squirrel-cage induction motors are found in some power drills, beater mixes, vacuum cleaners, electric saws, hair dryers, food processors and fan heaters, to name but a few.

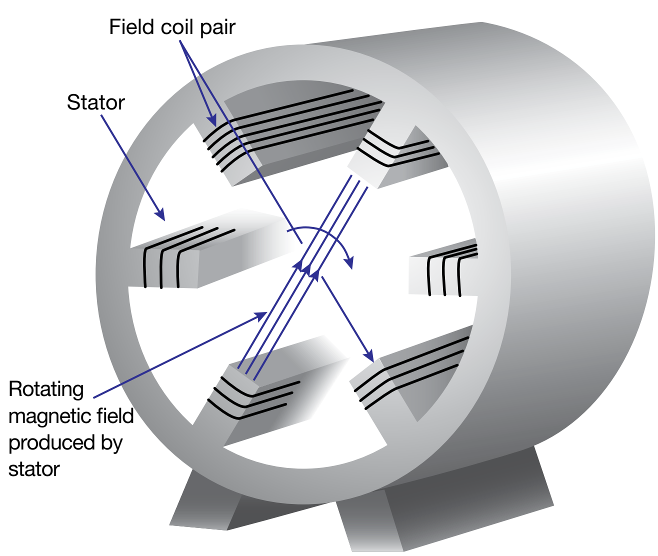
FIGURE 7.32 A mouse exercise wheel is similar to a squirrel cage.

The structure of AC induction motors

The easiest type of induction motor to understand is the three-phase induction motor. This operates by using each phase of AC electricity that is generated in power stations and supplied to factories.

Household electric motors are single-phase motors. This is because houses are usually supplied with only one phase of the three phases that are produced in power stations. It is not important to understand how the rotating magnetic field is achieved in single-phase AC induction motors; therefore, this topic will concentrate on the three-phase motor, as its workings are easier to visualise.

FIGURE 7.33 The rotating magnetic field set up by the stator. Note that in this stator there are three pairs of field coils and that each pair is connected.

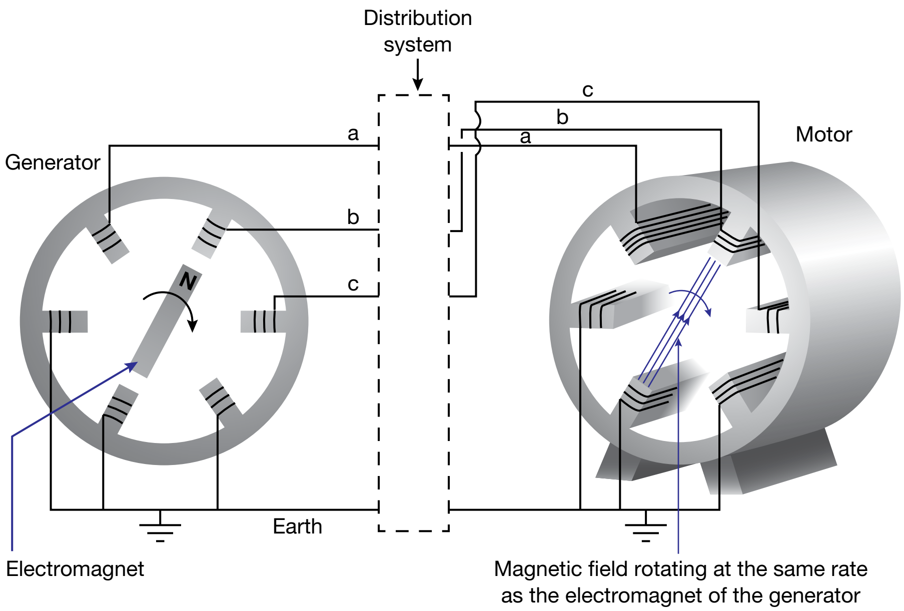


The stator of three-phase induction motors

In both single- and three-phase AC induction motors, the stator sets up a rotating magnetic field that has a constant magnitude. The stator of a three-phase induction motor usually consists of three sets of coils that have iron cores. The stator is connected to the frame of the motor and surrounds a cylindrical space in which it sets up a rotating magnetic field. In three-phase induction motors, this is achieved by connecting each of the three pairs of field coils to a different phase of the mains electrical supply. The coils that make a pair are located on opposite sides of the stator and they are linked electrically. The magnetic field inside the stator rotates at the same frequency as the mains supply; that is, at 50 Hz. A cutaway diagram of a stator is shown in Figure 7.33.

The magnetic field rotates at exactly the same rate as the electromagnet in the power station generator that provides the AC electricity. Each pair of coils in the stator of the generator supplies a corresponding pair of coils in the stator of the motor. Therefore, the magnetic field in the motor rotates at exactly the same rate as the electromagnet in the generator. This is represented in Figure 7.34.

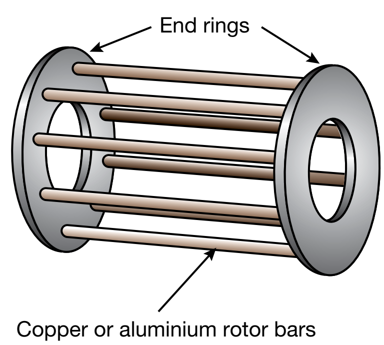
FIGURE 7.34 Supplying three-phase electrical power to the motor.



The squirrel-cage rotor

The rotor of the AC induction motor consists of a number of conducting bars made of either aluminium or copper. These are attached to two rings, known as end rings, at either end of the bars. This forms an object that is sometimes called a **squirrel-cage rotor** (see Figure 7.35). The end rings 'short-circuit' the bars and allow a current to flow from one side to the other of the cage.

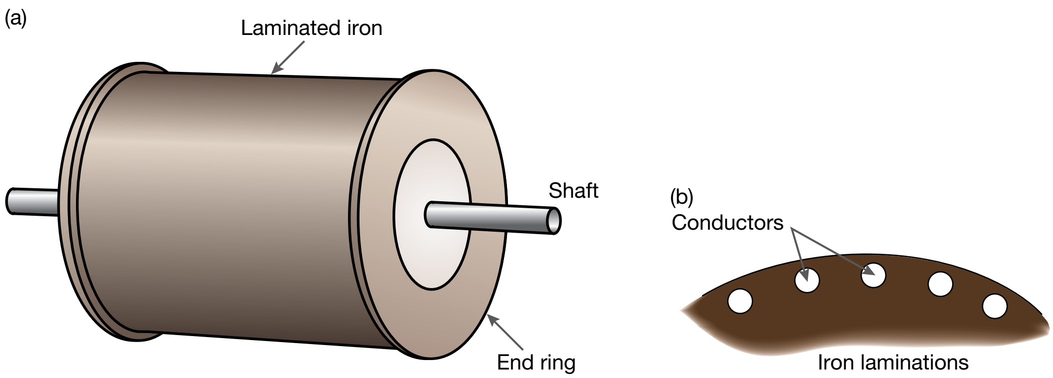
FIGURE 7.35 A squirrel-cage rotor.



The bars and end rings are encased in a laminated iron armature as shown in Figure 7.36. The iron intensifies the magnetic field passing through the conductors of the rotor cage and the laminations decrease the heating losses due to eddy currents. The armature is mounted on a shaft that passes out through the end of the motor. Bearings reduce friction and allow the armature to rotate freely.

Figure 7.37 shows a cutaway model of a fully assembled induction motor. Note the field coils of the stator and the squirrel-cage rotor with a laminated iron core. Also note that the shaft in this case is connected to a gearbox so that a lower speed than 3000 revolutions per minute can be achieved, and that the cooling fan is mounted on the shaft.

FIGURE 7.36 The rotor of an AC motor.



The operation of AC induction motors

As the magnetic field rotates in the cylindrical space within the stator, it passes over the bars of the cage. This has the same effect as the bars moving in the opposite direction through a stationary magnetic field. The relative movement of the bars through the magnetic field creates a current in the bars. Bars carrying a current in a magnetic field experience a force. The force in this case is always in the same direction as the movement of the magnetic field. The cage is then forced to 'chase' the magnetic field around inside the stator.

FIGURE 7.37 A cutaway view of an induction motor.

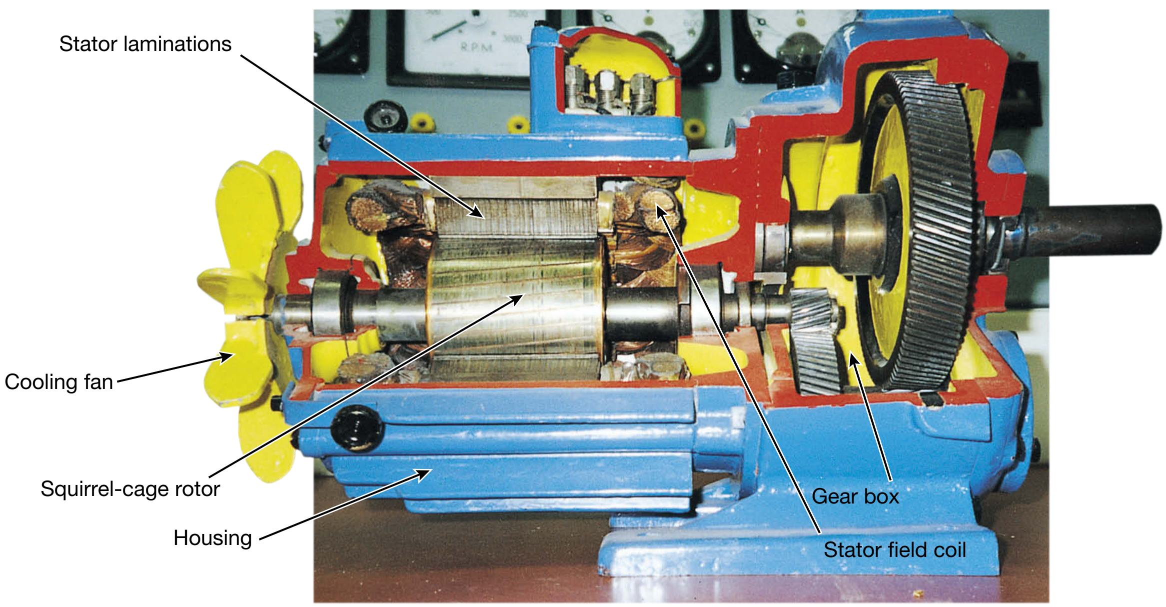
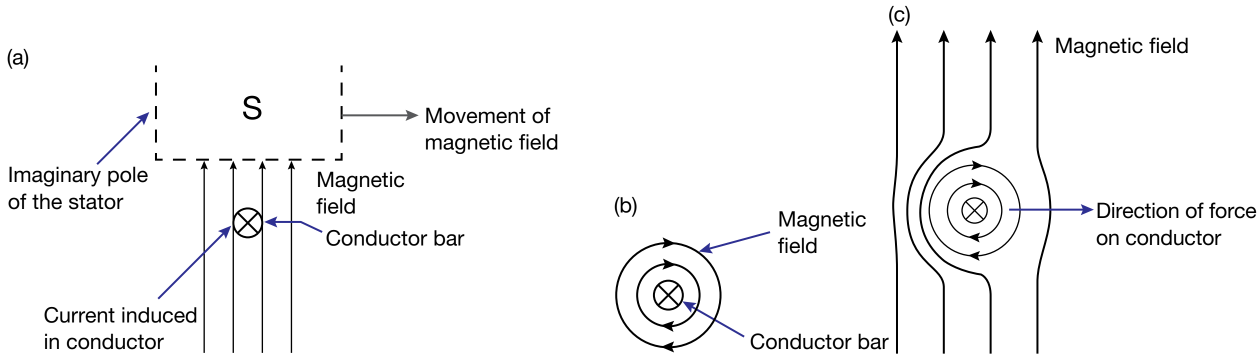


Figure 7.38a shows an end view of the magnetic field as it moves across a conductor bar of the squirrel cage. The magnetic field moving to the right across the conductor bar has the same effect as the conductor bar moving to the left across the magnetic field. You can use the right-hand push rule to determine the direction of the induced current in the conductor. The thumb points to the left (the direction of movement for positive charges relative to the magnetic field), the fingers point up the page (the direction of the magnetic field) and the palm of the hand shows the direction of the force on positive charges and consequently the direction of the induced current. This will show that the current in the bar is flowing into the page.

FIGURE 7.38 (a) The induced current in a conductor bar (b) The direction of the magnetic field of the induced current flowing in the bar (c) The force acting on the bar carrying the induced current.



There is now a current flowing in the conductor bar as shown in Figure 7.38b. The direction of the force acting on the induced current is determined using the right-hand push rule. Therefore, the force on the conductor is to the right, which is in the same direction as the movement of the magnetic field. This is shown in Figure 7.38c.

Slip

If the bars of the squirrel cage were to rotate at exactly the same rate as the magnetic field, there would be no relative movement between the bars and the magnetic field and there would be no induced current and no force. If the cage is to experience a force, there must be relative movement, such as the cage constantly 'slipping' behind the magnetic field. When operating under a load, the retarding force slows the cage down so that it is moving slower than the field. The difference in rotational speed between the cage and the field is known as the **slip speed**. This means that the rotor is always travelling at a slower speed than the magnetic field of the stator when the motor is doing work.

When any induction motor does work, the rotor slows down. You can hear this happen when a beater mix is put into a thick mixture or when a power drill is pushed into a thick piece of wood. When this occurs, the amount of slip is increasing. This means that the relative movement between the magnetic field and the conductor bars is greater and that the induced current and magnetic force due to the current are increased.

Power of AC induction motors

Power is the rate of doing work. Work is done when energy is transformed from one type to another. Induction motors are considered to produce low power because the amount of mechanical work they achieve is low compared with the electrical energy consumed. The electrical power consumed by a motor is calculated using the formula $P = VI$, where V is the voltage at the terminals of the motor, and I is the current flowing through the coils of the stator. The 'lost power' of induction motors is consumed in magnetising the working parts of the motor and in creating induction currents in the rotor.

7.8 Review

7.8.1 Summary

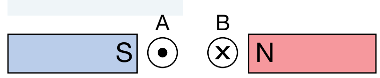
- A DC electric motor is one application of the motor effect.
- A DC electric motor has a current-carrying coil that rotates about an axis in an external magnetic field.
- Galvanometers and loudspeakers are other applications of the motor effect.
- Torque is the turning effect (moment) of a force. The magnitude of the torque is determined using the formula $\tau = Fd \sin \theta$, where θ is the angle between the force and the line joining the point of application of the force and the pivot axis.

- The torque acting on the coil of an electric motor is given by the formula $\tau = nIA_{\perp}B = nIAB\sin\theta$, where θ is the angle between the plane of the coil and the magnetic field.
- A generator is a device that transforms mechanical energy into electrical energy.
- One type of generator has a rotating coil in an externally produced magnetic field.
- An AC generator uses slip rings to connect its terminals to the coil.
- A DC generator uses a split-ring commutator to connect its terminals to the coil.
- AC electric motors are used in many machines in households and in industry.
- The basic operating principles of AC electric motors are the same as for DC motors. A current-carrying conductor in a magnetic field experiences a force, the direction of which is given by the right-hand push rule.
- The stator of an AC induction motor consists of field coils that set up a rotating magnetic field that has a constant magnitude. This field rotates 50 times every second.
- A squirrel-cage rotor has conductor bars that are short-circuited by two end rings. The bars are embedded in a laminated iron core. (The laminations reduce energy losses due to eddy currents.)
- A current is induced in the bars of the rotor as the magnetic field moves across them.
- The bars experience a force because they are carrying a current in a magnetic field. The direction of the force is the same as the direction of movement of the magnetic field. The rotor is therefore forced to chase the magnetic field.
- For the current to be induced in the rotor bars, the rotor must slip behind the magnetic field.

7.8.2 Questions

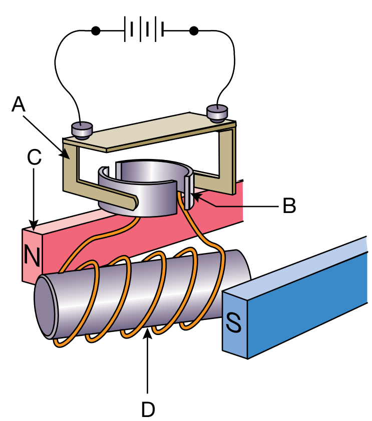
1. The diagram in Figure 7.39 represents a side view of a single loop in a DC electric motor. Identify the direction of the forces acting on sides A and B of the loop.

FIGURE 7.39



2. Figure 7.40 shows the functional parts of a type of DC electric motor.

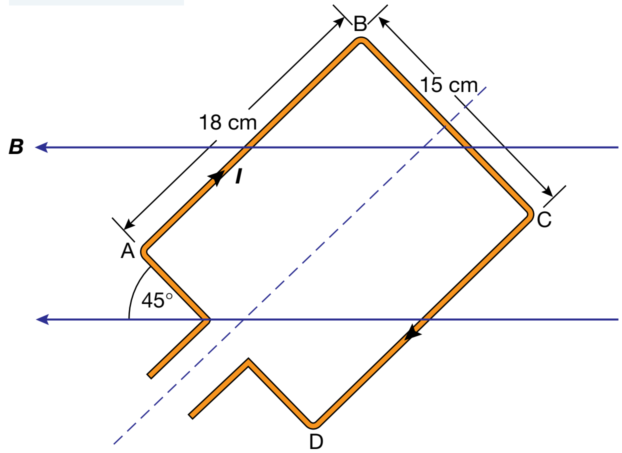
FIGURE 7.40



   (a) Name the parts labelled A to D in the diagram.
   (b) Describe the functions of the parts labelled A to D.
3. A coil is made up of 50 loops of wire and its plane is at an angle of 45° to the direction of a magnetic field of strength 0.025 T. The coil has the dimensions shown in Figure 7.41 and a current of 1.5A flows through it in the direction shown on the diagram.

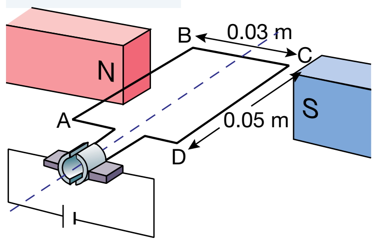
(a) Identify the direction of the force acting on side AB.
(b) Calculate the magnitude of the force acting on side AB.
(c) Calculate the area of the coil.
(d) Calculate the magnitude of the torque acting on the coil when it is in the position shown in the diagram.

FIGURE 7.41



4. A student makes a model motor. She makes a rectangular coil with 25 turns of wire with a length of 0.050 m and width 0.030 m. The coil is free to rotate about an axis that is represented by a dotted line in Figure 7.42.

FIGURE 7.42

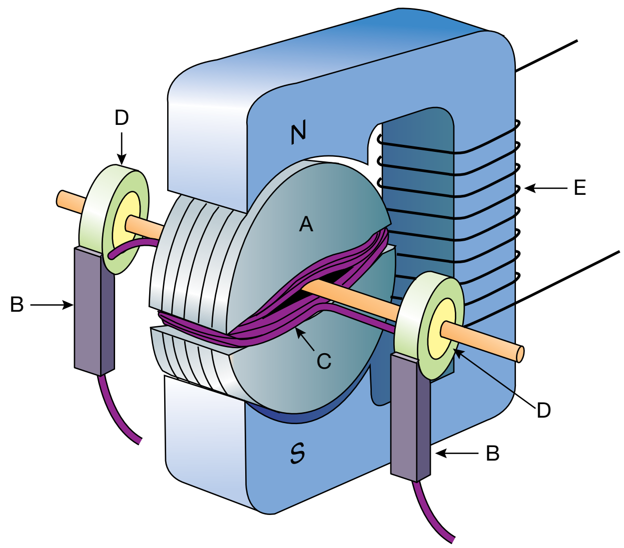


At the instant shown, the plane of the coil is parallel to the direction of the magnetic field. The magnetic field strength is 0.45 T. When the current to the coil is activated, it has a magnitude of 1.75 A in the direction ADCB.

(a) Calculate the magnitude and direction of the force acting on side CD when the current is flowing and the coil is in the position shown in the diagram.
(b) When the current begins to flow, the net force acting on the coil is zero yet the coil begins to rotate. Why does this occur?
(c) Describe what happens to the magnitude and direction of the force acting on side CD as the coil swings through an angle of 60°.
(d) Describe three things the student could do to get the coil to rotate at a faster rate.

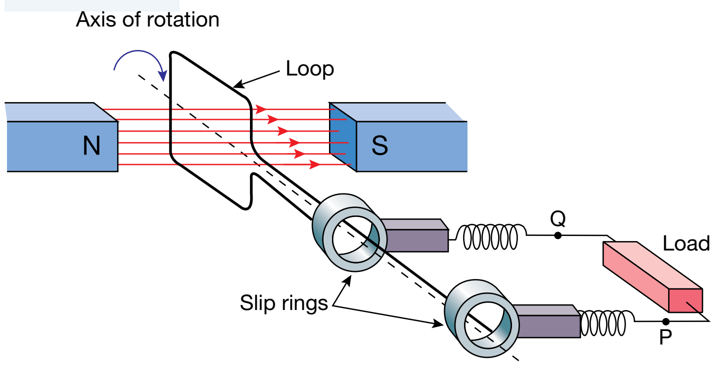
5. Identify the types of energy transformation that occur in electrical generators.
6. Figure 7.43 shows a generator.
   (a) Name the parts of the generator labelled A, B, C, D and E.
   (b) Describe the function of each of these parts.
   (c) What type of generator (AC or DC) is this?

FIGURE 7.43



7. A hand-turned generator is connected to a light globe in series with a switch. Is it easier to turn the coil when the switch is open or when it is closed? Explain your reasoning.
8. Describe the effect of the following changes on the size of the current produced by a generator:
   (a) the number of loops of the coil is increased
   (b) the rate of rotation of the coil is decreased
   (c) the strength of the magnetic field is increased
   (d) an iron core is used in the coil rather than having an air coil.
9. A student builds a model electric generator, similar to that shown in Figure 7.44. The coil consists of 50 turns of wire. The student rotates the coil in the direction indicated on the diagram. The coil is rotated a quarter turn (90°).
   (a) In which direction does the current flow through the external load?
   (b) Sketch a flux versus time graph and an emf versus time graph as the coil completes one rotation at a steady rate starting from the instant shown in the diagram.

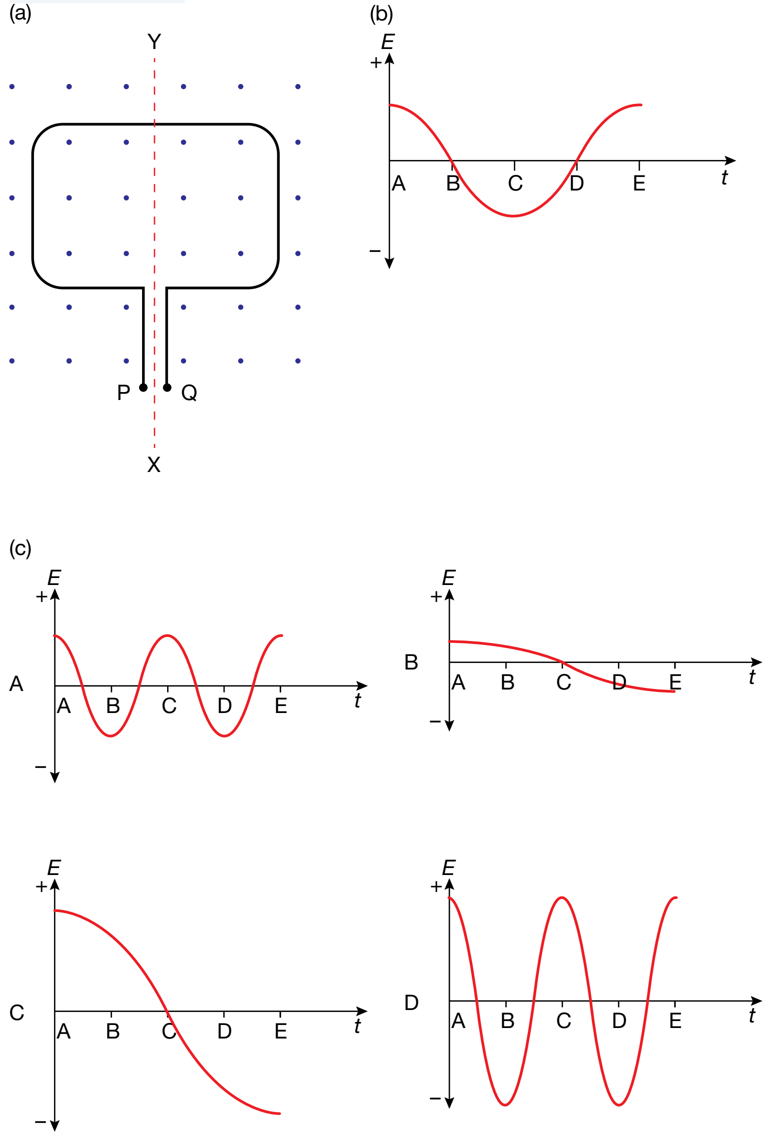
FIGURE 7.44



10. A rectangular coil of wire is placed in a uniform magnetic field, $\boldsymbol{B}$, that is directed out of the page. This is shown in Figure 7.45a. At the instant shown the coil is parallel to the page.
    The coil rotates about the axis XY at a steady rate. The time variation of the voltage drop induced between points P and Q for one complete rotation is shown in Figure 7.45b.

(a) At which time(s) could the coil be in the position shown in Figure 7.45a? Justify your answer.
(b) Which of the graphs in Figure 7.45c shows the variation of voltage versus time if the coil is rotated at twice the original speed?

FIGURE 7.45



11. In a power station, an electromagnet is rotated close to a set of coils. The electromagnet is supplied with a direct current. An idealised diagram of this arrangement is shown in Figure 7.46.
    (a) In this arrangement, which part(s) make up the rotor and which make up the stator?
    (b) Explain why it is necessary to rotate the electromagnet to produce an emf.
    (c) Explain why energy must be provided to the electromagnet to keep it rotating at a constant speed.
12. Describe the main difference between AC and DC generators.
13. Describe the main features of an AC induction motor.
14. Figure 7.47 shows a cutaway diagram of an AC induction motor.
    (a) Name the parts labelled A to E.
    (b) Explain the functions of the parts labelled A to E.
15. Briefly describe how the rotating magnetic field is produced in a three-phase AC electric motor.

FIGURE 7.46

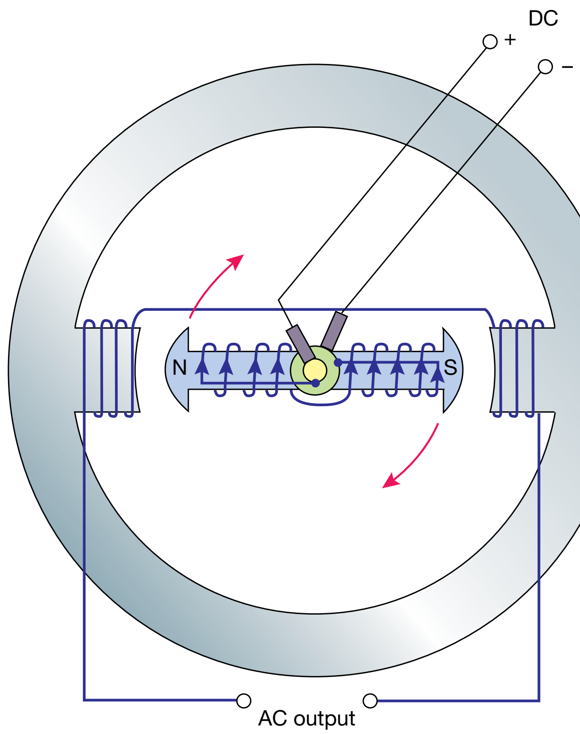
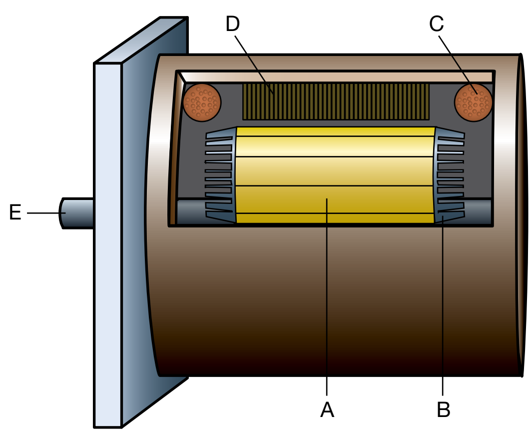


FIGURE 7.47



16. Describe the construction of a squirrel-cage rotor.
17. Explain how a current is produced in the rotor bars of an AC induction motor.
18. Describe the purpose of the end rings of a squirrel cage.
19. (a) What is slip?
    (b) Explain why slip is necessary for the operation of a squirrel-cage induction motor.
20. Account for the 'lost power' of induction motors.

eBookplus RESOURCES

Complete this digital doc: Investigation: A model DC motor
Searchlight ID: doc-26596

Complete this digital doc: Investigation: Demonstrating the principle of an AC induction motor
Searchlight ID: doc-26597

PRACTICAL INVESTIGATIONS

Investigation 7.1 A model DC motor

Aim

To build a model DC motor.

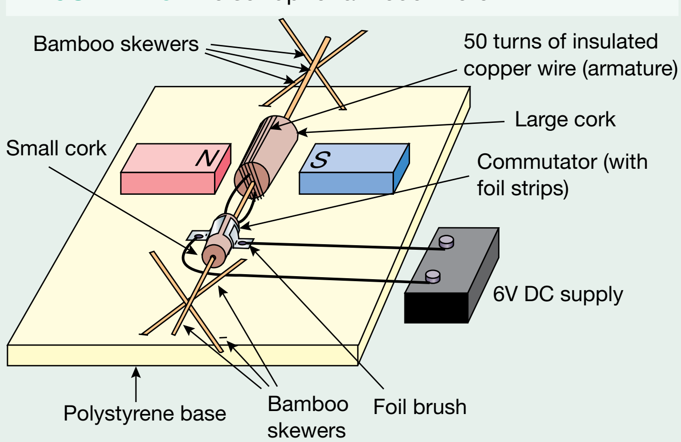
Apparatus

6 V, DC power supply
two bar magnets
thin insulated copper wire
two cylindrical corks, one thin and one thick
aluminium foil
glue
adhesive tape
five bamboo skewers
thick rectangular piece of polystyrene

Method

1. Push one of the skewers through the centres of the corks as shown in Figure 7.48.
2. Glue two pieces of foil onto the small cork with two thin gaps between them to make a split-ring commutator.
3. Wrap the thin copper wire around the thick cork 50 times, as shown. Hold in place with adhesive tape.
4. Strip the ends of the wire and connect to each of the foil strips of the commutator.
5. Make sure that the centres of the commutator strips line up with the centre of the coil windings.
6. Push the other skewers into the foil to support the coil and commutator, as shown.
7. Use the foil to make a set of brushes and use drawing pins to position them so that they touch opposite sides of the commutator, as shown.
8. Position the two magnets on opposite sides of the coil so that a N pole faces a S pole.
9. Connect the DC supply to the brushes and observe the motion of the armature.
10. Vary the spacing of the magnets.
11. Find two ways to reverse the direction of rotation of the armature.

FIGURE 7.48 The set-up for a model motor.



Analysis

1. Describe the effect of varying the spacing of the magnets on the speed of rotation of the armature. Account for this effect.

2. Describe two methods for reversing the direction of rotation of the armature.
3. Identify the role of the following:
    - magnets
    - coil.

Investigation 7.2 Demonstrating the principle of an AC induction motor

Aim

(a) To investigate the direction of a current in a conductor that is moving relative to a magnetic field

(b) To investigate the factors affecting the magnitude of a current in a conductor that is moving relative to a magnetic field.

Apparatus

a copper or aluminium rod. If none are available, a length of copper wire will do.
two bar magnets or a horseshoe magnet
a galvanometer
connecting wires

Theory

A current is induced in a conductor when there is relative movement between it and a magnetic field.

The magnitude depends on the orientation of the conductor to the field, the speed of the relative motion between the conductor and the magnetic field and the strength of the magnetic field.

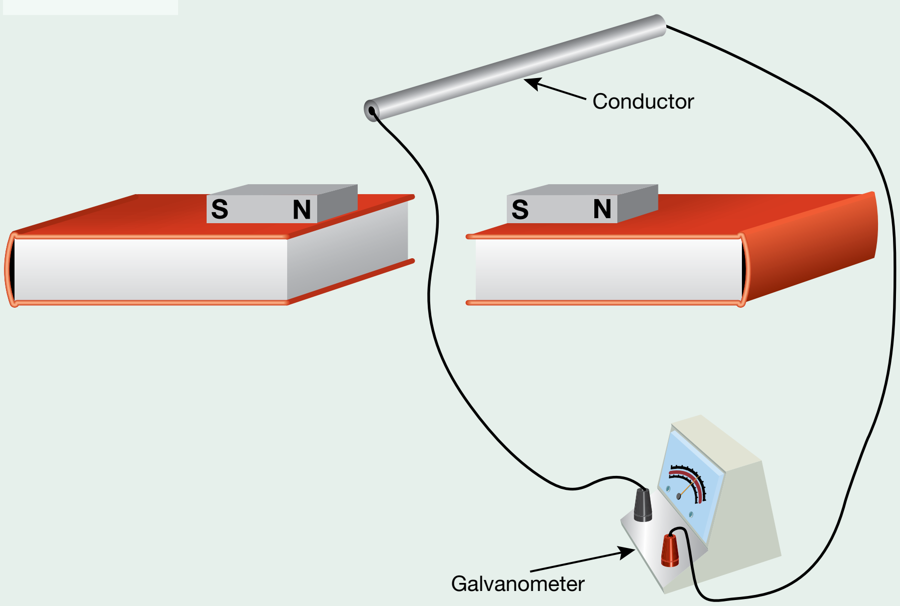
Method

Set up the apparatus as shown in Figure 7.49. Place two bar magnets on two books with an N pole on the left and an S pole on the right.

Note the separation of the poles, as this affects the strength of the magnetic field.

1. Connect the galvanometer to the conductor.
2. Move the conductor downwards slowly between the poles of the magnet. Note the magnitude and direction of the current through the conductor.
3. Move the conductor downwards quickly between the poles of the magnet. Note the magnitude and direction of the current through the conductor.
4. Move the conductor up slowly between the poles of the magnet. Note the magnitude and direction of the current through the conductor.
5. Move the conductor upward quickly between the poles of the magnet. Note the magnitude and direction of the current through the conductor.

FIGURE 7.49



6. Place the conductor at an angle to the magnetic field and move it upwards quickly between the poles of the magnet. Compare the magnitude of the current through the conductor with the result attained in step 5.
7. Double the separation of the magnetic poles. What effect will this have on the strength of the magnetic field?
8. Repeat step 5.
9. Return the magnetic poles to their original separation. Support the conductor and move the poles upwards past the conductor. Do you get the same direction of current as when the conductor moved downwards between the poles?

Analysis

Relate your observations to the theory presented in this topic.

TOPIC 8
Exploring the electromagnetic spectrum

8.1 Overview

8.1.1 Module 7: The nature of light

The electromagnetic spectrum

Inquiry question: What is light?

Students:

- investigate Maxwell's contribution to the classical theory of electromagnetism, including:
  - unification of electricity and magnetism
  - prediction of electromagnetic waves
  - prediction of velocity (ACSPH113)
- describe the production and propagation of electromagnetic waves and relate these processes qualitatively to the predictions made by Maxwell's electromagnetic theory (ACSPH112, ACSPH113)
- conduct investigations of historical and contemporary methods used to determine the speed of light and its current relationship to the measurement of time and distance (ACSPH082)
- conduct an investigation to examine a variety of spectra produced by discharge tubes, reflected sunlight or incandescent filaments
- investigate how spectroscopy can be used to provide information about:
  - the identification of elements
- investigate how the spectra of stars can provide information on:
  - surface temperature
  - rotational and translational velocity
  - density
  - chemical composition

FIGURE 8.1 A close-up of the Milky Way. Why do stars have different colours? What can we learn from the spectra of stars?

8.2 James Clerk Maxwell

8.2.1 Theory of electromagnetism

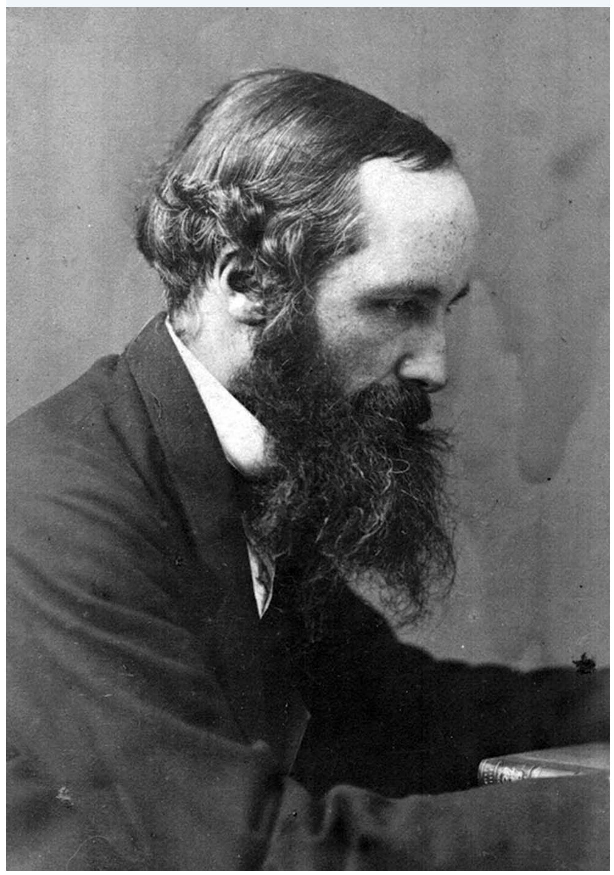
In Topic 4 we learned about electric and magnetic fields, and in Topic 5 we learned that charges moving through a magnetic field experience a force. Topic 6 introduced the concept of electromagnetic induction, and Topic 7 introduced how the forces on moving charges can be used to make electric motors. These discoveries of the early nineteenth century demonstrated that the phenomena of electricity and magnetism, which had been known though natural phenomena such as lightning, electric shocks and magnetic lodestone, were linked; changing magnetic fields causes electric charges to move, and moving electric charges generates magnetic fields.

One of the aims of physics is to come up with theories that make sense of a range of phenomena. The more phenomena that are explained, the more powerful the theory. Earlier topics introduced Isaac Newton's Law of Universal Gravitation and his three laws of motion, which provide a theory for helping to make sense of motion. Without a theory, what we have is experimental or empirical results. We see that a wire in a magnetic field moves when a current is passed through it. Another experiment shows that a loop of wire moving into a magnetic field has a brief flow of current through it. Through analysing the measurements of currents and forces, physicists come up with laws like Faraday's law. A theory takes the understanding to a new level. For example, a theory of electromagnetism links all the phenomena that have been observed and analysed during experimentation. A physicist with a theory can determine what to expect in an experiment. With a theory, the physicist might also be able to predict the existence of phenomena that haven't yet been observed.

Theories need to explain all the related phenomena that have been observed. The phenomena that theories explain can be tested by observing the predicted phenomena. A theory in science is quite different to the way the word *theory* is often used in general speech. 'I have a theory that ...' usually means a guess and might be more accurately described as a hypothesis. In science, a theory must be well tested and have considerable explanatory capability.

James Clerk Maxwell (1831–1879) was a Scottish physicist who developed one of the great theories of physics. His theory of electromagnetism, published in 1864, linked all the electric and magnetic effects that had been observed in earlier experiments and predicted new phenomena. In fact, his theory showed not only how electricity and magnetism are closely connected but how light is too. His theory was able to predict the speed of light and the existence of types of light that had not yet been observed in experiment. The theory of electromagnetism is encapsulated in four equations that are named after physicists and mathematicians who made major contributions to the understanding of the phenomena that the equations describe. Maxwell was able to express all that was known about electricity and magnetism mathematically in these four equations, which is an amazing achievement.

FIGURE 8.2 James Clerk Maxwell brought together all that was known about electricity and magnetism into a unifying theory.



AS A MATTER OF FACT

Maxwell's equations make use of mathematics called vector calculus.

Gauss's law: $\nabla . E = \frac{\rho}{\varepsilon_0}$

Gauss's law for magnetism: $\nabla . B = 0$

Faraday's law: $\nabla \times E = -\frac{\partial B}{\partial t}$

Ampère-Maxwell law: $\nabla \times B = \mu_0 J + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t}$

Maxwell's theory of electromagnetism is summarised by the four Maxwell equations. The actual equations involve vector calculus, which is beyond the scope of this physics course, so we will describe them qualitatively. They are named after those who made earlier contributions to them:

- Gauss's law: This law describes the electric flux produced by electric charges. Electric flux is the electric field multiplied by the area it passes through. Gauss's law states that the electric flux is proportional to the charge producing the flux: $Electric\ flux = \frac{q}{\varepsilon_0}$.

For example, a charge in a box determines the amount of electric flux passing through the box. If the box doubles in surface area, the density of field lines on any section of the box will halve but the total flux is unchanged. The total flux through the box only depends on the charge contained within the box. The field lines from charges external to the box will enter and then leave the box so the total contribution to the flux is zero.

If we consider this for a point charge, we get the special case known as Coulomb's Law. Coulomb's Law says that the strength of the electric field at any point around a point charge is given by $E = \frac{q}{4\pi r^2 \varepsilon_0}$, where E is the electric field strength a distance r from the charge; ε_0 is the permittivity of free space; and q is the charge. ε_0 is an important constant in electromagnetism that we will revisit.

Gauss's law is an excellent example of how a relatively simple idea can be expressed as a law that has far-reaching implications in our understanding of the universe.

The important aspects for us in this topic are that charges produce an electric field and there is an important constant called the permittivity of free space.

- Gauss's law for magnetic fields: This second law is very similar to the first but applies to magnetic rather than electric fields. This law states that the magnetic flux through any closed surface is equal to zero. Unlike the positive and negative charges supplying electric flux, magnetic poles cannot be separated. We will always have both the north pole and the south pole of the magnet.

FIGURE 8.3 The electric flux through a closed surface is determined by the electric charge contained within the closed surface. The red field line is due to charges external to the closed surface, so it passes in and then out of the sphere, making a zero contribution to the electric flux.

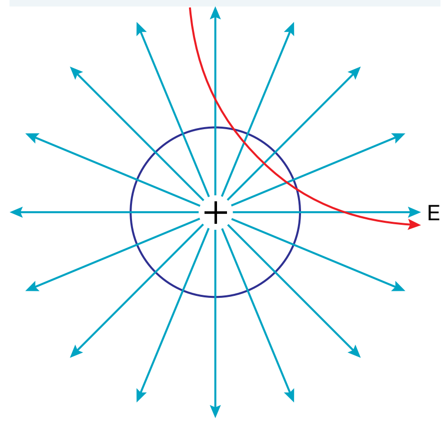
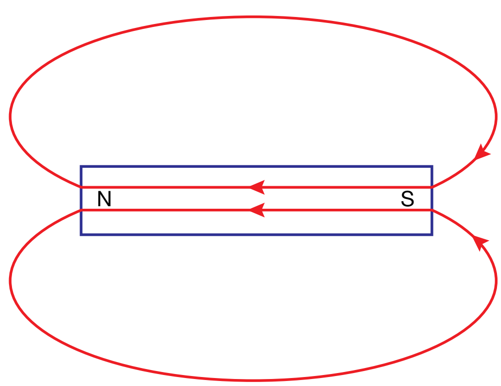


FIGURE 8.4 Gauss's law for magnetic fields states that the magnetic flux through any closed surface is zero. Any field lines leaving the surface are equalled by field lines entering the surface. This is another way of saying that magnetic monopoles are not found in nature, according to Maxwell's theory. So far this agrees with observation.



- Faraday's law: This describes the electric field induced by a changing magnetic field. This electric field does not begin and end in a charge like in Gauss's law but forms a loop. In Topic 6 we saw how Faraday's law can be applied to determine the emf induced by a change in flux through a coil using $\varepsilon = -\dfrac{\Delta\phi}{\Delta t}$.

  Faraday's law links electric and magnetic fields. The *emf* in the equation is determined by the strength of the electric field around the loop and the length of the loop. For a given loop, the *emf* is greater if the electric field around the loop is greater.

  The electric field around the loop is determined by the rate of change of magnetic flux through the loop. This is the right-hand side of the equation. For a given loop at a fixed orientation to the magnetic field, the change in flux is due to the rate of change in magnetic field. If there is no change in the magnetic field, there is no electric field around the loop.

  This is an important result for us in this topic because it shows that changing magnetic fields induce electric fields.
- Ampère-Maxwell law: This is a little like Faraday's law but it deals with changing electric flux. A magnetic field is produced by an electric current or by a changing electric flux. This law introduces another constant, the permeability of free space, μ_0, which has an exact value and indicates the direct proportions between the flux and the magnetic field.

Overall, the impact of Maxwell's theory of electromagnetism is that electricity and magnetism cannot be viewed as two separate phenomena. They are both features of a more general aspect of physics called electromagnetism. Physical theories have this unifying effect: the better the theory, the more aspects of nature it includes.

8.2 SAMPLE PROBLEM 1

A scientific theory shows the consistency between different observations and experimental results. What are some of the discrete phenomena linked by Maxwell's theory of electromagnetism?

SOLUTION:

Maxwell's theory links the phenomena of electric currents, forces between charges, magnetic forces, magnetic fields produced by electric currents, electric fields produced by changes in magnetic fields, and forces acting between electric currents.

8.2 SAMPLE PROBLEM 2

Explain what is meant by the permittivity of free space.

SOLUTION:

The permittivity of free space is the constant of proportionality between the electric flux passing through an enclosed space and the charge enclosed by the space.

8.2.2 An electromagnetic wave

Maxwell's theory synthesised all that was known about electrical and magnetic physics into four equations, but he went further. He showed mathematically, using Faraday's law and Ampère's law, that an electromagnetic wave was expected. A key strength of a good scientific theory is its ability to predict phenomena that are either previously not observed or not known to be related. The electromagnetic wave turned out to be a triumph of electromagnetic theory.

Waves, like particles, are an important category of phenomena in physics. The idea of a particle is quite well understood, a lump of matter that can be considered without concern for its structure. Balls or marbles are models of particles. By throwing a ball across a basketball court, energy is transferred along with the transfer of matter. By contrast, a wave is where a disturbance in a medium transfers energy but not matter.

Waves are visible in slinky springs. The spring acts as the medium. If we stretch out the spring and disturb one end by shaking it up and down, the disturbance travels along to the other end of the spring. However, no matter passes from the shaken end to the other end of the spring; the parts of the spring that moved up and down are still there at that end of the spring.

FIGURE 8.5 (a) A particle transfers matter and energy. (b) A wave transfers energy without a transfer of matter.

In Maxwell's time, it was assumed that a medium was required for electric and magnetic fields. People at the time called this supposed medium the luminiferous aether, but it was never observed. An electromagnetic wave is not a ripple in a slinky string or on the surface of water but, instead, is an oscillating electromagnetic field. By Maxwell's equations, this wave would perpetuate itself as changes in the electric field induce changes in the magnetic field. Waves that require a medium are called mechanical waves to distinguish them from electromagnetic waves.

8.2 SAMPLE PROBLEM 3

What are the characteristics of a wave?

SOLUTION:

Waves carry energy from one place to another without the transfer of matter.

8.2.3 The speed of electromagnetic waves

Waves can be mathematically described by the wave equation. Maxwell was able to produce wave equations using the Ampère-Maxwell law and Faraday's law. Wave equations include an expression for the speed of the wave. The speed of these electromagnetic waves was given by the following equation:

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

8.2 SAMPLE PROBLEM 4

Calculate the speed of electromagnetic waves given that:

i. The permittivity of free space is 8.854×10^{-12} C^2 s^2 kg^{-1} m^{-3}

ii. The permeability of free space is $4\pi \times 10^{-7}$ m kg C^{-2}

SOLUTION:

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

$$v = \frac{1}{\sqrt{4\pi \times 10^{-7} \times 8.854 \times 10^{-12}}}$$

$$= 2.998 \times 10^8 \text{ m s}^{-1}$$

Normally this figure is rounded to 3.0×10^8 m s^{-1} but the exact value is 299 792 458 m s^{-1}.
This figure is equal to the speed of light, a fact that was noticed by Maxwell and led to his proposal that light was an electromagnetic wave. He also proposed that the light we see is restricted to a particular range of wavelengths and that there would be other wavelengths of electromagnetic waves that are invisible to the naked eye. These have since been discovered and named in order from shortest wavelength to longest: gamma rays, X-rays, ultraviolet rays, visible light, infra-red radiation, microwave radiation and radio waves. This discovery has totally transformed communications, entertainment, medicine, astronomy and our understanding of the atom in the years since.

The shorter the wavelength, the greater the energy the wave carries if the intensity is constant. This energy is related to its ionising power, with gamma rays being highly ionising (and therefore damaging to life and other materials) and X-rays being less highly ionising. Ultraviolet light is also ionising, but visible light is not. Infra-red, microwaves and radio waves are not ionising. They simply do not deliver enough energy to individual electrons to allow them to escape their atoms. This does not mean, however, that these wavelengths are safe. At sufficient intensities, they are still able to cause burns. For example, infra-red rays from a fire are extremely dangerous and are often the cause of death in bushfires. However, the infra-red rays from the Sun are not sufficient to burn our skin. The damage to our skin in terms of sunburn and ultimately wrinkling, spots and skin cancer are the result of damage caused by ultraviolet rays from the Sun.

8.2 Exercise 1

1 Which of Maxwell's laws predicts the forces between two electric charges?
2 Explain what is meant by the permeability of free space.
3 What are the characteristics of an electromagnetic wave?
4 What is the medium for an electromagnetic wave?
5 The speed of sound (also modelled as a wave) in air is approximately 340 m s^{-1}. Compare the speed of sound with the speed of electromagnetic waves.
6 What is the speed of ultraviolet waves?

8.3 The production and propagation of electromagnetic waves

8.3.1 How to make an electromagnetic wave

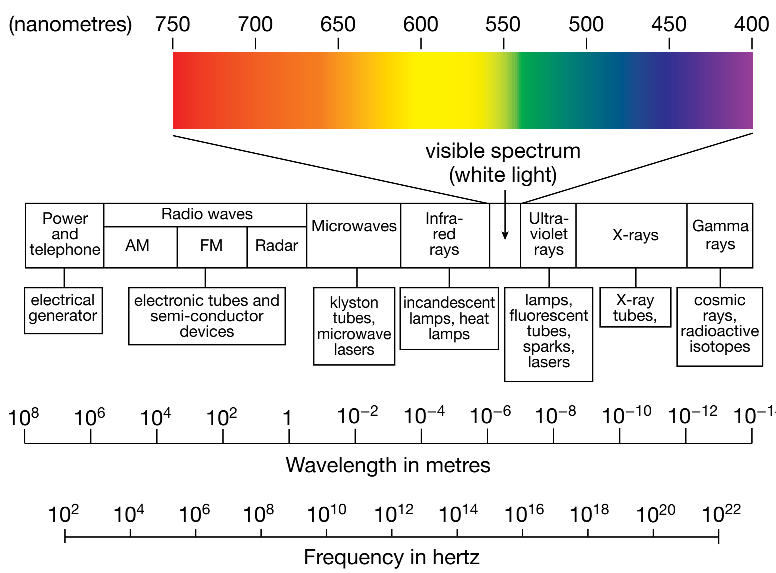
Maxwell's theory of electromagnetism showed that magnetic fields are formed in the presence of moving charges (an electric current). It also showed that magnetic fields can be produced from changing electric fields and that electric fields are produced by changing magnetic fields. The magnetic field is positioned at right angles to the electric field. To produce an electromagnetic wave, which propagates through space, an oscillating (vibrating) electric charge is required.

Imagine a charged sphere hanging from a spring. Pull the sphere down and let go. The sphere would oscillate up and down on the spring. The moving charge forms an electric current that is oscillating in magnitude and direction. This will induce a magnetic field that oscillates in magnitude and direction in proportion with the current. This changing magnetic field induces an electric field whose magnitude varies in proportion with the magnitude of the electric field. These oscillating fields are self-perpetuating and radiate out through space at a speed of $v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 2.998 \times 10^8\ \mathrm{m\,s^{-1}}$.

Charged spheres on springs are not the normal source of electromagnetic waves. There are many ways that charged particles are accelerated. For example, electrons moving in a radio transmitter antenna oscillate in response to an electrical signal produced by a sound recording or a disc jockey speaking into the microphone.

The acceleration of the electrons in different situations determines the wavelength of the electromagnetic waves produced. Ranges of wavelengths are categorised based on how they are produced and how they interact with matter. Visible light and infra-red were the only parts of the spectrum that scientists were aware of when Maxwell predicted the existence of electromagnetic waves. The others were discovered soon afterwards — in particular, Heinrich Hertz produced radio waves in 1887 by deliberately setting up the conditions to produce electromagnetic waves.

FIGURE 8.6 Forms of radiation and their place in the electromagnetic spectrum. The visible portion of the spectrum is shown enlarged in the upper part of the diagram.



8.3 SAMPLE PROBLEM 1

1. How could electromagnetic waves be produced by high-voltage power lines?
2. Does a constant DC current produce electromagnetic waves?

SOLUTION:

1. High-voltage power lines contain electrons moving backwards and forwards in an alternating current. These electrons, which move with a frequency of 50 to 60 Hz, will generate electromagnetic waves with a frequency of 50–60 Hz.

2. No. Changing electric fields produces changing magnetic fields. Constant magnetic and electric fields will not produce an electromagnetic wave.

8.3 Exercise 1

1 In cities where there are overhead power lines for trains, trams and electric buses, car drivers notice interference in their AM radio reception. These systems usually make use of direct current (DC) electricity. What might cause this interference?

A major source of electromagnetic radiation in the universe is the capture of electrons or the transition of electrons within atoms to lower energy levels. This will be addressed in greater detail in later topics but requires a quantum mechanical approach. Maxwell's electromagnetism is part of what is known as classical physics. This is a model that works well in many situations but, as physicists learned more about atoms and the way that they interact with light, they had to develop quantum theory, which treats light not as an electromagnetic wave but as a stream of photons or an excitation in what is known as the quantum field. The photon model retains elements of the wave model with its frequencies and wavelengths.

8.3.2 How electromagnetic waves travel

Electromagnetic waves contrast with mechanical waves resulting from the physical disturbance of a medium. The luminiferous aether considered to be the medium for electromagnetic waves in the nineteenth century was found to be non-existent and unnecessary. Electromagnetic waves travel most effectively when there is no material medium, such as in interstellar space.

When travelling through transparent media such as air, water or glass, an electromagnetic wave takes longer to cover a distance than it would in the near vacuum of space. This can be explained by the electromagnetic radiation interacting with atoms in the medium. The atoms absorb the energy of the wave, setting electrons within the atoms in motion, oscillating at the same frequency as the electromagnetic wave until they re-emit a wave a short time later. This slight delay slows down the propagation of the wave. The amount that the wave is delayed depends on features of the medium such as the distance between the atoms and the properties of the atoms in the medium.

This difference between transparent media is referred to as optical density. The more optically dense the medium, the greater the delay in the journey that light takes through the medium. Optical density is quantified by the refractive index of the material. We met this in Unit 1 and 2 with the formula $n = \frac{c}{v}$, where n is the refractive index, c is the speed of the electromagnetic radiation in a vacuum, and v is the effective speed of the radiation through the transparent material.

8.3 SAMPLE PROBLEM 2

An electromagnetic wave in the radio part of the spectrum with a frequency of 100 kHz is incident on an antenna for a radio receiver. What is the frequency that the electrons in the antenna will oscillate at due to the reception of these radio waves?

SOLUTION:

The electrons will oscillate at 100 kHz.

8.4 Measurement and the speed of light

8.4.1 Early measurements

In 1864, Maxwell predicted that light was an electromagnetic wave. This would be magnificent validation of the theory of electromagnetism. The electromagnetic waves that Maxwell predicted travelled at nearly $3 \times 10^8 \text{ m s}^{-1}$ according to his mathematical analysis. How did that match up with the experimental values for the speed of light?

Attempts to measure the speed of light can be traced back at least to Galileo Galilei over two hundred years earlier. Speed calculations require a method of measuring distance and time accurately. Galileo worked with an assistant, each carrying a lamp to the top of a hill a kilometre or so apart. Galileo was limited in his timekeeping apparatus but may have used a clock that worked on the basis of how much water or sand poured out in the time being measured. He uncovered his lantern at the same time as he started his clock. His assistant then uncovered his lantern as soon as he saw Galileo's lantern. Galileo then stopped his timing when he saw his assistant's lantern. Galileo determined that light must travel at more than ten times the speed of sound, but his method was inadequate to determine more than this.

FIGURE 8.7 Galileo used this method to measure the speed of light. He attempted to time, with his pulse, the delay between uncovering his lantern and seeing the light from his partner's lantern, which his partner uncovered at the moment when he saw the light from Galileo's lantern.

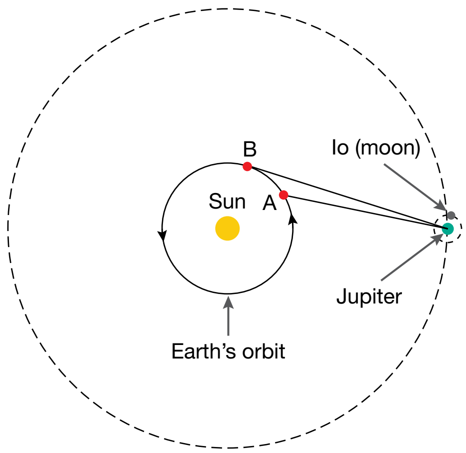
About fifty years later, at the Observatory of Paris, Ole Römer stumbled across a practical consequence of the finite speed of light that allowed him to calculate its value. The problem with Galileo's method was that the time taken for light to cover the distance was far too small to measure with the equipment available. Römer was able to increase this distance millions of times using the motion of the Earth.

What Römer was trying to do was use the regular eclipse of Jupiter's satellite Io, as it passed behind Jupiter, as a clock to help explorers determine what longitude they were at. Determining longitude was a major navigational problem before accurate clocks were invented. Early explorers of inland Australia, Burke and Wills, used a navigation guide that described this technique involving Io.

Römer noticed that the average period of revolution of Io appeared to change depending on the time of the year. At one point in Earth's orbit around the Sun, he measured the time for Io to complete four orbits of Jupiter by timing how long it took between successive appearances from behind Jupiter. This gave an average of 42 hours, 28 minutes and 31.25 seconds for the period of each orbit. Then Römer observed Io at 26 orbits of Jupiter later and found a difference between the time he expected Io to emerge from behind Jupiter and the time it actually appeared: a quarter of an hour.

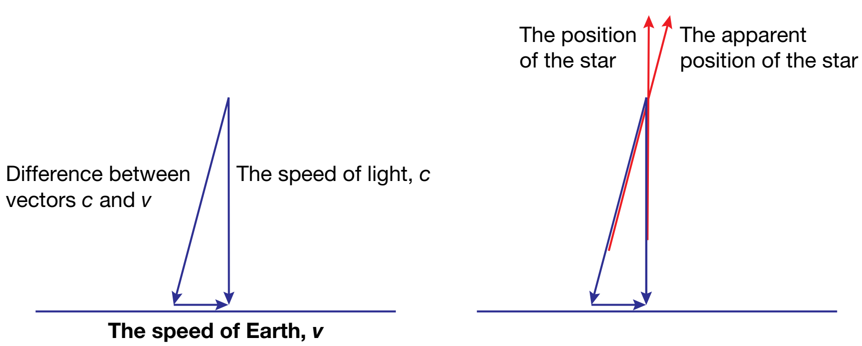
This was completely unexpected but Römer explained it. He accounted for the time difference by considering the motion of the Earth around the Sun and the finite speed of light. The first measurements were made when Earth was at its closest to Jupiter. By the time of the last measurement, Earth had moved in its orbit so that the distance to Jupiter was greater. Each time Io emerged from behind Jupiter, Earth had moved farther away and the light had to travel farther to reach him, adding over thirty seconds to the measured period of Io's orbit. A thirty-second difference would have been difficult to measure with confidence but, after 30 of these cycles, the time between the expected emergence of Io and when it actually became visible was a quarter of an hour. Using the best measurements for the diameter of Earth's orbit, scientists were able to determine the speed of light to be 210 000 kilometres per second.

FIGURE 8.8 It takes Io over 42 hours to orbit Jupiter. In that time, Jupiter's distance from the Sun has barely changed. However, Earth has moved in its orbit from point A to point B so that it is now significantly farther from Jupiter, resulting in light taking 30 seconds longer to reach the Earth. Therefore, the time observed for Io to orbit Jupiter is its period plus 30 seconds. As Earth continues to move away from Jupiter, this difference continues so that, if we consider 30 orbits of Io, as Römer did, Io emerges 15 minutes late from behind Jupiter based on what would be expected if the speed of light did not need to be considered.



More than fifty years later, in 1728, James Bradley found another practical implication of the speed of light. While trying to measure the parallax of stars due to the Earth revolving around the Sun, he instead observed an effect known as stellar aberration. This effect is the slight changes in the positions of stars overhead depending on the time of the year. As with falling rain, the light appears to be arriving from further in front, the faster you are moving. This effect would not exist if light travelled infinitely fast. Bradley determined light to be travelling at 3×10^8 m s^{-1} to cause the effect he saw.

FIGURE 8.9 Due to the Earth's motion through space, the position of an overhead star appears very slightly shifted. Bradley used this change in position and the known speed of the Earth to determine the speed of light accurately.



8.4.2 Modern values

Römer and Bradley's measurements were what we call indirect measurements. The phenomena they observed were explained by giving light a particular speed. In 1849, French physicist Hippolyte Fizeau invented a direct method. He had a disc with slots cut into it at regular intervals attached to a gear system that allowed him to rotate it very fast. He could then shine light through a slot in the disc. He placed a mirror about eight kilometres away to reflect the light back so that it came right back through the slots when he rotated the disc slowly. As he increased the speed of the disc, it started to appear transparent. As he continued to increase the speed, he reached a speed where the light no longer came back through the disc — it lost its transparency. By the time the light returned to the disc from its 16-kilometre journey, the disc had rotated slightly so that it blocked the light. Using the rotation rate of the disc and the fraction of the circumference needed for the disc to block the returning light, he could determine the time that the light took on its 16-kilometre journey. Using speed equals distance over time, he calculated the speed of light to be 313 000 km s^{-1}.

Another Frenchman, Leon Foucault, used a similar method, but instead of a rotating disc he used a rotating mirror. The returning light was reflected at an angle determined by how far the mirror had rotated during its journey. By 1862, Foucault had refined his measurements and determined the speed of light to be 298 000 km s^{-1}, extremely close to today's accepted value of 299 792.458 km s^{-1}

FIGURE 8.10 Fizeau used a disc with slots cut into it to help him time how long it took for light to travel across Paris.

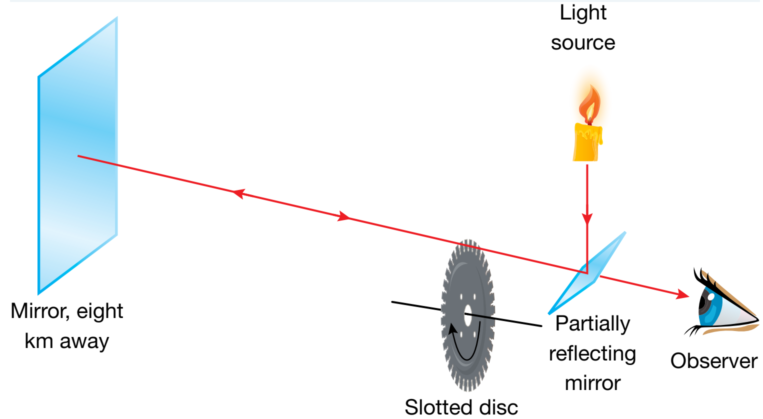
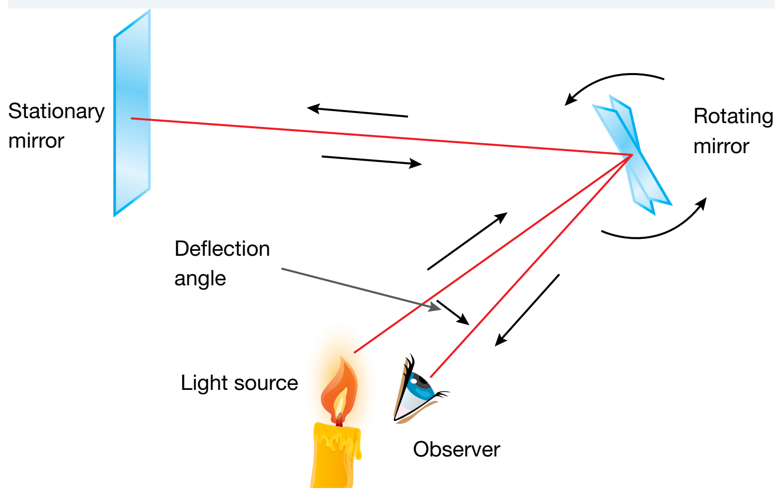


FIGURE 8.11 The reflected light in Foucault's apparatus was deflected by an angle that depended on how far the mirror had rotated in the time the light took to return to the mirror.



8.4 SAMPLE PROBLEM 1

1. What speed did Maxwell's equations predict for light?

SOLUTION:

Maxwell's equations predicted that light would travel at 2.998×10^8 m s^{-1}.

2. If Galileo had known the speed of light was about 3×10^8 m s^{-1}, estimate, to one significant figure, the time Galileo would have expected to elapse between the time he shone his lantern and when he saw his assistant shine the second lantern.

SOLUTION:

$$t = \frac{d}{v} = \frac{2000}{300\,000\,000} = 7 \times 10^{-6}\text{ s}$$

3. Using a technique like that of Foucault, what could be done to improve the precision of the measurement?

SOLUTION:

Using Foucault's apparatus, measuring an angle was required. The smaller the angle, the more uncertainty there was in the measurement. The size of the angle would depend on two factors: 1. the distance the light travelled, and 2. the rate of rotation of the mirror. Increasing either or both these factors would increase the size of the angle of deflection and therefore the precision for the speed of light. Any reduction in errors in the measurement of the distance that the light travelled or the rate of rotation of the mirror would also reduce the error in the measured value.

8.4 Exercise 1

1 Why was the measurement of the speed of light important for Maxwell's theory of electromagnetism?

2 If the slots in Fizeau's disc were 1 mm wide and the light shone through them at a radius of 20 cm from the centre of the disc, what is the minimum number of times the disc would need to rotate per second to block the light coming back from its 16 km journey?

3 What do you imagine were some of the difficulties that Fizeau and Foucault had to overcome to make their measurements?

Early in the twentieth century, the American scientist Albert A. Michelson (1852–1931) used a rapidly rotating eight-sided mirror. The light was reflected to a distant mirror approximately 35 kilometres away and then reflected back to the rotating mirror. For some particular rotation rates, this light was reflected by only one of the sides of the rotating mirror directly to the observer. The rotation rate can be used to calculate the speed of light. The value Michelson obtained was $2.99\,796 \times 10^8$ m s^{-1}. He actually measured the distance of 35 km to an accuracy of 2.5 cm.

8.4.3 The standard speed of light

In 1983, the speed of light was declared to be exactly 299 792 458 m s^{-1}. Previous values were based on definitions of the metre and the second. The standard for the metre had varied in the centuries since it had been devised. Given that the speed of light was constant, it was determined that the speed of light was a precise standard that everyone could agree on for the metre.

With the speed of light set at 299 792 458 m s^{-1}, one metre is defined to be how far light travels in $\frac{1}{299\,792\,458}$ seconds.

But what is one second? This is also defined against a very precise standard in nature. Caesium 133 atoms have an energy transition that occurs 9 192 631 770 times per second. This value has been found to be highly reliable and is the basis of the world's most accurate clocks, known as atomic clocks. One second is defined as the time taken for 9 192 631 770 of these transitions.

8.4 SAMPLE PROBLEM 2

1. Calculate how long it would take for light to travel 1 km.

SOLUTION:

$$t = \frac{d}{v} = \frac{1000}{299\,792\,458} = 3.336 \times 10^{-6}\ \text{s}$$

2. Why is it important to have a standard of measurement?

SOLUTION:

Whenever we measure something, we use a standard. When we say something is three metres long, we rely on having a common understanding of what a metre is. The speed of light provides a very precise standard for the metre.

8.4 Exercise 2

1 Calculate how far light would travel in one year.
2 What is the standard used for defining one second?

8.5 Exploring with the electromagnetic spectrum

8.5.1 Spectroscopes and spectroscopy

Isaac Newton (1642–1727) published his book *Opticks* in 1704. In the first volume, he demonstrated that light from the Sun can be dispersed into its constituent colours. Other theories about why rainbows formed, why prisms of glass produced a spectrum of colours and why soap bubbles appeared coloured involved the prism, raindrop or bubble altering the light. However, as Newton demonstrated, the prism, raindrop or bubble simply disperse the light according to its colour (wavelength), revealing information about the Sun. Newton's prisms showed the colour spectrum from the Sun to contain red, orange, yellow, green, blue, indigo and violet.

Note: the spectrum of the Sun is not observed directly but projected onto a screen to protect the observer's eyesight.

In 1802, William Wollaston (1766–1828) invented the spectroscope in an effort to explore the spectrum in more detail. He found the solar spectrum was not continuous but was crossed by a number of black lines. In 1814, Joseph von Fraunhofer (1787–1826) mapped the spectrum in much greater detail, finding 576 black lines. These have become known as Fraunhofer lines.

Spectroscopes use any process that disperses light into its constituent wavelengths. Typically, a prism is used. Light entering a medium such as glass is refracted towards the normal to the surface. Different wavelengths of light are refracted in different amounts, resulting in the light being dispersed, revealing its colours.

Another technique to disperse light is the diffraction grating. This technique uses the phenomenon that light passing through small holes in an object diffracts and that the diffraction depends on the wavelength of the light.

FIGURE 8.12 A continuous spectrum produced by a spectroscope.

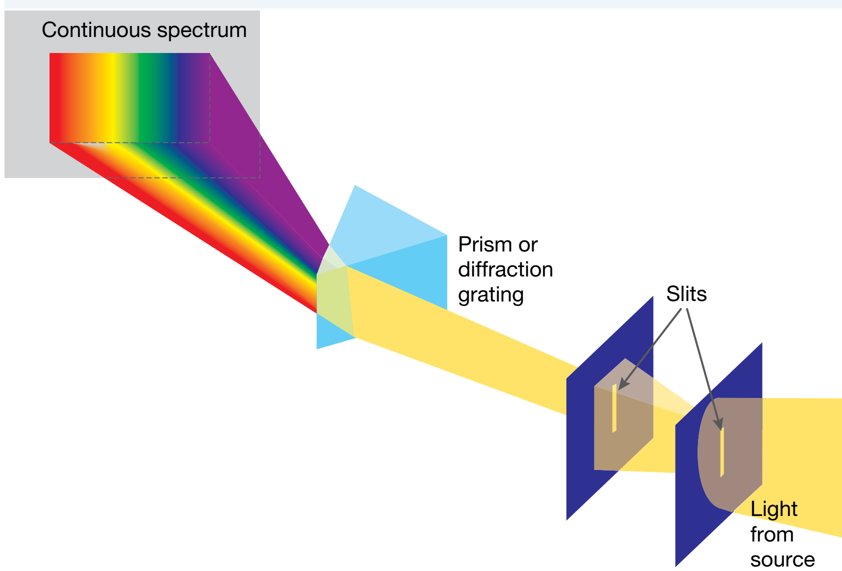


FIGURE 8.13 An absorption spectrum produced by shining light with a continuous spectrum through a cool gas.

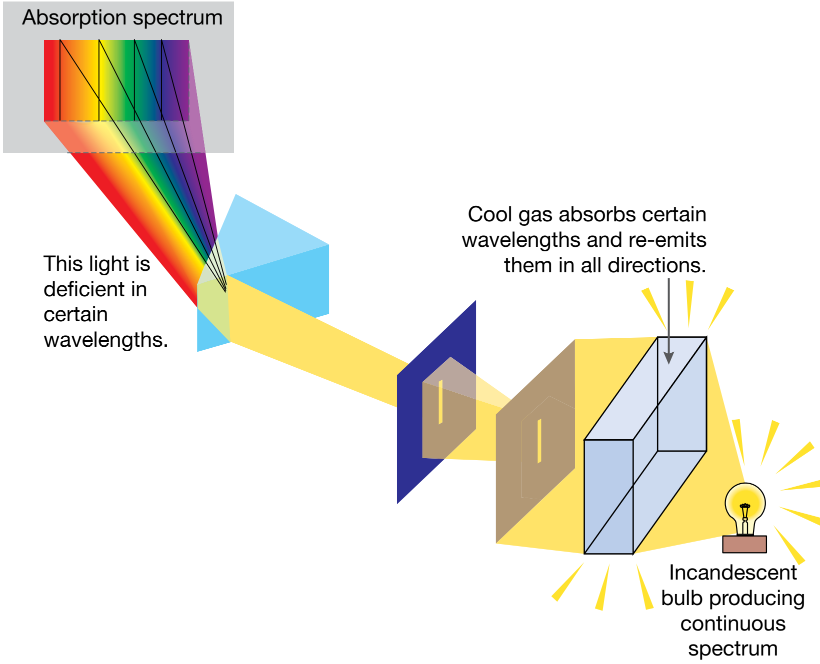


FIGURE 8.14 The absorption spectrum for hydrogen.

8.5 SAMPLE PROBLEM 1

1. What is the purpose of a spectroscope?

SOLUTION:
A spectroscope is an instrument designed to examine the spectrum of light coming from a source.

2. How does a spectroscope disperse light into its constituent colours?

SOLUTION:
A spectroscope uses the fact that a glass prism refracts different wavelengths of light by different amounts. When light enters the spectroscope, it is separated according to the wavelengths that make it up. This effect can also be achieved using a diffraction grating.

8.5.2 Discharge tubes

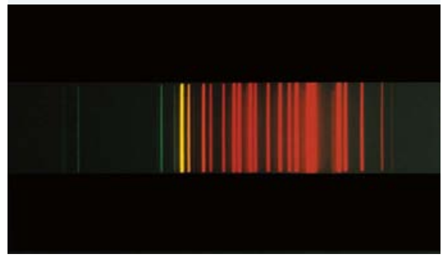
Discharge tubes are glass tubes that have most of the air pumped out of them. An electrode is placed at each end and a large DC voltage is applied. Above a certain voltage, atoms in the gas in the tube are ionised and a stream of electrons flows from the negative electrode (cathode) to the positive electrode (anode). These beams were first observed before the discovery of the electron and called cathode rays. This was the basis for old televisions and computer monitors (prior to flat screen televisions).

In this process, these tubes fluoresce as electrons in the atoms of the gas change energy levels. Different gases can be used in discharge tubes to produce different colours. Fluorescent lights and neon signs are examples of discharge tubes. Fluorescent lights emit in the ultraviolet part of the spectrum, but this part of the spectrum is invisible to our eyes. To get around this problem, the glass tubes are coated with a material that absorbs the UV light and re-emits it in the visible part of the spectrum.

The light produced by a discharge tube is known as an emission spectrum when passed through a spectroscope. Only colours corresponding to particular wavelengths of light are produced. Neon, for example, has emission lines in the blue, green, yellow, orange and red parts of the visible spectrum. However, the net result of this combination appears red to our eyes.

Discharge tubes provided further evidence of the electromagnetic waves that Maxwell predicted. William Röntgen was studying the effects of a discharge tube in 1895. He worked in a darkened room with the discharge tube covered. Two metres away, he happened to have a screen covered in fluorescent material that he noticed would glow when he applied a very high voltage across the electrodes of his tube. He discovered there was a form of radiation produced by the discharge tube that could pass through the covers over the tube. He then took photographs of his wife's hand using this 'X-ray' light, showing that the X-rays passed through her skin but not so well through her bones or her wedding ring. This new discovery was almost immediately put to use in medicine.

FIGURE 8.15 Discharge tubes like neon signs produce emission spectra.



The high-energy electromagnetic waves observed by Röntgen were caused by the high energies of the electrons resulting from the high voltage he applied to the electrodes. These electrons accelerated rapidly and stopped when they hit the anode, releasing energy in the form of X-rays.

FIGURE 8.16 The colours in neon signs are produced by including the gases of different elements.

FIGURE 8.17 William Röntgen's 1895 X-rays passed through his wife's skin but not so well through her bones or her wedding ring.

8.5 SAMPLE PROBLEM 2

1. Compare the features of an absorption spectrum for neon and its emission spectrum.

   SOLUTION:
   The absorption spectrum is a continuous spectrum with dark lines cutting through it. The emission spectrum only has light at particular frequencies. These correspond with the dark lines in the absorption spectrum.

2. The kinetic energy of an electron accelerated by a voltage is given by $E_k = \mathrm{q}V$ where q is the charge on the electron, 1.6×10^{-19} C. Calculate the energy of an electron accelerated by 100 kV.

   SOLUTION:
   $E_k = \mathrm{q}V = 1.6 \times 10^{-19} \times 100 \times 10^3 = 1.6 \times 10^{-14}$ J (or 100 keV).

8.5.3 Reflected sunlight

Most of what we see during daytime is due to sunlight reflecting off a surface into our eyes. The sunlight arrives as a nearly white light. When it reflects, it shows all the different colours that we see. As Newton discovered over three hundred years ago, the light from the Sun contains a full spectrum of colours. What we see in reflected light is the combination of colours that reflect. Some of the wavelengths are absorbed by molecules in the material we are looking at. The molecules absorb some frequencies and reflect others. This results in some parts of the spectrum being more intense than others in producing the colours that appear.

FIGURE 8.18 The glow of an incandescent light globe is caused by an electric current heating up a tungsten wire filament.

8.5.4 Incandescent filaments

Incandescence is the effect of dense objects emitting electromagnetic radiation due to their temperature. A continuous spectrum is emitted but this spectrum is mostly in the infra-red and needs to be very hot before there is enough intensity in the visible part of the spectrum for it to be seen. Incandescent light bulbs were a major source of lighting until fluorescent and LED lighting replaced them. Fluorescent and LED lights can be made to produce wavelengths predominantly in the visible part of the spectrum, making them much more energy efficient than incandescent globes.

Incandescent globes emit electromagnetic waves that peak in intensity in the infra-red part of the spectrum. As a result, the ‘white’ light they produce is stronger in the red end of the spectrum. This gives the light they produce a warmth that is more appealing than many of the fluorescent lights. Fluorescent lights often come with a ‘warm white’ option that tries to mimic the experience provided by incandescent lights.

The electrons in the tungsten metal of incandescent light bulbs are able to move relatively freely in the metal so they can produce electromagnetic waves at a wide range of frequencies, unlike those in gases, which are restricted to the allowed energies of the atoms of the gas.

FIGURE 8.19 A comparison of the 'warm white' and 'cool white' light produced by an incandescent bulb on the right and a compact fluorescent bulb on the left.

8.5 SAMPLE PROBLEM 3

Compare the spectra of light from fluorescent and incandescent lights.

SOLUTION:

The spectrum from a fluorescent light is a series of emission lines. The spectrum from an incandescent globe is continuous.

8.5 Exercise 1

1. What did Wollaston observe in the solar spectrum?
2. When do these features appear in a spectrum?
3. Compare the energy per electron in a discharge tube producing X-rays with those of electrons in a radio antenna.
4. Contrast what a scientist would see viewing the fluorescent light from a neon discharge tube through a spectroscope to the light from a star.
5. Investigate why it is difficult to replicate the light of an incandescent lamp using fluorescent or LED lights.

8.6 Signatures of the elements

8.6.1 Unique signatures

Spectroscopy, the analysis of the electromagnetic spectrum, has proved to be a very powerful tool in science. It has helped reveal the structure of atoms and the elements that make up stars and galaxies. In 1859, Gustav Kirchhoff (1824–1887) with his friend Robert Bunsen (1811–1899) used Bunsen's burner to burn elements and clearly describe the cause of these spectral lines. They found that:

- a continuous colour spectrum is produced by glowing solids or dense gaseous bodies like the Sun (a continuous black body spectrum)

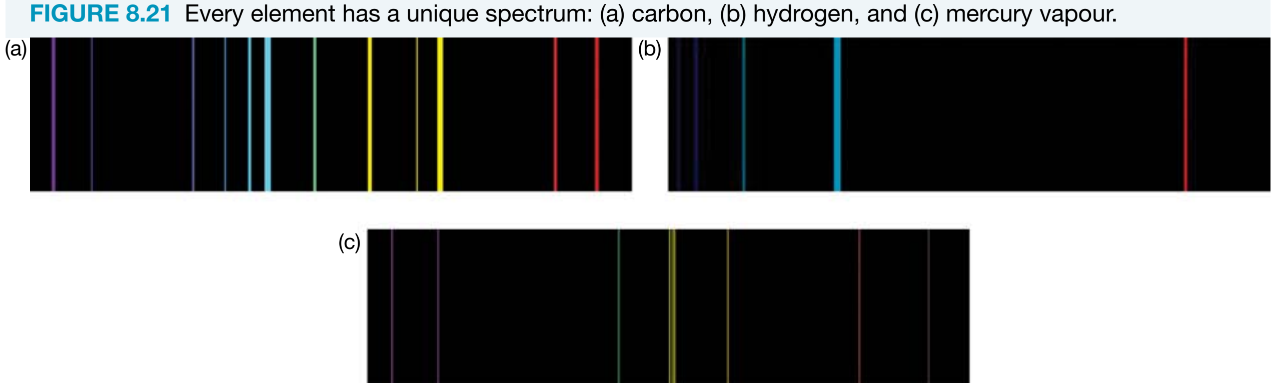
- if a gas exists between the light source and the spectroscope, light is absorbed from the continuous spectrum at wavelengths or colours characteristic of the chemical components of the gas (an absorption spectrum)
- a glowing gas produces bright lines on a dark background at wavelengths or colours characteristic of the chemical components of the gas (an emission spectrum).

They also found that the lines produced in the absorption and emission spectra were characteristic of the atoms that were heated or that the light passed through. No two elements produced the same spectrum, but every atom of a particular element produced the same spectrum. This made the spectroscope a powerful tool for analysing chemicals in the laboratory and in stars to determine the elements they contained. It also helped develop the understanding of the structure of the atom and introduce a whole new understanding of physics: quantum mechanics.

In 1868, Joseph Norman Lockyer (1836–1920) detected some Fraunhofer lines in the solar spectrum that did not correspond with any known element on Earth. He predicted that there must be an as yet undiscovered element in the Sun and called it helium, after the Greek *helios*, meaning Sun. William Ramsay (1852–1916) confirmed this in 1895, when he isolated the gas helium in the laboratory. Although the second most common element in the universe, helium is rare on Earth because it has so little mass that, even at normal temperatures, it has sufficient energy to escape the atmosphere, a property that makes it useful in blimps and party balloons. Helium is also extremely unreactive with other chemicals, so it forms few compounds, unlike hydrogen, which is even less massive but occurs in many compounds on Earth, including water and organic compounds.

FIGURE 8.20 Burning elements in a flame is called a flame test. Different elements can produce different and vivid colours when burned in a flame.

FIGURE 8.21 Every element has a unique spectrum: (a) carbon, (b) hydrogen, and (c) mercury vapour.



8.7 Stellar spectra

8.7.1 Surface temperature

In the 1920s, astronomers such as Cecilia Payne determined how stellar temperatures related to spectra. The spectrum of a star is not of equal intensity for all colours. The astronomers found that hot stars radiate more energy at short wavelengths than cooler stars. Short wavelengths correspond to the blue end of the visible spectrum, while longer wavelengths correspond to the red end of the spectrum.

We are all familiar with this relationship between colour and temperature. In the school laboratory, you have probably used a Bunsen burner. These burners have two settings, one a cool, yellow flame and the other a hot, blue flame. The flames emit more than just the colours we see. If you hold your hand in the air about 30 cm from the flame, you can feel heat, indicating that the flame is emitting infra-red radiation that we can feel but cannot see.

Astronomers generally like to use the Kelvin temperature scale. Every degree in temperature difference is the same as the Celsius scale, but 0 K is set at the coldest possible temperature, which corresponds to −273.15 °C. This means that 0 °C is 273.15 K.

On a clear night it is possible to see some variation in the colour of stars, but the rods in the retinas of our eyes that are responsible for distinguishing colours are not very sensitive to dim light. A photograph will show the colours much more clearly. We notice that some stars are red and some are white or blue. The Sun is a yellow star, indicating that it is neither particularly hot nor cool in the range of star temperatures. The colour of a star indicates the area of the spectrum of the star that is most intense. Some stars are so hot that they emit most of their radiation at very short wavelengths of ultraviolet light, making them invisible from Earth. They must be observed using UV telescopes in orbit because the atmosphere absorbs most UV radiation, preventing it from reaching ground-based telescopes.

The temperature of a star's outer layers determines its colour. The core of the star is much hotter than the outer layers, due to fusion reactions and gravitational energy.

FIGURE 8.22 Emission at different wavelengths for objects of different temperatures.

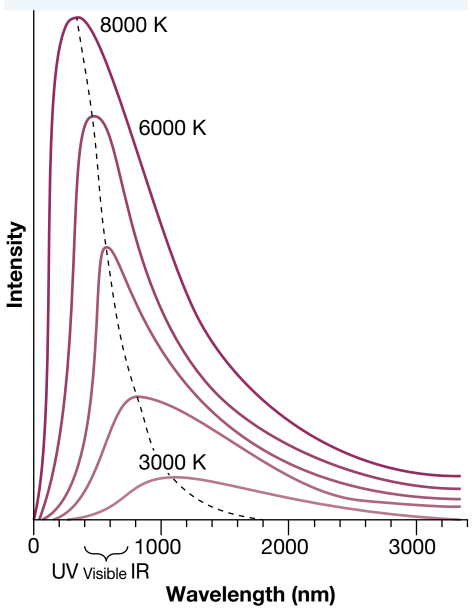


FIGURE 8.23 A portion of the night sky including the constellation of Orion. Four of the brightest stars in the sky can be seen. Notice the different colours: Betelgeuse is a red supergiant, Rigel is a blue supergiant, Sirius and Procyon are white main sequence stars.

8.7.2 More stellar characteristics

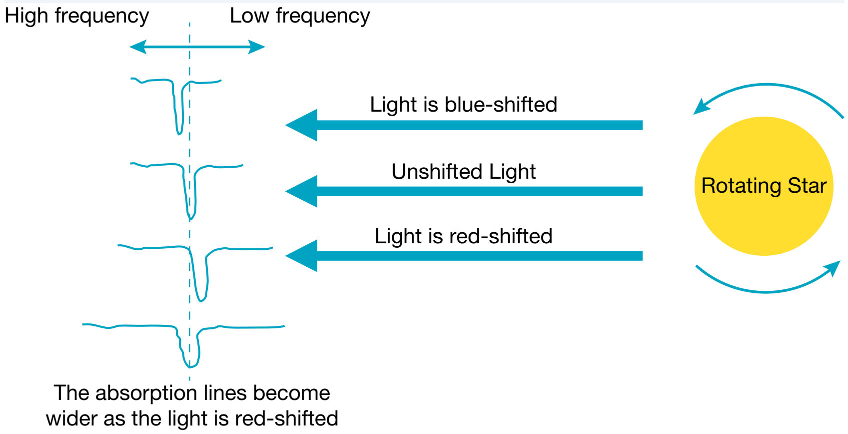
A star rotates on an axis positioned through the middle of it, similar to the Earth's axis of rotation running through the north and south poles. This rotation affects the Fraunhofer lines in the absorption spectrum from the star. The lines appear a little wider due to what is known as the Doppler effect, an effect that changes the wavelengths of waves whose source is moving towards or away from us. When a wave source is moving away from the observer, its frequency is measured to be lower by the observer. This means that the features in the light spectrum (such as absorption lines in the yellow part of the spectrum) are shifted towards the blue end if a light source like a star is moving towards us, and shifted towards the red end if the star is moving away from us. These effects on the frequency (or wavelength) of the spectra are called blue-shift and red-shift respectively. The Doppler effect is described in more detail in Topic 12 as it has been one of most powerful tools that we have for understanding the universe. Using the Doppler effect, astronomers can detect the rotations of stars that appear in the largest telescopes as mere points of light.

The absorption lines of stars are also affected by the density of the gases in the star's outer layers. We will learn in Topic 12 that stars expand greatly towards the end of their life cycle due to the heating of the outer layers of gas. This greatly reduces the density of this part of the star. As a result, there are fewer collisions between atoms than in more dense stars. For stars with denser outer layers, there is a broadening of the absorption lines. Astronomers need to distinguish whether the broadening is due to the star's rotation or due to the density of the star's outer layers.

As mentioned earlier, in the discovery of helium, the spectra from stars reveals the chemical composition of the star. Through spectroscopy, astronomers can determine the proportions of elements in different stars and relate that information to the material that the star must have been made from.

The electromagnetic spectrum predicted by Maxwell a little over 150 years ago has been one of the most important discoveries in the history of humanity. It has provided us with communication and entertainment through radio, television, wi-fi, bluetooth and remote controls. It has enabled us to explore the structure of the atom in a way that led to the quantum mechanics on which electronics, chemistry and our understanding of matter is based. It was the basis for Einstein's Theory of Special Relativity and our understanding of the structure and nature of the universe.

FIGURE 8.24 The rotation of a star is revealed by the widening of the spectral lines.



As well as determining the rotation of a star, the Doppler effect provides a powerful tool for determining the radial velocity of the star. The radial velocity is the component of the star's velocity along the line between the observer on Earth and the star. The actual translational velocity of the star is the vector sum of the radial velocity and the tangential velocity, the component of velocity at right angles to the radial velocity.

FIGURE 8.25 The translational motion of a star.

8.8 Review

8.8.1 Summary

- In the 1860s, James Clerk Maxwell produced his theory of electromagnetism, which encapsulated what was known about electricity and magnetism to that time.
- Maxwell's theory included the electric force between charges, the forces acting between electric currents and Faraday's law of electromagnetic induction.
- Maxwell's theory unified the phenomena of electricity and magnetism — they are both aspects of electromagnetism.
- Maxwell's theory predicted the existence of electromagnetic waves whose speeds are determined by $v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$ where ε_0 is the permittivity of free space, $8.854 \times 10^{-12}\ \mathrm{C^2 s^2 kg^{-1} m^{-3}}$, and μ_0 the permeability of free space is $4\pi \times 10^{-7}\mathrm{m\ kg\ C^{-2}}$.
- This determines the speed of electromagnetic waves to be $2.998 \times 10^8\ \mathrm{m\,s^{-1}}$, which is the same as the speed of light. Maxwell suggested that light is an electromagnetic wave and that there are similar waves of different frequencies that we cannot see.
- Electromagnetic waves are produced by accelerating charges. For example, electrons oscillating at high frequency in a radio antenna produce radio waves. These waves are of the same nature as light but have lower frequency.
- Over time, electromagnetic waves over the entire spectrum, such as X-rays, ultraviolet waves, gamma rays, infra-red waves and radio waves, were discovered as predicted.
- Electromagnetic waves are not disturbances in a material medium; they are not mechanical waves like sound or water waves. Instead, they propagate because a changing magnetic field produces a changing electric field at right angles to it, which produces a changing magnetic field. These fields radiate through space at the speed of light.
- Early attempts to measure the speed of light were made by Galileo and Römer. More accurate measurements were made by scientists such as Fizeau and Foucault in the eighteenth century, using rotating discs or mirrors to accurately time how long it took for light to travel a known distance.
- Spectroscopes separate light into its range of colours. Light from discharge tubes reveals emission spectra; reflected sunlight reveals a continuous spectra with stronger intensities relating to the colour of the reflected light; and incandescent filaments produce continuous spectra where the peak of intensity is in the infra-red (frequencies lower than visible light).
- Spectroscopes can reveal emission or absorption lines in spectra. Each element has a unique combination of emission and absorption lines. This enables the elements present in the gas producing the light, or that the light has passed through, to be determined.
- The surface temperatures of stars can be determined by measuring the frequency of the peak in the intensity spectrum of the light from that star. The higher the frequency, the hotter the surface of the star.
- The rotation of the star can be determined by the width of the absorption lines in the spectrum.

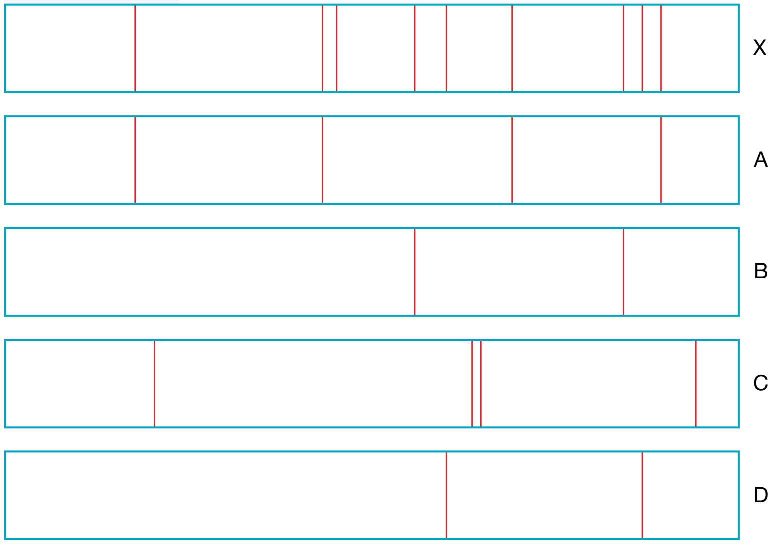
- The radial component of the translational velocity of a star can be determined by the blue-shift or red-shift of the star. The tangential component needs to be determined in other ways.
- The density of a star affects the width of the spectral lines. This is important in determining whether a star is on the main sequence or at another stage of its life cycle.
- The chemical composition of stars is revealed by the lines in the stellar spectra. Each element produces a unique series of lines in the spectrum.

8.8.2 Questions

1. Who was the key nineteenth-century scientist who produced a unified theory of electromagnetism?
2. Provide examples that demonstrate that electricity and magnetism are not separate phenomena.
3. What is a major prediction of the theory of electromagnetism?
4. How was the prediction of electromagnetic waves a good test for the theory of electromagnetism?
5. When was the theory of electromagnetism developed?
6. What formula did Maxwell determine from his equations for the speed of electromagnetic waves?
7. Did the value for the speed of electromagnetic waves calculated by Maxwell match the value measured by Fizeau or Foucault more closely?
8. Foucault's accurate measurement of the speed of light was published a few years before Maxwell's equations, which enabled the prediction of the speed of light. How might the measurement of the speed of light have influenced Maxwell's work?
9. How are electromagnetic waves produced?
10. Use your understanding of electromagnetism to describe the difference between what is happening in a radio transmitter's antenna and a radio receiver's antenna.
11. What is the difference between an electromagnetic wave and a mechanical wave such as sound?
12. A disturbance in the medium propagates a sound wave. How does an electromagnetic wave propagate through space?
13. Given the concept of a wave, why can an electromagnetic wave be called a wave even though there is no medium required for it to propagate?
14. Not only electromagnetic radiation reaches Earth from the Sun. Particles such as electrons and protons also are emitted by the Sun and reach Earth. What is the difference between particles and waves?
15. What two measurements are required for a direct measurement of the speed of light?
16. (a) Light takes 8 minutes and 20 seconds to reach Earth from the Sun. How far away is the Sun?
    (b) Light takes 1.3 seconds to reach Earth from the Moon. How far away is the Moon?
    (c) Light takes about 5.75 hours to reach Earth from Pluto. How far away is Pluto?
    (d) If your friend travelled to Pluto, how long would it take to receive a message from her by radio?
17. Can the current value of 299 792 458 m s^{-1} for the speed of electromagnetic waves be improved on by more accurate measurements? Explain.
18. (a) How is the metre currently defined?
    (b) What distance is 1 light second?
19. (a) What is the difference between the source of light in an incandescent light and a discharge tube?
    (b) Explain why discharge tubes need to be coated in a fluorescing surface to be useful as house lighting.
20. (a) Describe the electromagnetic spectrum of objects that appear green when sunlight reflects from them.
    (b) Use the electromagnetic spectrum to explain the appearance of objects having different shades of green.
21. Name two other parts of the electromagnetic spectrum that we receive from the Sun and the effects these have on humans.
22. How can the emission or absorption spectra be used to identify elements present in a gas?

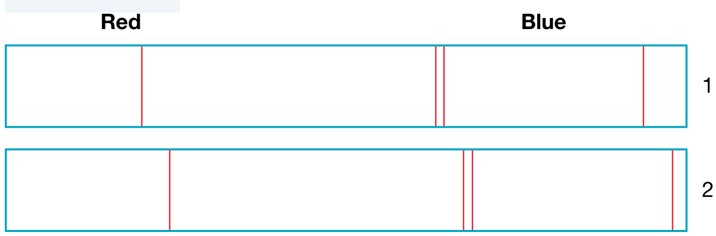
23. (a) Identify the elements (A, B, C, D) present in the gas that the light in the spectrum shown as X has passed through. The lines shown are absorption lines.

FIGURE 8.26



(b) Sketch the spectrum of an element that is in X but is not listed in A, B, C or D.
24. What instrument is used for creating a spectrum for the analysis of light?
25. (a) What are two components that can be used in the instrument mentioned in question 24 to disperse the light into its spectrum?
(b) For each, name the wave-like phenomenon that is employed to disperse the light.
26. What happens to the intensity peak in the spectrum from an incandescent source as the source gets hotter?
27. Name three common incandescent light sources.
28. How can the surface temperature of stars be determined?
29. Sketch the intensity spectrum of the Sun and add to your sketch the intensity spectrum of a hotter star.
30. List three ways that electromagnetic waves have helped us to learn about distant stars.
31. How does the Doppler effect help astronomers to determine the translational velocity of stars?
32. The following diagrams show 1) a spectrum of gases in the laboratory and 2) the spectrum from a star. What can we determine about the star from these spectra?

FIGURE 8.27



33. How was the element helium discovered?
34. How could an astronomer distinguish between a spectrum of a new element in a star and a star's spectrum that had been red-shifted or blue-shifted due to radial motion relative to Earth?

35. Astronomers claim that all the stars in the universe are made from the same elements of matter as those that make up our solar system. Explain how they can make this claim.
36. List as many uses of electromagnetic waves in your house that you can and identify which parts of the electromagnetic spectrum they use.
37. Write an argument for the discovery of electromagnetic radiation being one of the greatest scientific discoveries ever.
38. Read some reports of purported health risks due to electromagnetic waves from mobile phones. State your own opinion on these risks based on your readings and support it with the information in the readings and your understanding of electromagnetic waves.
39. Investigate the radio wave part of the electromagnetic spectrum. How is each part used and why are different frequency bands more suited to different purposes?
40. Investigate why your AM radio does not work well in a concrete building.
41. Research what scientists have learned about the universe from telescopes that detect non-visible parts of the electromagnetic spectrum.

eBook*plus* RESOURCES

Complete this digital doc: Investigation: The colour of stars
Searchlight ID: doc-17421

PRACTICAL INVESTIGATIONS

Investigation 8.1: The colour of stars

Aim
To model the colour of stars by connecting an incandescent globe to different voltages.

Apparatus
- 12 V globe
- variable 12 V power supply
- hand spectroscope.

Method
1. Darken the room and set the power supply on the 12 V setting.
2. Use the spectroscope to observe the light produced; record the most prominent colours featured in the spectrum.
3. Repeat the process with successively lower voltages until the light is no longer visible.
4. Record your results in a table like the one shown here.

Power supply setting	Red	Orange	Yellow	Green	Blue	Indigo	Violet
1							
2							

Analysis
1. What happens to the temperature of the filament as you decrease the power setting?
2. Describe what happens to the colours of the light produced by the globe as the filament's temperature is decreased.
3. Relate your observations to the colour of stars.

TOPIC 9
The wave model of light

9.1 Overview

9.1.1 Module 7: The nature of light

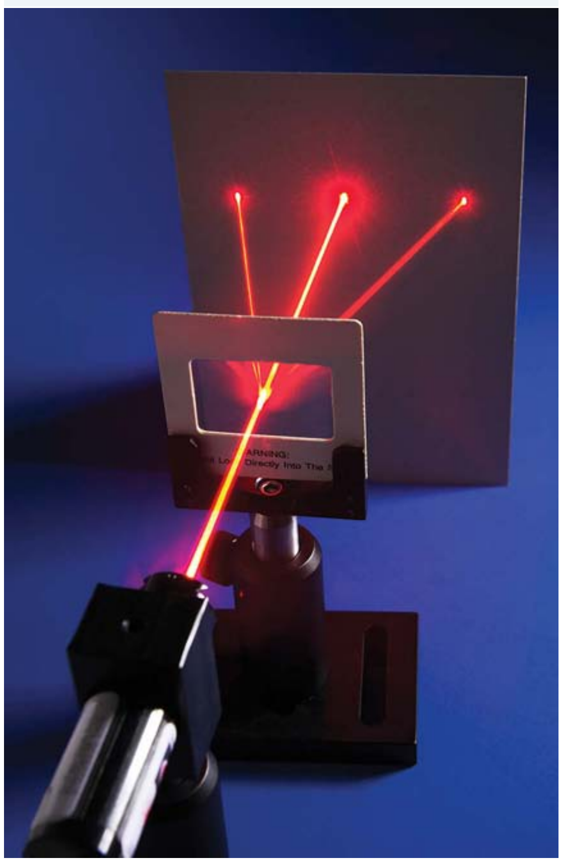
The wave model of light

Inquiry question: What evidence supports the classical wave model of light and what predictions can be made using this model?

Students:

- conduct investigations to analyse qualitatively the diffraction of light (ACSPH048, ACSPH076)
- conduct investigations to analyse quantitatively the interference of light using double slit apparatus and diffraction gratings $d \sin\theta = m\lambda$ (ACSPH116, ACSPH117, ACSPH140)
- analyse the experimental evidence that supported the models of light that were proposed by Newton and Huygens (ACSPH050, ACSPH118, ACSPH123)
- conduct investigations quantitatively using the relationship of Malus's Law, $I = I_{max} \cos^2\theta$, for plane polarisation of light, to evaluate the significance of polarisation in developing a model for light (ACSPH050, ACSPH076, ACSPH120).

FIGURE 9.1 Laser light being fired through a diffraction grating.



9.2 Competing theories of light

9.2.1 Newton and Huygens

Physicists were once convinced that light was made up of a shower of tiny particles flying together through space. What was it that changed their minds and started the world talking about light waves instead?

In the seventeenth century, there were two competing theories of light. The first theory was put forward by the esteemed Isaac Newton, and it stated that light consisted of small particles. Newton called them 'corpuscles' and his theory meant that, on the smallest scale, a light ray consisted of a shower of tiny particles. Newton was well respected, and, in his book *Opticks*, he used his theory to explain behaviours of light. Consequently, his theory was quite convincing. The alternate theory of light was put forward by Dutch physicist Christiaan Huygens. He argued that light consisted of waves, which was able to explain diffraction and interference — something that Newton's theory could not do. Unfortunately, the necessary experimental evidence for Huygens's theory did not come until well after his death, and so Newton's argument prevailed at the time. However, that situation changed over the following century due to the work of Young and Fresnel.

FIGURE 9.2 Huygens proposed that light travelled outwards from a source like circular ripples on a pond caused by dropping small stones into the water.

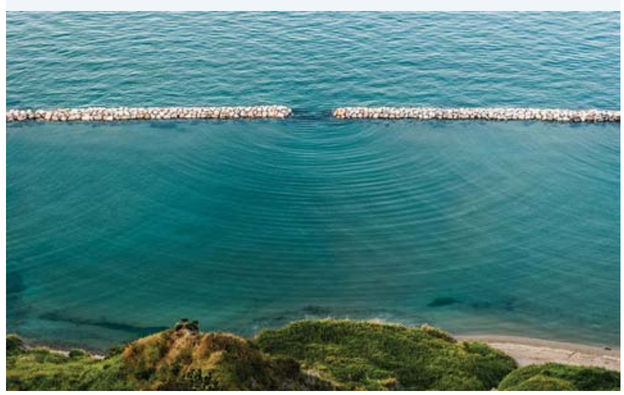
This topic focuses on the wave model of light. Diffraction and interference, as well as polarisation, remain phenomena that can only be satisfactorily explained and analysed using wave theory. Diffraction and interference are strong evidence in favour of this model.

9.3 Diffraction and interference

9.3.1 Diffraction

Diffraction and interference are closely linked, and usually occur together. Diffraction of a waveform is the spreading out of that wave when it passes a barrier or passes through a small opening. It can be difficult to observe light diffraction, because the size of the opening or barrier must be similar to the wavelength of the wave. Diffraction of water waves is much easier to observe, as shown in Figure 9.3. The reason is that water waves have much longer waves than light.

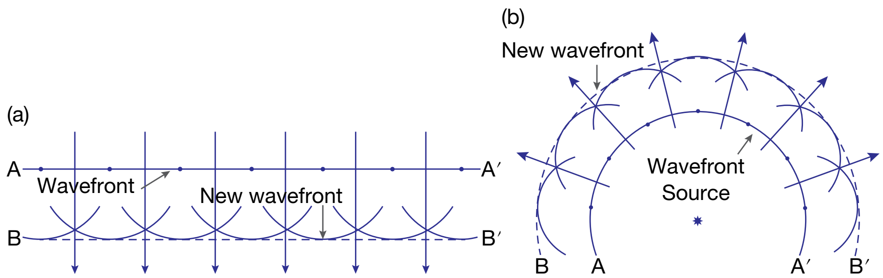
FIGURE 9.3 This aerial photograph shows water waves passing through a gap in a breakwater. The waves that have already passed through can be seen to be fanning out in semicircular shapes. This is diffraction.



9.3.2 Huygens's Principle

Christiaan Huygens proposed what has become known as Huygens's Principle, which states 'Every point on a **wavefront** may be considered to act as a source of circular secondary wavelets that travel in the direction of the wave. The new wavefront will be tangential to the wavelets.' Huygens's Principle is shown in Figure 9.4.

FIGURE 9.4 (a) Huygens's Principle explains the propagation of a plane wave. Each point acts as a point source and a new plane wave is formed. AA' is the original wavefront and the new wavefront is BB'. (b) Each point on the curved wavefront acts as a point source, and a new curved wavefront is formed.



This principle can be used to derive the laws of reflection and refraction. It also helps to explain diffraction.

If we use Huygens's Principle to explain the propagation of a wave (the formation of a new wavefront), it is necessary that all the point sources contribute to the production of the new wavefront. When a barrier blocks part of the wave, not all the point sources will be able to contribute to the new wavefront and the new wavefront on the edge is able to form a circular shape that fans out.

9.3 Exercise 1

1. Draw a barrier with a gap in it. Next draw a wavefront approaching the barrier and then add successive wavefronts that show how they progress through and diffract out of the gap.
2. Draw a series of wavefronts that show how waves can diffract around an obstacle.

9.3.3 Constructive and destructive interference

Diffraction rarely occurs without some interference. Look at Figure 9.2. The waves caused by several dropped stones are intersecting one another. They *interfere* with one another. When two waves try to act upon the same point in a medium, they combine. They add up so that two crests intersecting will produce an extra high crest and two troughs intersecting produce an extra low trough. This is called **constructive interference**. If a crest from one wave interferes with the trough of another identical wave, then the result will be no wave displacement at all. This is called **destructive interference**. (See Figure 9.5 for an explanation of this effect.)

Interference is a wave effect. Two identical water waves interfering will produce patches of extra high waves and patches of no waves at all. Two identical sound waves interfering will produce areas of extra loud sound and areas of silence. If interference occurred with light then it should produce bands of light and dark, and this would prove that light consists of waves. Huygens did observe these bands faintly around the edges of shadows, and he correctly explained them in terms of light diffracting around the edges of obstacles and then interfering to form patterns. See Figure 9.6 for examples of this optical effect. However, Huygens was not able to analyse the interference patterns quantitatively, so his explanations were not sufficiently strong. The definitive quantitative evidence of the interference of light was provided by Thomas Young almost ninety years later. He achieved this by designing an experiment that used just two point sources of monochromatic (single-wavelength) light.

FIGURE 9.5 Interference of two identical waves. (a) Where the crest of one wave meets the crest of another wave, constructive interference occurs and the resultant displacement will be twice that of one wave. (b) Where the crest from one wave meets the trough of another wave, destructive interference occurs and there will be no displacement.

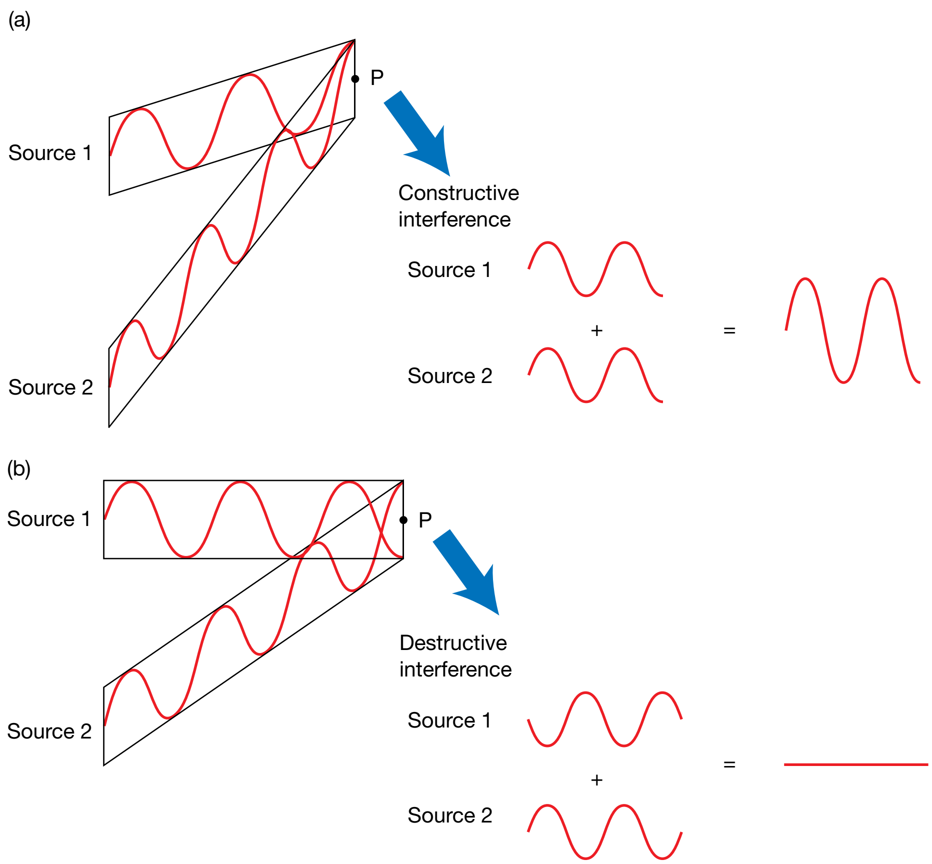
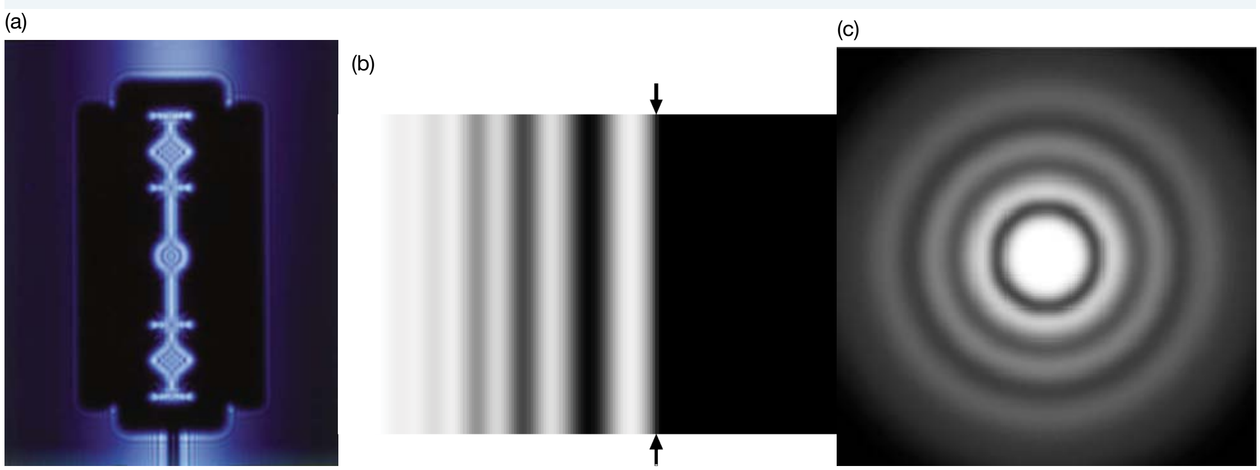


FIGURE 9.6 (a) Diffraction of monochromatic light by a straight edge, a razor blade. (b) An enlargement of the shadow shows bright and dark lines. The arrows indicate the edge of the geometric shadow. A small amount of light passes behind the straight edge but this is too faint to be observed. A series of bright and dark lines are observed next to the edge. (c) Diffraction of light by a small circular opening. This time a series of bright and dark circles is observed.



9.3.4 Young's experiment

Thomas Young (1773–1829) was keenly interested in many things. He has been called 'the last man who knew everything'. He was a practising surgeon as well as a very active scientist. He analysed the dynamics of blood flow, explained the accommodation mechanism for the human eye and proposed the three-receptor model for colour vision. He also made significant contributions to the study of elasticity and surface tension. His other interests included deciphering ancient Egyptian hieroglyphics, comparing the grammar and vocabulary of over 400 languages, and developing tunings for the twelve notes of the musical octave. Despite these many interests, the wave explanation of the nature of light was of continuing interest to him.

Young had already built a ripple tank to show that the water waves from two point sources with synchronised vibrations show evidence of interference. He was keen to see if he could observe interference with two beams of light. He held a fine hair close to his eye while staring past it at a distant candle. The light from the candle flame passed on both sides of the hair to reach his eyes. He did not notice a scattering of light in all directions as predicted by the particle model. Instead, a beautifully coloured pattern of bands parallel to the hair spread out across his view of the candle. Young's interpretation of what he saw was that light behaved like waves as it spread out from the candle.

It occurred to me that their cause must be sought in the interference of two portions of light, one reflected from the fibre, the other bending round its opposite side, and at last coinciding nearly in the direction of the former portion.

Young described this and other experiments in lectures at the Royal Institution in London in 1801 and 1802. He did not convince his audience! His listeners were reluctant to remove their confidence from the particle model that Newton apparently supported. Young was determined to produce quantitative evidence of the phenomenon that he had observed. He analysed the published results of similar experiments performed by Newton and made further measurements of his own.

In one of his experiments Young made a small hole in a window blind. He placed a converging lens behind the hole so that the cone of sunlight became a parallel beam of light. He then allowed light from the small hole to pass through two pinholes that he had punctured close together in a card. On a screen about two metres away from the pinholes he again noticed coloured bands of light where the light from the two pinholes overlapped. Figures 9.7 and 9.8 show Young's experimental arrangement.

FIGURE 9.7 Young's experiment.

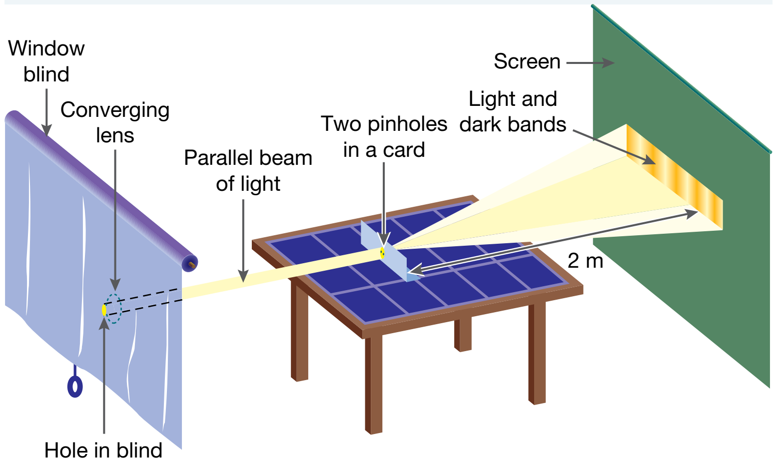
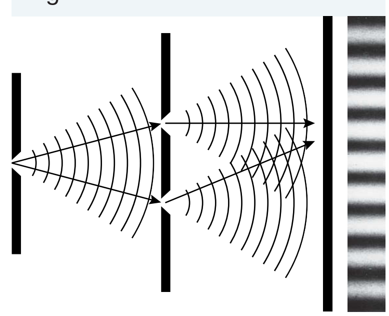
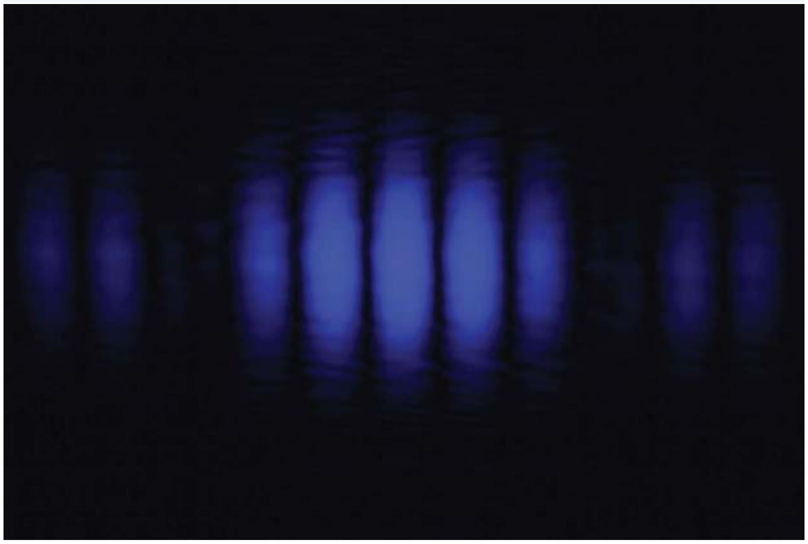


FIGURE 9.8 Young's pinhole arrangement to obtain two coherent point sources of light.



Young deliberately had just one source, the hole in the blind, because he wanted the one wavefront to arrive at the two pinholes, so that light coming through one pinhole would be synchronised with the light coming though the other pinhole. Today we would describe light coming from the two pinholes as **coherent**, which means that they are monochromatic and in phase. If Young had used two separate sources of light, one for each pinhole, their light would have been incoherent, with a random relationship between the light coming from the two pinholes and no discernible pattern on the screen.

FIGURE 9.9 A light pattern produced by a modern performance of Young's experiment.



eBook*plus* RESOURCES

Watch this eLesson: Young's experiment (interference effects with white light)
Searchlight ID: eles-0027

Try out this interactivity: Young's experiment (interference effects with white light)
Searchlight ID: int-0051

9.3.5 Interpreting Young's experiment

Young used the wave model for light to analyse his observations. Each hole in the window blind is a source of spherical waves. When these waves pass through the pinholes, each pinhole becomes a source of spherical waves. Waves from the two pinholes overlap on the screen, and their effects add together to produce the pattern. In reaching a particular point on the screen, waves have travelled from the source along two alternative routes, through one pinhole or the other. The difference between the lengths of the two paths is called the **path difference**. If the path difference results in the crests of the wave from one pinhole always meeting the troughs of the wave from the other pinhole (that is, exactly out of phase) then destructive interference occurs and that place on the screen is a dark band. Destructive interference occurs when the path difference is a whole number, minus one half, multiplied by the wavelength of the light: $(n - 0.5)\lambda$ where $n = 1, 2, ...$ is the number of bright bands from the central bright band. A bright band occurs when, in spite of a path difference, the waves are in phase: crests reinforcing crests and troughs reinforcing troughs.

This constructive interference occurs when the path difference is a whole number multiple of the wavelength of the light, $n\lambda$, again where $n = 1, 2, ...$ is the number of bright bands from the central bright band.

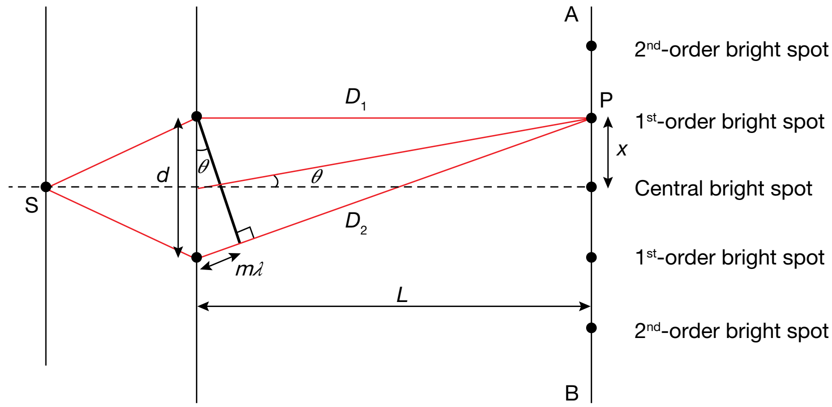
Refer to Figure 9.10 for the geometry of Young's experiment. This figure shows two rays D1 and D2 meeting at point P and interfering constructively to produce a maximum, that is, a bright spot. The extra distance travelled by ray D2 must be equal to $m\lambda$. It can also be seen that this path difference is equal to $d \sin\theta$ by trigonometry. Hence, for constructive interference or maximum, $d \sin\theta = m\lambda$ where $m = 1, 2, 3 \ldots$

For small angles, $\sin\theta \approx \tan\theta$, so the following formula also works well in practice:

$$\sin\theta = \frac{m\lambda}{d} = \frac{x}{L}$$

For destructive interference, or minimum, $d \sin\theta = \left(m - \frac{1}{2}\right)\lambda$ *where* $m = 1, 2, 3 \ \ldots$

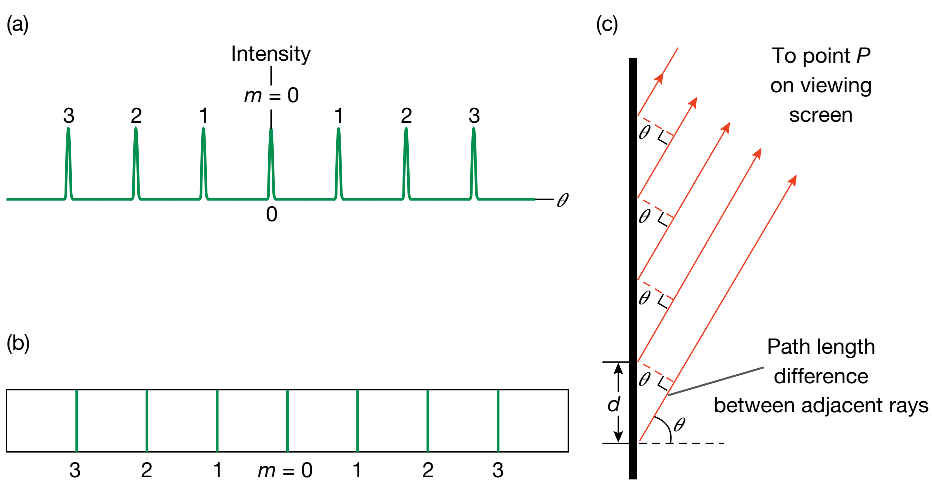
FIGURE 9.10 The geometry of Young's experiment.



9.3.6 The diffraction grating

A diffraction grating for visible light is a device for producing interference effects such as spectra. A grating consists of a large number of equidistant parallel lines engraved on a glass or metal surface. The distance between the lines is of the same order as the wavelength of the light. Note that the larger the number of slits on a grating, the sharper the image obtained. This is why a diffraction grating produces a well-separated pattern of narrow peaks (see Figure 9.11).

FIGURE 9.11 A diffraction grating is a set of accurately ruled lines on a glass that produces a sharp pattern of maxima and minima.



A transmission diffraction grating is a transparent material that has many fine lines ruled across it. (A grating may have many thousands of lines per centimetre.) These lines can be considered to be breaking the wavefront into point sources. The interference of light from these many point sources produces a diffraction pattern.

Figure 9.1, at the start of this chapter, shows a laser being directed into a diffraction grating. The rays emerging from the grating are actually very distinct maxima in the interference pattern.

9.3.7 Diffraction and interference patterns with white light

What would happen if you did not use monochromatic light but used white light instead? Red light has a longer wavelength than blue light, which leads to larger diffracting angles than blue. For example, if using a gap separation of 10 μm the first maximum for each colour can be calculated:

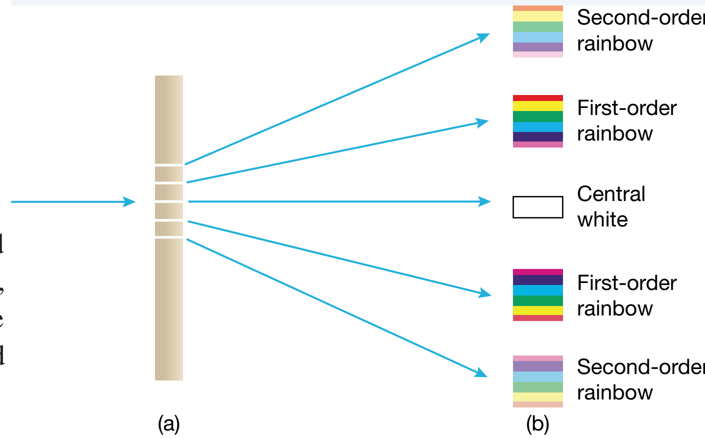
$$d \sin\theta = m\lambda \text{ so } \theta = \sin^{-1}\left(\frac{m\lambda}{d}\right)$$

$$\theta_{red} = \sin^{-1}\left(\frac{1 \times 700 \times 10^{-9}}{10 \times 10^{-6}}\right) = 4.0°$$

$$\theta_{blue} = \sin^{-1}\left(\frac{1 \times 450 \times 10^{-9}}{10 \times 10^{-6}}\right) = 2.6°$$

The result is that each maximum is dispersed into a small spectrum of coloured maxima, beginning with blue and running out to red. The diffraction grating does this very effectively and is often used in spectrometers just for its dispersing capability.

FIGURE 9.12 White light results in dispersion of the maxima as each colour has its own wavelength and dispersion angle.



9.3.8 Diffraction at a single slit

With better technology, interference patterns were found in some very interesting places. It turned out that interference patterns similar to those in Young's experiment are even produced by a single slit or pinhole, as each point across the width of a gap acts as a point source of light that will interfere with all the others. This also happens at the edge of a shadow. Looking back to Figure 9.6 these interference patterns can be identified.

9.3 SAMPLE PROBLEM 1

A sodium lamp produces almost-monochromatic light of wavelength 589 nm. If this light falls on parallel, vertical slits that are 0.100 mm apart, the light will produce an interference pattern on a screen 1.50 m away. The pattern shows a series of bright and dark bands.

(a) Determine the angle of the first-order maximum, and the distance on the screen of that maximum from the central maximum.

(b) Determine the distance from the central maximum to the third-order maximum.

SOLUTION:

(a) $d \sin\theta = m\lambda,\ so\ \theta = \sin^{-1}\left(\frac{m\lambda}{d}\right)$

$$\theta = \sin^{-1}\left(\frac{1 \times 589 \times 10^{-9}}{1.0 \times 10^{-4}}\right) = 0.337°$$

The distance to the first-order maximum can be found using

$$x = L\sin\theta = 1.5 \times \sin 0.337°$$

$$= 0.0088\ \text{m} = 8.8\ \text{mm}$$

(b) The maxima are equally spaced, so the distance to the third maximum is three times the distance to the first.

$\therefore\ distance\ \ to\ \ third\ \ maximum = 3 \times 8.8\ \ \text{mm} = 26.4\ \ \text{mm}$

9.3 Exercise 2

1 Light of wavelength 625 nm is incident on a pair of vertical slits. The distance between the centre of the two slits is 46.0 μm. The screen is placed 0.920 m from the slits. What is the distance from the central maximum to the first-order bright spot on the screen?

2 A diffraction grating is made with 3000 lines per cm. A green laser of wavelength 530 nm is directed at the grating.
   (a) What is the distance between each diffraction grating line? Since the gaps between the lines function as slits, this value is the slit separation, *d*.
   (b) What is the angular deviation of the first-order maximum?
   (c) The pattern on a screen 1.25 m away will appear as a row of dots. Calculate the separation, *x*, of the dots.

9.4 Comparing the experimental evidence for the two models of light

PHYSICS IN FOCUS

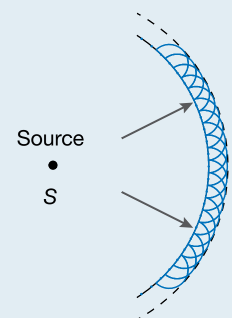
HOW DO THE TWO MODELS EXPLAIN THE PROPERTIES OF LIGHT?

How light travels

Newton's particle model: Once ejected from a light source the particles continue in a straight line until they hit a surface.

Huygens's wave model: Huygens proposed a basic principle: 'Every point in the wavefront is a source of a small wavelet. The new wavefront is the envelope of all the wavelets.'

FIGURE 9.13 Every point in the wavefront is a source of a small wavelet. The new wave-front is the envelope of all the wavelets.



Reflection of light

Newton's particle model: As particles approach a surface they are repelled by a force at the surface that slows down and reverses the normal component of the particle's velocity, but does not change its tangential component. The particle is then reflected from the surface at an angle equal to its angle of approach. The same process happens when a billiard ball hits the cushion.

Huygens's wave model: As each part of the wavefront arrives at the surface, it produces a reflected wavelet. The new wavelets overlap to produce the next wavefront, which is travelling away from the surface at an angle equal to its angle of approach.

FIGURE 9.14 Newton's particle model of reflection.

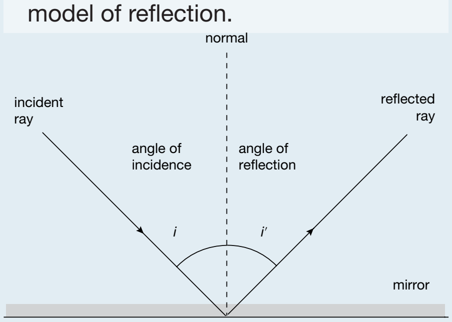
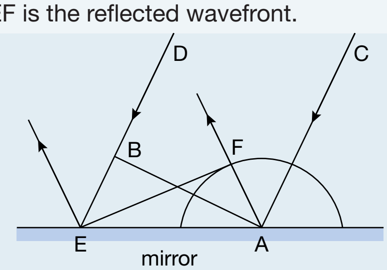


FIGURE 9.15 The wave model of reflection. C and D are parallel, incoming rays. AB is the wavefront. When A hits the mirror a circular wavelet is produced. By the time B has reached the mirror at E, the reflected wavelet has travelled out to F. The line EF is the reflected wavefront.



Refraction of light

Newton's particle model: In approaching a denser medium, the particles experience an attractive force that increases the normal component of the particle's velocity but does not affect the tangential component. This has the effect of changing the direction of the particles, bending them towards the normal where they are now travelling faster in the denser medium. Snell's Law can be explained by this model.

Huygens's wave model: When the wavefront meets a heavier medium the wavelets do not travel as fast as before. This causes the wavefront to change direction. In this case the wavefront bends towards the normal when it enters a medium where the wave is slowed down. Snell's Law can be explained by this model.

FIGURE 9.16 The particle model of refraction. The particles are pulled towards the denser medium, resulting in a change in direction.

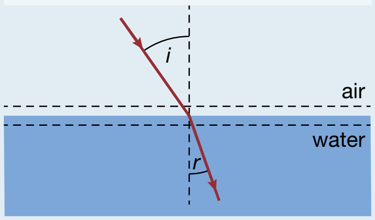
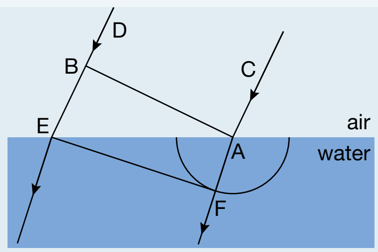


FIGURE 9.17 The wave model of refraction. C and D are parallel, incoming rays. AB is the wavefront. When A hits the surface a circular wavelet of slower speed and so smaller radius is produced. By the time B has reached the surface at E, the refracted wavelet has only gone as far as F. The line EF is the refracted wavefront, heading in a direction bent towards the normal compared to the incoming wavefront, AB.



A point of difference

Now, with these two explanations of refraction, there is a clear distinction between the two models. When light bends towards the normal as it enters water (a denser medium), the particle model says it is because light travels faster in water (the denser medium), whereas the wave model says it is because the light is travelling slower.
In the seventeenth century they did not have the technology to measure the speed of light in water. However, the particle model became the accepted explanation, partly because of Newton's status, and partly because Huygens's Principle suggested that light should bend around corners like sound, and there was no evidence of this at the time. (Newton himself actually thought that the particles in his model needed to have some wave-like characteristics to explain some of his other observations of light and colour.)

New evidence emerges

In 1802, Thomas Young (1773–1829) showed that in fact light could bend around an edge. This was convincing evidence for the wave model, as the particle model had no mechanism to explain how particles could bend around a corner. However, the status of Newton was such that not all were convinced by Young's results. It was suggested that conclusive evidence would be to measure the speed of light in water and see if it was faster or slower than that in air. Jean Bernard Leon Foucault (1819–1868) and Hippolyte Fizeau (1819–1896) competed to measure the speed of light in water; in 1850, both of them showed that light was slower in water, though Foucault won by seven weeks.

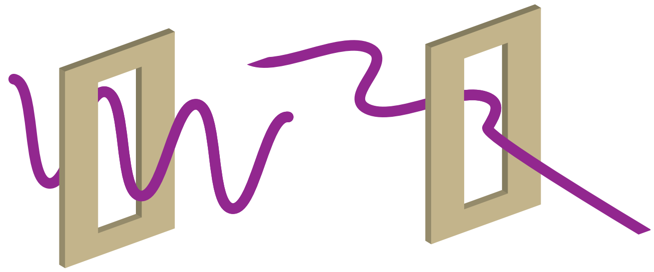
9.5 Polarisation

9.5.1 Polarisation and transverse waves

The work of Huygens, as well as Young and Fresnel, convinced the scientific world that light is a form of wave, but what type of wave? Huygens had actually proposed that light travelled as longitudinal waves, like sound waves. However, early nineteenth-century physicists experimenting with light became aware of polarisation, a phenomenon that can only occur with transverse waves. A transverse wave has a plane of oscillation — the plane along which the wave oscillates as it travels.

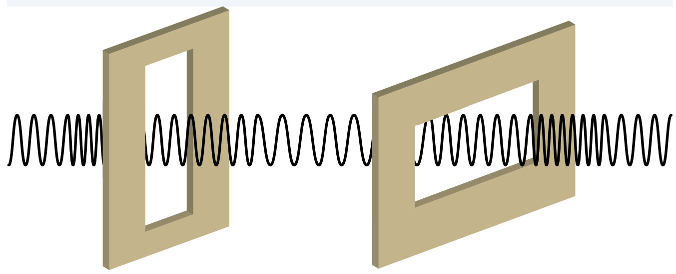
Figure 9.18 shows what happens when a transverse wave in a vertical plane passes through a vertical slit. A transverse wave in a horizontal plane is unable to pass through a vertical slit. If transverse waves in many planes were to approach the slit, only the waves in the vertical plane would pass through. This blocking of waves except for a single plane is called **polarisation**. Figure 9.19 shows how a longitudinal wave can pass through both slits. Longitudinal waves cannot be polarised. Polarisation only works with transverse waves.

FIGURE 9.18 Waves in a vertical plane pass through the slit. The waves in a horizontal plane cannot pass through.



Observations of the polarisation of light show that light is a transverse wave rather than a longitudinal wave, as longitudinal waves cannot be polarised. The polarisation of light is observed when it passes through some materials. These materials, which allow light waves in one plane to pass while blocking light in all other planes, are called polarising filters.

FIGURE 9.19 Longitudinal waves can pass through both vertical and horizontal slits.



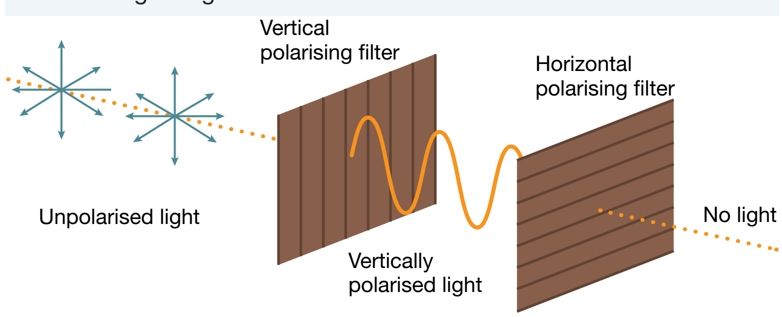
Light from a common source (the Sun or a lamp) is unpolarised light. Every plane of oscillation is represented in a ray of unpolarised light. In Figure 9.20, unpolarised light is passed through a polarising filter with a vertical axis, and only waves with vertical axes of oscillation pass through unimpeded. If a wave has an axis at an angle different to the filter's axis, only the vertical component of the wave is passed through. The result is that the intensity of the light passing through the filter is half the intensity of the original light:

$$I = \frac{1}{2} I_O$$

where I = the intensity if the emerging polarised light

I_O = the intensity of unpolarised light falling on the polariser

FIGURE 9.20 Light passed through crossed polarisers — polarising filters at right angles to each other.

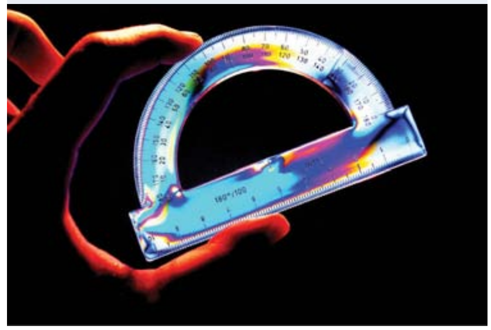


In a double filter arrangement, such as that shown in Figure 9.20, the first filter is usually called the *polariser* and the second is the *analyser*. When the polariser and the analyser are crossed, as shown, none of the polarised light will be able to pass through the analyser. If the analyser were rotated 90°, the two filters would be aligned and all the polarised light would pass through the analyser. If there is some other angle θ between the axes of the polariser and the analyser, the intensity of light emerging from the analyser is given by the following equation, discovered by Etienne Malus and known as Malus's Law:

$$I = I_O \cos^2 \theta$$

where I = the intensity of the emerging polarised light
I_O = the intensity of polarised light falling on the analyser
= the maximum light intensity available, I_{max}
θ = the angle between the axis of the incident polarised light and the axis of the analyser
= the angle between the axes of the two filters.

FIGURE 9.21 Light can be polarised by reflection off hard surfaces such as water, glass or plastics.



Polarisers can be made of certain crystals or filters constructed of stretched long-molecule plastics (sometimes called *polaroid filters*). Light can also be polarised by reflection off hard surfaces such as water or glass, and by scattering from atoms or molecules.

As a result, polarising sunglasses (which have a vertical axis) can cut though the horizontally polarised glare off a pond but not through the vertically polarised glare off a shop window. In addition, as blue sky is partially polarised due to the scattering of sunlight, polarising filters can be fitted to camera lenses to deepen this colour for effect.

FIGURE 9.22 This polarising filter for a camera deepens the blue colour of the sky and sea.

9.5 SAMPLE PROBLEM 1

(a) Unpolarised light falls upon a polariser with intensity I_O. What will be the intensity of the emerging polarised light (call this I_a) expressed as a percentage of the original light?
(b) The polarised light then falls upon an analyser with an axis at 60° to the polariser. What will be the intensity (call it I_b) of the emerging light, again expressed as a percentage of the original light?
(c) The emerging light then falls upon a second analyser with an axis 30° to the first analyser (90° to the polariser). Once more, find the intensity (I_c) of the emerging light as a percentage of I_O.
(d) The first analyser is now removed. What will be the relative intensity (I_d) of the light emerging from the remaining analyser as a percentage of I_O?

SOLUTION:

(a) $\dfrac{I_a}{I_o} = 0.5 = 50\%$

(b) $I_b = I_a \cos^2 (60°)$
$I_b = 0.25 I_a$ and $I_a = 0.5 I_o$, so $I_b = (0.25 \times 0.5)\, I_o$

$\therefore \dfrac{I_b}{I_o} = 0.125 = 12.5\%$

(c) $I_c = I_b \cos^2 (30°)$
$I_c = 0.75 I_b$ and $I_b = 0.125 I_o$, so $I_c = (0.75 \times 0.125)\, I_o$

$\therefore \dfrac{I_c}{I_o} = 0.094 = 9.4\%$

(d) $I_d = I_a \cos^2(90°)$

$I_d = 0$

$\therefore \frac{I_d}{I_o} = 0 = 0\%$

You should be able to confirm this experimentally but bear in mind that plastic polarising filters are not perfect polarisers.

9.5 Exercise 1

1 Sheryl decided that she would use two polaroid filters to observe a very bright light source. She placed the filters together and rotated them so that their axes were at 45° to each other, then looked at the light source through them.

(a) What fraction of the original intensity of the light will Sheryl observe?

(b) This light was much too bright and so Sheryl rotated the second filter until the intensity of the transmitted light was one-tenth of the unfiltered light. What was the new angle between the axes of the two filters?

(c) This light was still too bright so she continued rotating the second filter to reduce the intensity of the light to one-twentieth of the original intensity. What is the final angle between the axes of the filters?

9.6 Review

9.6.1 Summary

- In the seventeenth century, there were two competing models of light — Isaac Newton's particle model and Christiaan Huygens's wave model.
- The variety of known behaviours of light at that time could be explained by both theories.
- Experiments that followed in the eighteenth century demonstrated light behaving as a transverse wave. The phenomena investigated were diffraction, interference and the polarisation of light.
- Diffraction is the spreading out of a wave after it has passed through a gap or around an obstacle. It can be used to create multiple coherent waves from a single source.
- Interference occurs when two or more waves act on a point at the same time. It may be constructive, creating a net wave with greater amplitude, or destructive, creating a net wave with smaller amplitude or zero amplitude.
- Young's experiment used two coherent point sources of light to show that the light waves from the two sources interfered with each other to produce a pattern of bright and dark spots, referred to as maxima and minima. The formula that describes this behaviour is $d \sin \theta = m\lambda$.
- A diffraction grating is a piece of glass with many fine lines ruled upon it. There may be thousands of lines per centimetre. When a beam of light is shone into the grating, the gaps between the lines act as point sources of light and a very sharp interference pattern is produced.
- If white light is used, each maximum is dispersed into a spectrum since each wavelength of light has a different diffraction angle.
- Even a single source of light will produce an interference pattern that blurs the edges of shadows and images.

- Plane polarisation is a wave behaviour specific to transverse waves; it is the restricting of the plane of oscillation of a transverse wave to just one direction. Light from common sources is unpolarised, with a plane of oscillation represented. Passing that light through a polarising filter produces light that has just one plane present. Passing this polarised light through another polarising filter, called an analyser, will produce polarised light with intensity given by $I = I_O \cos^2 \theta$, known as Malus's law.

9.6.2 Questions

FIGURE 9.23 Examples of diffraction and interference in water waves.

1. Refer to the images in Figure 9.23. Identify the areas where you can see evidence of diffraction and also areas showing interference. Explain what is happening in the areas you have identified.
2. It is said that one of the side effects of building a breakwater next to a beach is that it turns the waves around to face the beach. Discuss this suggestion.
3. White light passed through a narrow slit and projected onto a distant screen shows bright and dark bands with coloured fringes.
   (a) Explain how the coloured fringes arise.
   (b) Red fringes are observed at the further extent from the central white maximum. Why?
4. List several different path differences that would produce constructive interference for infra-red radiation with a wavelength of 1.06 μm. Now list several path differences that would produce destructive interference.
5. A student shines a helium–neon laser, which produces light with a wavelength of 633 nm, through two slits and produces a regular pattern of light and dark patches on a screen as shown below. The centre of the pattern is the band marked A. Using a wave model for light we can describe light as having *crests* and *troughs*.

FIGURE 9.24

   (a) Use these terms to explain:
      i. the bright band labelled A in the diagram above
      ii. the dark band labelled B.
   (b) What is the difference in the distance light has travelled from the two slits to:
      i. the bright band labelled A
      ii. the dark band labelled B
      iii. the bright band labelled C?

(c) Using the same experimental setup, but replacing the laser with a green argon ion laser emitting 515 nm light, what changes would occur to the interference pattern?
(d) The helium–neon laser is set up again. The distance between the two slits is now increased. What changes to the interference pattern shown in the diagram above would occur?
(e) The screen on which the interference pattern is projected is moved further away from the slits. What changes to the interference pattern shown in the diagram would occur?

6. Light of wavelength 430 nm falls on a double slit of separation 0.500 mm. What is the distance between the central bright band and the third bright band in the pattern on a screen 1.00 m from the double slit?
7. A double slit is illuminated by light of two wavelengths, 600 nm and the other unknown. The two interference patterns overlap with the third dark band of the 600 nm pattern coinciding with the fourth bright band from the central band of the pattern for the light of unknown wavelength. What is the value of the unknown wavelength?
8. Diffraction and interference are wave properties, so the theory in this chapter also applies to other waveforms, such as sound. A stage is set up for an outdoor performance. Two large speakers are positioned 13 m apart and point out towards the audience area. If the speakers play a note of 500 Hz and are in phase with each other (and so are coherent), an interference pattern will be created in the audience space.
   (a) If a person were to walk across the audience area from left to right, what would they hear?
   (b) Assume that the speed of sound is 340 m s^{-1}. What is the wavelength of this sound?
   (c) Assume the audience area is 95 m deep. Identify three locations at the back of this area that would be particularly good for hearing this sound.
9. Two colours of light are shone onto a diffraction grating with exactly 5000 lines per cm. One colour is violet of wavelength 398 nm and the other is red of wavelength 652 nm. The resulting pattern falls on a screen 89.0 cm from the grating.
   (a) What will be the colour of the central bright spot?
   (b) What is the distance from the central maximum to the first-order bright spot of the violet light on the screen?
   (c) What is the distance from the central maximum to the first-order bright spot of the red light on the screen?
   (d) What is the separation of the second order of each colour on the screen?
10. A polariser and analyser are set up so that the intensity of the light emerging from the analyser is 20% of the original unpolarised light striking the polariser. Determine
   (a) the intensity of the light between the two filters
   (b) the angle between the polarising axes of the two filters.
11. A polaroid filter is used to produce polarised light of intensity I_o. Another polaroid filter is placed in front of that light and the intensity drops to $\frac{1}{8}I_o$. What is the angle between the axes of the two filters?
12. Three polaroid filters are stacked so that the angle between each is 45°, with 90° between the first and the last. An unpolarised beam of light is shone into the stack. What proportion of the original beam emerges from the stack?

eBook*plus* RESOURCES

Complete this digital doc: Investigation: The diffraction of light
Searchlight ID: doc-26598

Complete this digital doc: Investigation: The interference of light with diffraction gratings
Searchlight ID: doc-26599

Complete this digital doc: Investigation: The interference of light — Double slits
Searchlight ID: doc-26600

Complete this digital doc: Investigation: The plane polarisation of light
Searchlight ID: doc-26601

PRACTICAL INVESTIGATIONS

Investigation 9.1: The diffraction of light

Aim

To qualitatively observe the diffraction of light by using a single slit.

Apparatus

A laser pointer

A single slit apparatus — this may be a purpose-built mechanism that can vary the size of the slit.

Theory

Diffraction refers to the ability of the waves to spread out after passing through a gap or going past an obstacle. The gap needs to be comparable to the wavelength of the wave for this to occur. Consequently, it is relatively easy to demonstrate the diffraction of water or sound waves. Diffraction of light is more difficult but can be shown by incorporating an interference effect.

If monochromatic light, such as that from a laser, shines on a narrow single slit or gap in a barrier, then each point along the width of the slit will act as a new source of secondary wavelets of light, each of which will spread out and then interfere with one another. Under the right conditions, this will produce a pattern on a screen. In this case, the pattern will consist of a bright central band flanked by fainter, closely spaced bright bands. This pattern is evidence of the diffraction and then interference of the laser light.

Method

1. Set up the laser pointer so that it shines onto a screen at least a metre away.
2. Mount the single slit in front of the laser so that the beam shines through the slit.
3. Darken the room and then adjust the size of the slit until the diffraction pattern appears on the screen.

Questions

1. Describe the appearance of the diffraction pattern.
2. Measure the distance between the central bright band and the adjacent band, and then between the fainter bands. How do these gaps compare?
3. If time permits, you may research the equation that mathematically describes this pattern and then compare this to the equation that describes double-slit and diffraction grating interference.

Investigation 9.2: The interference of light with diffraction gratings

Aim

To observe the diffraction and interference of light using diffraction gratings.

Apparatus

A laser light source such as a laser pointer

A diffraction grating set (many of these have three different rulings on one slide)

Metre rule or tape measure

Theory

A diffraction grating has many fine lines etched onto a glass surface. When monochromatic light shines through the glass, the gaps between the lines act as sources of secondary wavelets that interfere to produce a pattern of light and dark bands, called an interference pattern. If a laser is used, the pattern will be a series of bright dots on a screen. The geometry of the pattern is described mathematically by

$$\sin\theta = \frac{m\lambda}{d} = \frac{x}{L}$$

where θ is the angle of deviation to the first-order bright spot, *m* is the order (so 1 for the first-order bright spot), λ is the wavelength of the light used, *d* is the distance between adjacent lines on the grating, *x* is the distance between the dots, and *L* is the distance from the grating to the screen. From this, an expression can be derived to determine the wavelength of the laser light:

$$\lambda = \frac{dL}{x}$$

The relationship between the number of lines per centimetre on the grating N and the distance in metres between the bright dots d is given by $d = \frac{1}{100\,N}$.

Method

1. Use supports such as retort stands to set up the laser pointer so that it shines perpendicularly onto a screen, wall or board at least one metre away.
2. Mount the diffraction grating directly in front of the laser pointer so that a regular row of dots appears on the screen.
3. Measure the values of x and L, and record these in your results table, along with the N value for your grating.
4. Repeat this procedure for each grating of different N value.
5. Analyse the data to determine the wavelength of the laser pointer.

Results

Diffraction grating	N (lines per cm)	d (m)	x (m)	L (m)	λ (m)	λ (nm)
A						
B						
C						

Questions

1. Calculate an average value for λ from the data table. Does this value fit within the accepted range of wavelengths for visible light of the colour of your laser?
2. If you have been able to use a number of gratings, use the greatest deviation from the average value to determine the accuracy of your measurement. If you only have one grating available, use the number of significant figures within your data to estimate the accuracy of your measurement.
3. Rearrange the formula used here to make x the subject. How would x change if a different colour laser was used? Extend this idea to explain what would happen to the interference pattern if a beam of white light was used instead of the monochromatic laser.
4. The formula $\sin\theta = \frac{m\lambda}{d} = \frac{x}{L}$ makes an assumption. What is that assumption and is it valid in this experiment?

Investigation 9.3: The interference of light — Double slits

Aim

To use the diffraction and interference of monochromatic light, as observed by Thomas Young, to determine the separation of two close slits.

Apparatus

A laser light source such as a laser pointer, of known wavelength
A double-slit apparatus, such as a prepared photographic slide
Metre rule or tape measure

Theory

Thomas Young first discovered that monochromatic coherent light shone on two closely separated slits would produce an interference pattern on a screen. The pattern consists of a series of light and dark bands and is caused by the slits acting as new sources of secondary wavelets, which then diffract and interfere with each other. Young needed a clever arrangement to produce the monochromatic coherent light, but we can simply use a laser pointer for the same purpose, provided that the width of the laser beam can cover both slits.
The separation of the bright bands in the interference pattern is described mathematically by

$$\sin\theta = \frac{m\lambda}{d} = \frac{x}{L}$$

where θ is the angle of deviation to a bright band, m is the order number of that band, λ is the wavelength of the light used, d is the slit separation, x is the distance between the bands on the screen, and L is the distance from the grating to the screen. From this, an expression can be derived to determine the separation of the slits:

$$d = \frac{\lambda L}{x}$$

Here we are considering the distance on the screen of the first-order band from the central band, so that $m = 1$.

Method

1. Set up the laser pointer so that it shines perpendicularly onto a screen, wall or board approximately two metres away.
2. The double-slit slide may have a selection of slits available. Choose the closest pair of slits and mount the slide in front of the laser pointer so that an interference pattern appears on the screen. You should be able to discern a distinct series of closely spaced bright spots.
3. Identify the central (brightest) spot and measure the values of x (distance to the next bright spot) and L (distance from the slits to the screen), and record these in your results table, along with the wavelength of your laser light.
4. Repeat this procedure for other slit sets that may be available.
5. Analyse your data to determine the slit separations.

Results

Wavelength of laser light $\lambda =$ __________ m
Distance from slits to screen $L =$ __________ m

Slit set	Band separation x(m)	Slit separation d(m)
A		
B		
C		

Questions

1. If several slit sets were available, describe how the pattern changes as the slit separation increases. A useful activity is to graph the slit separation d against $\frac{1}{x}$. This should be a linear relationship with a gradient equal to λL.
2. How does the pattern change if the wavelength is different? You can try this if you have a laser pointer with a different colour.

Investigation 9.4: The plane polarisation of light

Aim

To observe plane polarisation of light using polarising filters, and to verify Malus's law.

Apparatus

A light source, such as an LED torch
A light sensor
3 polarising filters
A protractor

Theory

A polarising filter produces plane polarised light from unpolarised light by restricting the electric field vector in the electromagnetic waves to one plane. This necessarily cuts down the intensity of the transmitted light to half of the original intensity. If the polarised light of intensity I_{max} is directed into another polarising filter with an axis at an angle θ to the first filter, the intensity of the transmitted light should be $I = I_{max} \cos^2 \theta$. This is Malus's law.

Method

1. Work out the axes of polarisation of your polarising filters. The easiest way to do this is to observe light from a ceiling lamp reflecting off a shiny desk surface. The light will be partially polarised horizontally. Place a polarising filter in front of your eyes so that the reflected glare is seen through the filter, and then rotate the filter. You should notice that the transmitted light is darker or lighter depending on its orientation. Find the darkest position. In this orientation, the axis of polarisation of the filter is vertical. Use a marker to draw a vertical line along the side of the filter so that you can identify the axis. Repeat this process for each of your filters.
2. Darken the room.

3. Use the light sensor to measure the light intensity produced by the light source.
4. Place one of the polarising filters in front of the light source and measure the intensity of the transmitted light. This value should be about half the value of the original value. For the following analysis, call this new value $I_{\max}$.
5. Place a second polarising filter behind the first. The second filter is called an analyser. Keep the axes of the filters parallel and measure the intensity of the transmitted light. Note this value in the results table.
6. Rotate the axis of the analyser to $30°$ from the polariser and measure the light intensity again. Repeat this process for each angle indicated in the results table.
7. Use Malus's law to calculate the theoretical value for the intensity of the transmitted light for each of the angles in the results table.
8. Finally, place the two polarising filters together with their axes $90°$ apart so that minimal light is transmitted. Slide a third polarising filter between them, with its axis at $45°$ to each of the others. What do you observe?

Results

Angle between filters	Measured intensity	Calculated intensity
$0°$		
$30°$		
$45°$		
$60°$		
$90°$		

Questions

1. The first filter polarises the light and should reduce the light intensity by 50%. Did your measurements confirm this in step 4 above?
2. Referring to the results table, how did your measured intensities compare to those calculated using Malus's law?
3. Do your results support or refute Malus's law? Justify your answer.
4. Malus's law assumes ideal polarisers; however, plastic polarising filters are less than ideal. Research the reasons for this.
5. In the final part of this experiment, you slid a third polarising filter between two other crossed filters. Explain the effect that you observed.

TOPIC 10 The quantum model of light

10.1 Overview

10.1.1 Module 7: The nature of light

Light: Quantum Model

Inquiry question: What evidence supports the particle model of light and what are the implications of this evidence for the development of the quantum model of light?

Students:

- analyse the experimental evidence gathered about black body radiation, including Wein's Law, $\lambda_{max} = \frac{b}{T}$, related to Planck's contribution to a changed model of light (ACSPH137)
- investigate the evidence from photoelectric effect investigations that demonstrated inconsistency with the wave model for light (ACSPH087, ACSPH123, ACSPH137)
- analyse the photoelectric effect ($E_{k,\,max} = hf - W$) as it occurs in metallic elements by applying the law of conservation of energy (ACSPH119)

FIGURE 10.1 Radio telescopes such as these at Narrabri in NSW are pointed into space. They collect radio waves and other electromagnetic radiation from galaxies. The SETI Project (Search for Extra-Terrestrial Intelligence) utilises radio telescopes to 'listen' for intelligent signals from other intelligent beings. The electromagnetic spectrum we know today extends from the wavelength of gamma rays, as small as 10^{-14} m, through to radio waves with wavelengths of 10^{5} m. That knowledge has come from the work of two of the giants of science, James Clerk Maxwell and Heinrich Hertz.

10.2 Light as a type of wave

10.2.1 Maxwell's theory of electromagnetic waves

The passage of light across the vast universe is apparent to all who looked towards the heavens and the stars. Explaining how that could be was another matter. According to previous wave theory, waves were propagated through a medium. What was the medium in which light travelled?

One of the problems that nineteenth-century scientists had in understanding how electromagnetic waves carry energy through the vacuum of space was that mechanical waves vibrate in a medium. Was there a medium filling space? One name given to this unproven medium was the luminiferous *aether*. The presence or absence of the aether was hotly debated. Proof of its presence or absence became one of the great goals of science. In the end, Albert Einstein simply said that its existence or absence was irrelevant. It could not be detected and made no difference to the passage of light.

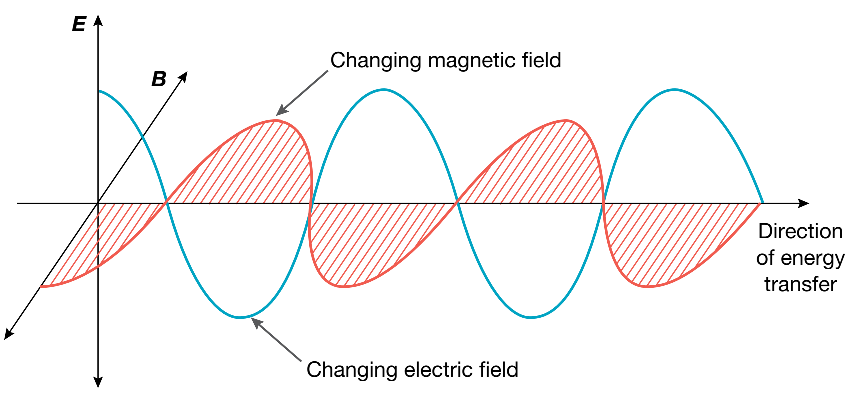
FIGURE 10.2 James Clerk Maxwell (1831–1879).

Based on observations that a changing magnetic field induces an electric field in the region around a magnet, and that a magnetic field is induced in the region around a conductor carrying an electric current, James Clerk Maxwell concluded that the mutual induction of time- and space-changing electric and magnetic fields should allow the following unending sequence of events.

- A time-varying electric field in one region produces a time- and space-varying magnetic field at all points around it.
- This varying magnetic field then similarly produces a varying electric field in its neighbourhood.
- Thus, if an electromagnetic disturbance is started at one location (for example, by vibrating charges in a hot gas or in a radio antenna) the disturbance can travel out to distant points through the mutual generation of electric and magnetic fields.
- The electric and magnetic fields propagate through space in the form of an 'electromagnetic wave' (illustrated in Figure 10.3).

The upshot of this sequence of events was clear. Light, and indeed any electromagnetic wave, does not need a medium to propagate. Electromagnetic waves are self-propagating. Once started, they have the capacity to continue forever without continuous energy input. You are probably familiar with the idea that the light we see coming from distant stars took millions, if not billions, of years to reach the Earth. It is possible that the origin of that light no longer exists, yet you can still see the light that emanated from the object.

FIGURE 10.3 A diagrammatic representation of an electromagnetic wave.

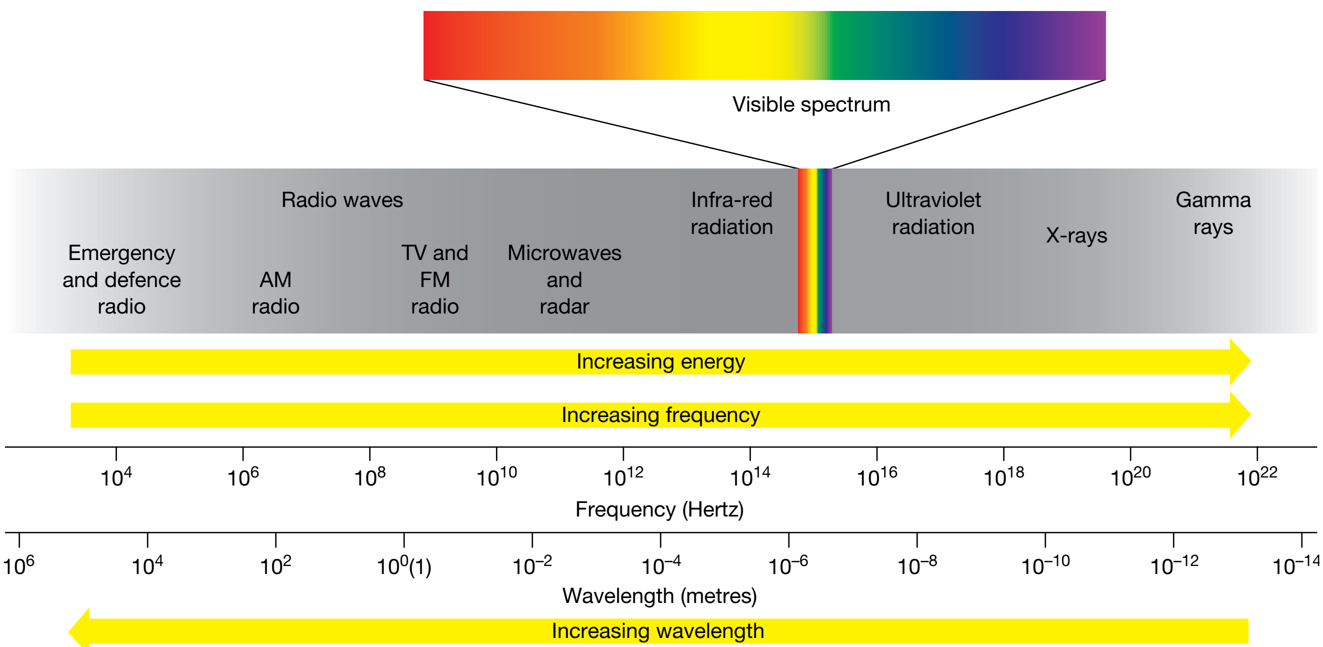


Maxwell's theory gave a definite connection between light and electricity. In a paper titled 'A dynamical theory of the electromagnetic field', which he presented before the Royal Society in 1864, Maxwell expressed four fundamental mathematical equations that have become known as 'Maxwell's equations'.

Maxwell's equations predicted that light and electromagnetic waves must be transverse waves and that the waves must all travel at the speed of light. They also implied that a full range of frequencies of electromagnetic waves should exist. In other words, the equations suggested the existence of an electromagnetic spectrum.

At the time of these predictions, only light and infra-red radiation was known and confirmed to exist. One look at the spectrum shown in Figure 10.4 allows you to see how little was known of the complete electromagnetic spectrum known to exist today. Maxwell's equations also suggested that the speed of all waves of the full electromagnetic spectrum, if they did exist, was a definite quantity that he estimated as $3.11 \times 10^8\ \mathrm{m\,s^{-1}}$. Maxwell's theoretical calculations were supported by the experimental data of French physicist Armand Hippolyte Louis Fizeau (1819–1896) who had determined a figure very close to this for the speed of light. In 1849, Fizeau's experiments to measure the speed of light had obtained a value of $3.15 \times 10^8\ \mathrm{m\,s^{-1}}$.

FIGURE 10.4 The electromagnetic spectrum.



10.2 Exercise 1

1 Use the figure of the electromagnetic spectrum to estimate the following:
   (a) the frequency of TV and FM radio
   (b) the wavelength of ultraviolet radiation.
2 A type of electromagnetic wave has a frequency of 1×10^{19} Hz. What type of electromagnetic radiation would it be best classified as?
3 An electromagnetic wave of wavelength 1×10^4 m travels at speed x in a vacuum while an electromagnetic wave of frequency 1×10^{22} Hz travels at speed y in a vacuum. Determine the ratio $\frac{x}{y}$ and explain your numerical answer.

10.2.2 Heinrich Hertz: Further evidence that light is a type of wave

Maxwell's two most important predictions were that:

- electromagnetic waves could exist with many different frequencies
- all such waves would propagate through space at the speed of light.

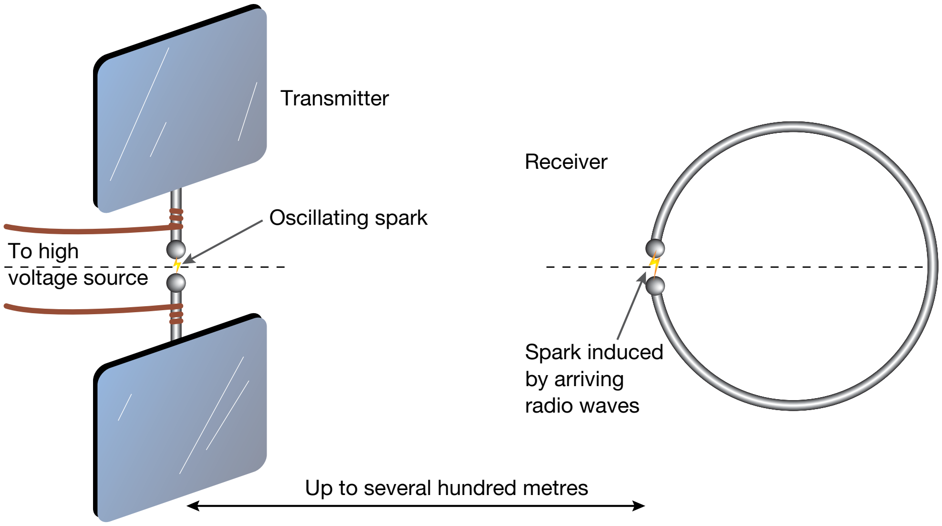
In 1886, Heinrich Hertz conducted a series of experiments that verified these predictions. Unfortunately, Maxwell had died in 1879 and did not see this experimental confirmation of the theoretical predictions of his equations.

Hertz reasoned that he might be able to produce some of the electromagnetic waves with frequencies other than that of the visible light predicted by Maxwell's equations. He thought he could produce some of these electromagnetic waves by creating a rapidly oscillating electric field with an induction coil that caused a rapid sparking across a gap in a conducting circuit.

In his experiments that confirmed Maxwell's predictions, Hertz used an induction coil to produce sparks between the spherical electrodes of the transmitter. He observed that when a small length of wire was bent into a loop so that there was a small gap and it was held near the sparking induction coil, a spark would jump across the gap in the loop. He observed that this occurred when a spark jumped across the terminals of the induction coil (see Figure 10.6). This sparking occurred even though the loop was not connected to a source of electrical current. Hertz concluded this loop was a detector of the electromagnetic waves generated by the transmitter. This provided the first experimental evidence of the existence of electromagnetic waves.

FIGURE 10.5 Heinrich Hertz (1857–1894).

FIGURE 10.6 Hertz, using an induction coil and a spark gap, succeeded in generating and detecting electromagnetic waves. He measured the speed of these waves, observed their interference, reflection, refraction and polarisation. In this way, he demonstrated that they all have the properties characteristic of light.



Hertz then showed that these new electromagnetic waves could be reflected from a metal mirror, and refracted as they passed through a prism made from pitch. This demonstrated that the waves behaved similarly to light waves in that they could be reflected and refracted.

Additionally, Hertz was able to show that, like light, the new electromagnetic waves could be polarised. Hertz showed that the waves originating from the electrodes connected to the induction coil behaved as if they were polarised by rotating the receiver loop. When the detector loop was perpendicular to the transmitter gap, the radio waves from the gap produced no spark (see Figure 10.7). The spark in the receiver was caused by the electric current set up in the conducting wire. When the detector loop was parallel to the spherical electrodes attached to the induction coil (see Figure 10.8), the spark in the receiver was at maximum. At intermediate angles it was proportionally less. This was a behaviour similar to that shown by polarised light waves after the light has passed through an analyser, such as a sheet of polaroid. It demonstrated that the newly generated electromagnetic waves were polarised.

FIGURE 10.7 No spark was detected when the detector loop was rotated.

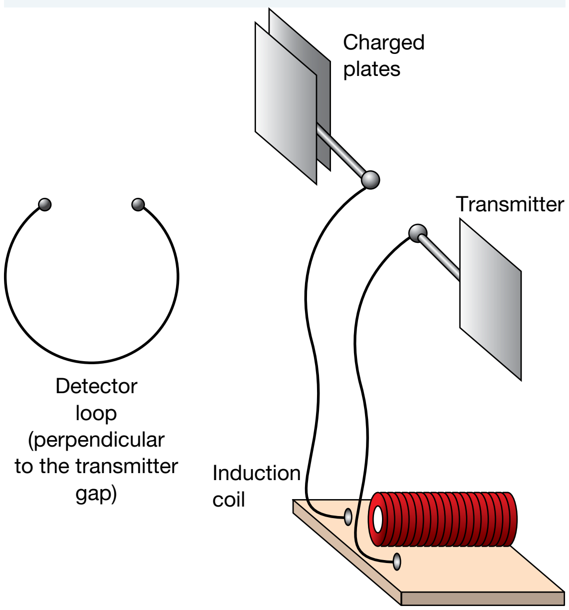
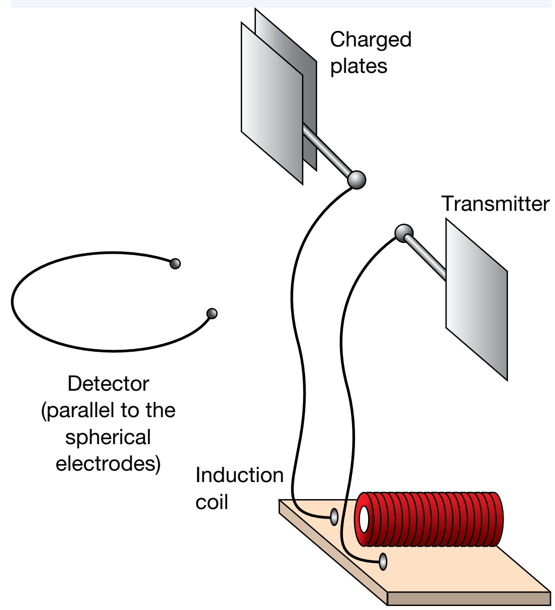


FIGURE 10.8 Hertz detected the waves when the detector loop was placed like this.



If Hertz's interpretation was correct, and electromagnetic waves did travel through space from the coil to the loop, he reasoned there must be a small delay between the appearance of the first and second spark. The spark in the detector cannot occur at exactly the same time as the spark in the induction coil because even travelling at the speed of light it takes a finite time for the wave to move from one point to another. Hertz measured this speed in 1888 by using a determined frequency from an oscillating circuit and a measured wavelength, as determined by interference effects for the waves produced, and found that it corresponded to the speed predicted by Maxwell's equations; it was the same as the speed of light.

To measure this wavelength, Hertz connected both the transmitter and the detector loop with a length of wire. He had already shown, by rotating the second loop, that the waves produced by the sparking behaved as if they were polarised. The spark in the receiver was caused by the electric current set up in the conducting wire. At intermediate angles, interference of the currents provided a measure of the wavelength of the radio waves through the air.

The speed of transmission of the sparks was measured using a technique taken from light. Lloyd's mirror uses interference of two separate beams of light. One beam travels directly from the source to a detector. The other reflects a beam from the source from a mirror set at a small angle. Both beams interfere constructively and destructively when they arrive at the detector. It is possible to use the pattern produced to determine the wavelength of the waves.

Hertz carried out a modification of this experiment, reflecting the sparks from a metal plate. He suggested that the waves produced had a wavelength larger than light, which should make measurements easier. Knowing the frequency of the sparks and their wavelength, he obtained a value for the speed of transmission. His value was similar to the speed of light measured by Fizeau.

The invention of this set of experiments and procedures was the first time that electromagnetic waves of a known frequency could be generated. Today these waves originally produced by Hertz in his experiments are known as radio waves. Hertz never transmitted his radio waves over longer distances than a few hundred metres. The unit for frequency was changed from 'cycles per second' to the 'hertz', honouring the contribution of Heinrich Hertz.

Using the microwave apparatus available in schools, large wax prisms and metal plates, students can reproduce many of these results — demonstrating that electromagnetic waves have all the properties of light.

Radio waves and their frequencies

The discovery of radio waves was made by Hertz; however, the development of a practical radio transmitter was left to the Italian, Guglielmo Marconi (1874–1937). Marconi's experiments showed that for radio waves:

- long wavelengths penetrate further than short wavelength waves
- tall aerials were more effective for producing highly penetrating radio waves than short aerials.

The earliest radio messages were sent in 1895 by Marconi across his family estate, a distance of approximately three kilometres. By 1901 he was sending radio messages across the Atlantic Ocean from Cornwall, England to Newfoundland, Canada.

Different frequency radio waves can be generated easily and precisely by oscillating electric currents in aerials of different length. This is because the frequency of the waves generated faithfully matches the frequency of the AC current generating them. This ease of generation allows radio waves to be utilised extensively for many purposes. Applications include communication technologies, such as radio, television and mobile phones, and other technologies such as microwave cooking and radar. Essentially, the only difference between any of these waves used for these different purposes is the frequency of the waves generated by the transmitting aerials. For communications, sections of the available electromagnetic spectrum or bands of spectrum are used (see Figure 10.4). These chunks of spectrum use many single frequencies to transmit without interference.

10.2.3 The emission of light by hot objects: black body radiation

When an object such as a filament in a light globe is heated (but not burned), it glows with different colours: black, red, yellow and blue-white as it gets hotter. To understand how radiation is emitted for all objects, and how the wavelength of the radiation varies with temperature, creative experiments involving the behaviour of standard objects called 'black bodies' were required. A black body is one that absorbs all incoming radiation. The use of black bodies was necessary because all objects behave slightly differently in terms of the radiation they emit at different temperatures. Scientists could use the standard black body in experiments to study the nature of radiation emitted at different temperatures, and then extrapolate their findings for other objects.

As an example of an object used to model a black body, imagine you drilled a very small hole through the wall of an induction furnace (an efficient oven in which the temperature can be set to known values). At a temperature of 1000 °C, the walls of such a model black body will emit all types of radiation, including visible light and infra-red and ultraviolet radiation, but they will not be able to escape the furnace except through the small hole. They will be forced to bounce around in the furnace cavity until the walls of the furnace absorb them. As the walls absorb the radiation, they will increase in energy. This causes the walls to release radiation of a different wavelength, eventually establishing an equilibrium situation. All radiation entering through the small hole is absorbed by the walls, so the radiation leaving the hole in the side of the furnace is characteristic of the equilibrium temperature that exists in the furnace cavity. This emitted radiation is given the name black body radiation.

As Figure 10.9 shows, the radiation emitted from a black body extends over all wavelengths of the electromagnetic spectrum. However, the relative intensity varies considerably and is characteristic of a specific temperature.

Black bodies absorb all radiation that falls on them. That energy is spread throughout the object. The cavity walls within the black body also get hotter. As the walls of the cavity get hotter, the emission of more intense, shorter wavelength radiation from the cavity occurs. Physicists used a spectrometer to measure how much light of each colour, or wavelength, was emitted from the hole in the side of the black body models they constructed. The shape of the radiation versus intensity curves on the graphs they created presented a problem for the physicists attempting to explain the intensity and wavelength variations that occurred quantitatively.

The problem was how to explain the results theoretically. The traditional mathematics based on thermodynamics predicted that the pattern of radiation should be different to that which the physicists found occurred.

The 'classical' wave-theory of light predicted that, as the wavelength of radiation emitted becomes shorter, the radiation intensity would increase. In fact, it would increase without limit. This would mean that, as the energy (that was emitted from the walls of the black body and then re-absorbed) decreased in wavelength from the visible into the ultraviolet portion of the spectrum, the intensity of the radiation emitted from the hole in the black body would approach infinity. This increase in energy level would violate the principle of conservation of energy and could not be explained by existing theories. This effect was called the 'ultraviolet catastrophe'.

The experimental data from black body experiments (see Figure 10.10) showed that the radiation intensity curve corresponding to a given temperature has a definite peak, passing through a maximum and then declining. This could not be explained.

The German scientist, Max Planck, arrived at a revolutionary explanation for the nature of the radiation emitted in experiments. Planck proposed that energy would be exchanged between the particles of the black body and the equilibrium radiation field. Using an analogy with the transmission of radio waves (with an aerial of specific length indicating the frequency of the radio wave radiation produced), the relatively high frequency of light emitted by a black body required an 'aerial' of a size similar to that of the atom for its production. The question was, how did this come to be?

Planck came up with a revolutionary idea to explain the results observed in experiments. He assumed that the radiant energy, although exchanged between the particles of the black body and the radiant energy field in continuous amounts, may be treated statistically as if it was exchanged in multiples of a small 'lump'. Each lump is characteristic of each frequency of radiation emitted. He described this small, average packet as a 'quantum' of energy that could be described by h*f*, where *f* was the frequency, and h a small constant, now called 'Planck's constant' ($\mathrm{h} = 6.63 \times 10^{-34}$ J s).

Therefore:

$$E = \mathrm{h}f$$

where

E = energy, measured in joules

h = Planck's constant = 6.63×10^{-34} J s

f = frequency in hertz.

FIGURE 10.9 The peak in intensity moves to lower wavelength and higher frequency radiation with increasing temperature.

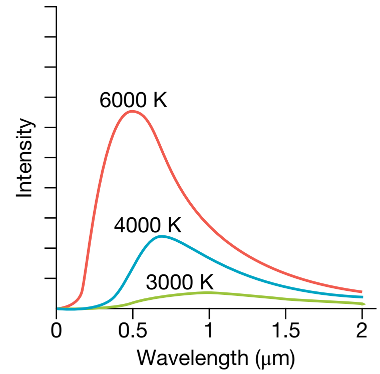
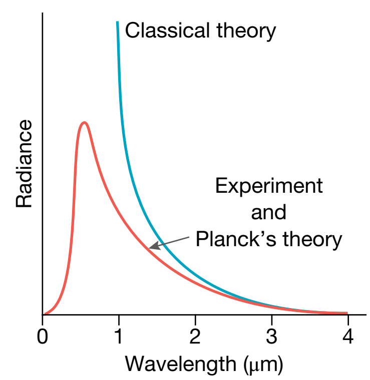


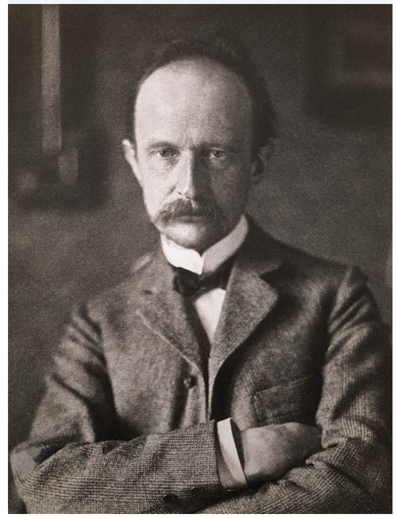
FIGURE 10.10 The predicted curve for the classical model and the actual curve obtained from experiment.



This equation models the quantum relationship. This modification was seen by Planck as a small correction to classical thermodynamics. It turned out to be a most significant step towards the development of a totally new branch of physics: the quantum theory. Although he is considered to be the first to introduce this theory, Planck was never comfortable with the strict application of the quantum theory. He had invented the quantum theory but believed that all he had really done was to invent a mathematical trick to explain the results of black body radiation experiments. He failed to accept the quantisation of radiation until later in his career when the quantum theory was backed up with more examples and supporting evidence.

Armed with the Planck relationship and the knowledge that the speed of electromagnetic radiation ($c = 3 \times 10^8$ m s^{-1}) was the product of the frequency of the radiation and the wavelength of the radiation ($c = f \times \lambda$), the energy in one quantum, or **photon**, of light of any known wavelength was then able to be determined.

FIGURE 10.11 Max Planck (1858-1947) has come to be recognised as one of the key people involved in the development of modern physics.



10.2 SAMPLE PROBLEM 1

What is the energy of an ultraviolet-light photon, wavelength $= 3.00 \times 10^{-7}$ m?

SOLUTION:

$$c = f\lambda \text{ so } f = \frac{c}{\lambda}$$

$$\begin{aligned} E &= hf \\ &= h\frac{c}{\lambda} \\ &= \frac{6.63 \times 10^{-34} \times 3.00 \times 10^{8}}{3.00 \times 10^{-7}} \\ &= 6.63 \times 10^{-19}\ \text{J} \end{aligned}$$

In this way the energy of a light photon of any known wavelength of light can be determined.

10.2 Exercise 2

1 Red light has a wavelength of 650 nm.
- (a) Determine the frequency of this light as a wave.
- (b) Determine the energy of a photon of light associated with this red light.

2 A stream of photons, each with energy 3.2×10^{-18} J strikes a solar cell.
- (a) Calculate the frequency of this light modelled as a wave.
- (b) Determine the wavelength of this light.
- (c) Classify the type of electromagnetic wave this stream of photons is associated with.

10.2.4 Wein's law

The hotter objects are, the more light they emit — hence, the more energy they emit per second and the brighter they appear. Additionally, the wavelength of the light with the greatest intensity is decreased — that is, the frequency of this light is higher. The wavelength having maximum intensity is given the symbol λ_{max} and is only dependent on the temperature of the body T, stated in kelvin.

Wein's law asserts that $\lambda_{max} = \dfrac{b}{T}$ and, also

$$\lambda_{max}T = \text{constant} = 2.9 \times 10^{-3}\ \text{K m}$$

This equation is a result of the Stefan-Boltzmann equation for black body radiators, which is reliant on the notion that charged particles undergoing collisions emit single quanta of radiation according to Planck's equation $E_\gamma = hf$ rather than accelerated charged particles emitting light as a wave in all directions. Without physicists yet knowing the importance of this result, the birth of quantum mechanics was taking shape. The excellent agreement between the observed emission spectrum of a black body radiator with the Stefan Boltzmann equation and the lack of any ultraviolet catastrophe meant that the quantum model of light was here to stay. The only problem was that the quantum model of light was contradictory to the well-established wave model of light.

10.2 SAMPLE PROBLEM 2

(a) A hotplate has a glow that is dark red. The wavelength of the most intense light is 4.0 μm. What is the surface temperature of the hotplate in kelvin?
(b) The skin temperature of a normal human is approximately 32 °C. What is the wavelength of the most intense light emitted by the skin of a normal human?

SOLUTION:

(a) Rearrange Wein's law; thus, $T = \dfrac{2.9 \times 10^{-3}}{4.0 \times 10^{-6}} = 725\ \text{K}$

(b) Rearrange Wein's law; thus, $\lambda_{\text{max}} = \dfrac{2.9 \times 10^{-3}}{32 + 273.15} = 9.5 \times 10^{-6}\ \text{m}$.

10.2 Exercise 3

1 The surface of the Sun is approximately 5800 K. What is the wavelength of the most intense light emitted by the surface of the Sun?
2 The temperature of a glass of cold water is 10 °C, and hot water from a tap is measured to be 60 °C. Find the ratio $\dfrac{\lambda_{\text{max, hot}}}{\lambda_{\text{max, cold}}}$ where λ_{max} is the wavelength of the most intense light.

10.3 The photoelectric effect

10.3.1 The need for a new model for light

Classical physics can be described, in broad terms, as physics up to the end of the nineteenth century. It relied on Newton's mechanics and included Maxwell's theories of electromagnetism. Classical physics still applies to large-scale phenomena and to the motion of bodies at speeds very much less than the speed of light. The quantum theory applies to the very small scale, particularly at the atomic level. Energy is believed to occur in discrete 'packets' or 'quanta'. Energy packets can be absorbed by an atom, and then re-radiated.

Classical physics predicts that the emission of electromagnetic radiation is continuous; that is, it can occur in any amount.

A useful analogy is that of a slippery dip. Classical physics says that the difference in height from the top to the bottom is a continuous 'slide'. Quantum mechanics says that there are a set number of small steps (the ladder) between the top of the slide and the ground.

The difference between the classical description of 'continuous' energy and the new discrete, or quantised description of energy may be understood with a simple 'thought experiment'. Consider an old-fashioned balance, pivoted in the centre, with large pans suspended from each side. On one side we place a bucket with water and on the other we add house bricks to balance the bucket and water. We can increase the weight of bricks by adding or subtracting one brick at a time. If each brick has a mass of 2 kg, then our smallest packet, or quantum, of mass is 2 kg. On the other side, we can increase the weight by adding water. Without extending this example too far, we have a discrete variable (the bricks) and a continuous variable (the water) that can be added in smaller and smaller drops.

The continuous variable (water) is similar to the classical model of energy, and the bricks correspond to the quantum model in which energy can be exchanged in multiples of a small number. This simple example uses only one size of brick, or quantum. In fact the size of the quantum value varies with the energy effect we are studying.

Why don't we see these 'jumps' in light? The constant, h, now called 'Planck's constant', is equal to 6.63×10^{-34} J s. The quantum of energy is very small and, for all practical purposes, it is too small to be observed.

10.3.2 The photoelectric effect: Evidence that light is not a type of wave

The **photoelectric effect** is the name given to the release of electrons from a metal surface exposed to electromagnetic radiation. For example, when a clean surface of sodium metal is exposed to ultraviolet light, electrons are liberated from the surface.

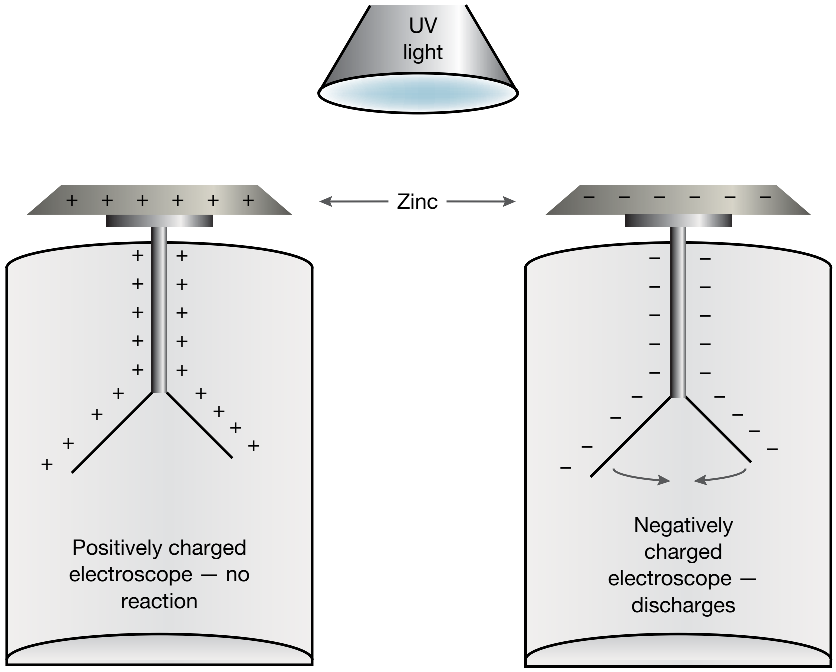
The photoelectric effect is one of several processes for removing electrons from a metal surface. The effect was first observed by Hertz in 1887 when investigating the production and detection of electromagnetic waves using a spark gap in an electric circuit. Hertz used an induction coil to produce an oscillating spark. Hertz called the transmitting loop, spark A, and the detecting loop, spark B. In Hertz's own words, describing his detection of the photoelectric effect:

'I occasionally enclosed spark B in a dark case so as to more easily make the observations; and in so doing I observed that the maximum spark length became decidedly smaller inside the case than it was before. On removing, in succession, the various parts of the case, it was seen that the only portion of it which exercised this prejudicial effect was that which screened the spark B from spark A. The glass partition exhibited this effect not only in the immediate neighbourhood of spark B, but also when it was interposed at greater distance from B between A and B.'

Hertz had discovered the photoelectric effect. He had found that illuminating the spark gap in the receiving loop with ultraviolet light from the transmitting gap gave stronger sparks in the receiving loop. Glass used as a shield between the transmitting and receiving loops blocked the UV. This reduced the intensity of sparking in the receiving loop. When quartz was used as a shield, there was no drop in the intensity of sparking in the receiving loop. Quartz allowed the UV from the transmitted spark to fall on the detector.

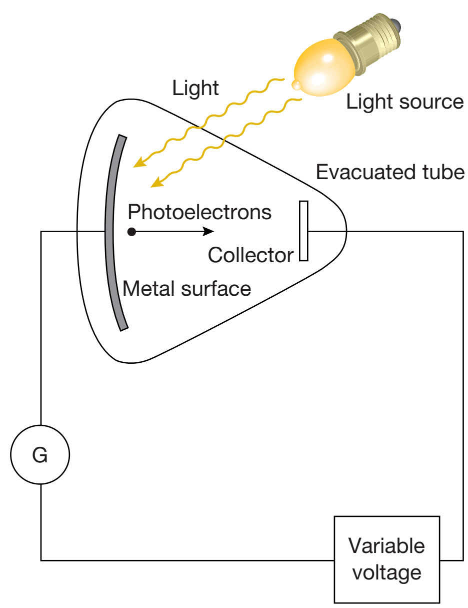
Wilhelm Hallwachs (1859–1894) read the extract written by Hertz describing the photoelectric effect in a journal and designed a simpler method to measure this effect. He placed a clean plate of zinc on an insulating stand and attached it by a wire to a gold leaf electroscope. He charged the electroscope negatively and observed that the charge leaked away quite slowly. When the zinc was exposed to ultraviolet light from an arc lamp, or from burning magnesium, charge leaked away quickly. If the electroscope was positively charged, there was no fast discharge (see Figure 10.12).

FIGURE 10.12 A positively charged electroscope is not affected by illumination with UV light, while the charge on a negatively charged electroscope discharges.



In 1899, J.J. Thomson established that the ultraviolet light caused electrons to be emitted from a sheet of zinc metal and showed that these electrons were the same particles found in cathode rays. He did this by enclosing the metal surface to be exposed to ultraviolet light in a vacuum tube (see Figure 10.13).

FIGURE 10.13 Apparatus used to demonstrate the photoelectric effect.



The new feature of this experiment was that the electrons were ejected from the metal by radiation rather than by the strong electric field used in the cathode ray tube. At the time, recent investigations of the atom had revealed that electrons were contained in atoms and it was proposed that perhaps they could be excited by the oscillating electric field.

In 1902, Hungarian-born German physicist Philipp von Lenard (1862–1947) studied how the energy of emitted photoelectrons varied with the intensity of the light used. He used a carbon arc lamp with which he was able to adjust the light intensity. He found in his investigations using a vacuum tube that photoelectrons emitted by the metal cathode struck another plate, the collector.

When each electron struck the collector, a small electric current was produced that could be measured. To measure the energy of the electrons emitted, Lenard charged the collector negatively to repel the electrons. By doing so, Lenard ensured that only electrons ejected with enough energy would be able to overcome this potential hill (see Figure 10.14). Surprisingly, he found that there was a well-defined minimum voltage, V_{stop} (see Figure 10.15).

FIGURE 10.14 The voltage applied across the variable resistor opposes he motion of the photoelectrons. The electrons that reach the opposite electrode create a small current, measured by the galvanometer. The value of the voltage at which the current drops to zero is known as the stopping voltage.

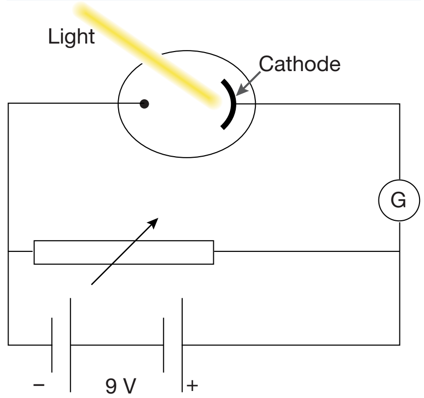
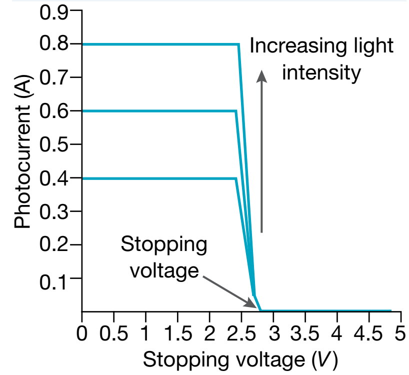


FIGURE 10.15 For a given frequency, photoelectrons are emitted with the same maximum kinetic energy because the electrons are all stopped by the same voltage. Increasing the intensity of the light increases the number of electrons released from the surface, causing an increase in the photocurrent.

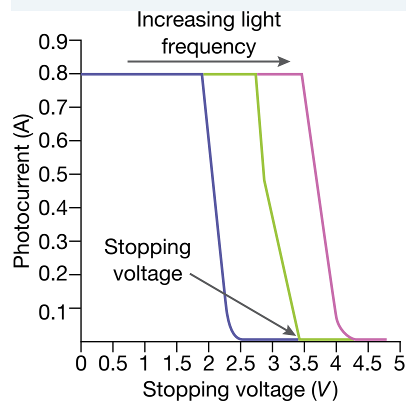


Lenard was also able to filter the arc light to investigate the effect that different frequencies of incident electromagnetic radiation had on photoelectron emission.

Lenard observed that:

- doubling the light intensity would double the number of electrons emitted
- there was no change in the kinetic energy of the photoelectrons as the light intensity increased
- the maximum kinetic energy of the electrons depends on the frequency of the light illuminating the metal, as Figure 10.16 shows.

FIGURE 10.16 For a given light intensity, increasing the frequency of the light increases the maximum kinetic energy with which the photoelectrons are emitted.



10.3 SAMPLE PROBLEM 1

(a) Electrons are emitted from a surface with a kinetic energy of 2.6×10^{-19} J. What is the size of the stopping voltage that will remove all of this energy from the electrons?

(b) What energy electrons will a 4.2 V stopping voltage stop?

SOLUTION:

(a) The kinetic energy of each electron is 2.6×10^{-19} J. The charge on an electron is 1.6×10^{-19} C.

$$E_\mathrm{k} = qV_0$$
$$2.6 \times 10^{-19}\ \mathrm{J} = 1.6 \times 10^{-19}\ \mathrm{C} \times V_0$$
$$V_0 = \frac{2.6 \times 10^{-19}\ \mathrm{J}}{1.6 \times 10^{-19}\ \mathrm{C}}$$
$$= 1.62\ \mathrm{V}$$
$$= 1.6\ \mathrm{V}(\text{accurate to 2 significant figures})$$

A stopping voltage of 1.6 V will stop the electrons emitted from the surface.

(b) The stopping voltage is 4.2 V. The charge of an electron is 1.6×10^{-19} C.

$$E_\mathrm{k} = qV_0$$
$$= 1.6 \times 10^{-19}\ \mathrm{C} \times 4.2\ \mathrm{V}$$
$$= 6.72 \times 10^{-19}\ \mathrm{J}$$
$$= 6.7 \times 10^{-19}\ \mathrm{J}\ (\text{accurate to 2 significant figures})$$

A stopping voltage of 4.2 V will stop electrons with energy 6.7×10^{-19} J.

10.3.3 Explanation of the photoelectric effect: Enter Albert Einstein

Classical physics was unable to explain the photoelectric effect. Maxwell had predicted electromagnetic radiation, and Hertz had confirmed that light was electromagnetic radiation. Light was a wave phenomenon, and according to the classical theory of Maxwell, somehow or other the light waves falling on the surface of a metal would cause the emission of electrons.

The key observations a theory had to explain are:

1. If the radiation falling on the metal surface is going to cause emission of photoelectrons, they may be emitted almost immediately after the light falls on the metal. There may be no significant time delay, even if the intensity of the light, and hence the rate at which energy is being transferred to the metal surface, is very low.
2. There is a cut-off (or threshold) frequency. Radiation of lower frequency than a particular value will not cause the emission of photoelectrons, regardless of how bright the light source is. (Very intense light, which carries a large amount of energy, cannot cause emission of photoelectrons if the frequency of the light is less than the threshold frequency.)
3. If the light does cause the emission of photoelectrons, increasing the intensity of the light will increase the number of photoelectrons emitted per second.
4. The energy of the photoelectrons does not depend on the intensity of the light falling on the surface of the metal, but it does depend on the frequency of the light. Light of higher frequency causes the emission of photoelectrons with higher energy.

Classical physics had difficulties with three of these four observations.

- According to a wave theory of light, the light waves would distribute their energy across the whole of the metal surface. It might be expected that all electrons in the atoms of the metal (or at least all of the outer electrons) would gain energy from the light waves. If the light was very faint, it could take a considerable time for the electrons to gain sufficient energy and for one electron to be able to escape from the metal surface, but this is not what is observed.
- There was no way to explain the cut-off frequency.
- There was no way to explain the relationship between the frequency of the light and the kinetic energy of the most energetic electrons.
- The only observation that could be explained by a wave theory was the fact that increasing the intensity of light increases the rate of emission of photoelectrons.

The problems associated with explaining the photoelectric effect were solved by Einstein in 1905. Although his paper is often referred to as his photoelectric paper, it was in fact a paper on the quantum nature of light, and the photoelectric effect was just one of several examples used by Einstein to illustrate the quantum nature of light.

In this paper, Einstein states: 'The simplest conception is that a light quantum transfers its entire energy to a single electron.' In other words, the light quantum is acting as a particle in a collision with an electron.

This 'light quantum' model of light is able to explain all four observations of the photoelectric effect. (The term photon was not introduced until 1926, but we will use it at this stage to refer to a 'light quantum'.)

If a photon strikes a metal surface, it will collide with a single electron in an atom in the metal. All of the energy of the photon will be passed on to the electron. This electron may gain sufficient energy to escape from the metal surface, so there is no problem with a time delay. Even with very faint light, the first photon to strike the metal surface could possibly cause the emission of an electron.

The energy of a photon is related to its frequency ($E = \mathrm{h}f$). A certain minimum energy is required to cause the emission of an electron from the metal surface (this is the **work function** for that metal). If the energy of the photon is less than this value, the photon cannot cause an electron to be emitted. This explains the threshold frequency.

Increasing the intensity of the light increases the rate at which photons fall on the metal surface, hence the rate of emission of photoelectrons increases.

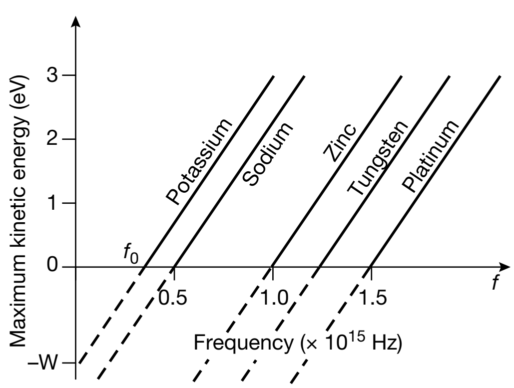
As the frequency of the light is increased beyond the cut-off frequency, more energy is provided to the electrons, hence the kinetic energy of the most energetic electrons increases.

This last idea played an important part in the development of the particle nature of light in the photoelectric effect. In 1905, observations of sufficient accuracy were not available to Einstein to test his equation; it was not until 1916 that Millikan, reluctantly, provided that evidence. Little attention had been paid to Einstein's ideas of the photoelectric effect in the intervening ten years. Millikan announced his results in 1916 but said, 'The Einstein equation accurately represents the energy of the electron emission under irradiation with light [but] the physical theory upon which the equation is based [is] totally unreasonable.' However, Millikan also stated that his results, combined with Einstein's equation, provided 'the most direct and most striking evidence so far obtained for the reality of Planck's h'.

As previously discussed, when different metal surfaces are illuminated with monochromatic light, electrons may be ejected from the metal surface. These electrons are called photoelectrons. Different metals hold electrons with different forces. Providing the photons of light illuminating the metal have sufficient energy (are of a high enough frequency) to overcome the energy holding the electrons in the metal, the electrons may be emitted. Only a small proportion of such electrons will in fact escape from the metal surface, and the emitted electrons will have a spread of energies, as some electrons may have required energy to move them to the metal surface. We will deal with the most energetic electrons emitted.

If a graph is plotted of the maximum kinetic energy of the emitted electrons versus the frequency of the light, the gradient of the lines representing different metals is the same (see Figure 10.17). The point at which the lines intercept the frequency axis is a measure of the threshold frequency for that metal. If the frequency of the monochromatic light is below this threshold frequency, no photoelectrons will be emitted from the metal surface.

FIGURE 10.17 This graph shows the maximum kinetic energy with which the photoelectrons are emitted versus the frequency of light, for five different metals. Note that the gradient of all the lines is equal to Planck's constant.



The lines for all of the metals are parallel and have a gradient equal to Planck's constant.

If we apply the general gradient equation $y = mx + b$ to any of the lines on this graph, we find that:

$$E_{\text{k max}} = \text{h}f - W$$

This equation is an energy equation:

$E_{\text{k max}}$ = the maximum kinetic energy of an emitted electron

W = the minimum energy required to remove the electron from the metal surface (the work function of the metal)

hf = the energy of the incident photon.

The energy of Einstein's 'light quantum' is hf, so this equation represents an interaction between an individual quantum of light (a photon) and an individual electron.

Of course, we now have light behaving as a particle in the photoelectric effect but as a wave in other phenomena (such as interference and diffraction). The photon has a dual wave and particle nature.

PHYSICS FACT

Both Planck and Einstein lived in Germany during the early part of the twentieth century. Working in the same area of physics, they were firm friends. However, during World War I, this friendship became strained. Einstein was a pacifist, while Planck strongly supported the German cause, even though he lost his son in battle in 1915.

With the rise of Hitler and the anti-Semitism movement, Einstein, who was Jewish, emigrated to the United States during the 1930s. Planck was able to continue his academic career in Berlin, even in the face of the hostility of anti-Semitism groups towards the 'decadent Jewish science' of relativity and the quantum theory.

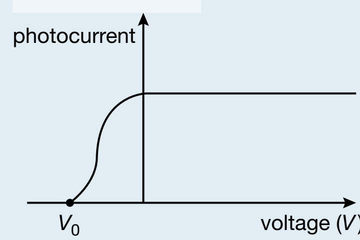
It is difficult to understand the pressures experienced by Planck, who tried to protect his Jewish friends and students throughout the period. Unlike Einstein, he did not see the moral imperative of opposing Hitler but tried to compromise and work within the system. Einstein remained a pacifist, yet some would say that he compromised some of these ideals with his support of the development of the atom bomb. He later wrote of the pain he experienced when the bombs were finally used.

10.3 SAMPLE PROBLEM 2

The diagram below shows the current-versus-stopping voltage curve for a typical photoelectric cell using green light.

The colour is changed to blue, but with a lower intensity. Sketch the curve that would result from these changes.

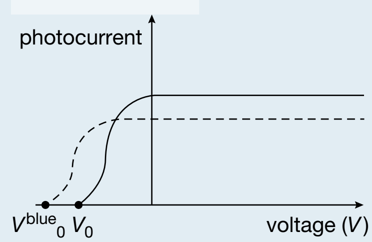
FIGURE 10.18



SOLUTION:

Because blue light has a higher frequency than green light, the stopping voltage would be greater. The lower intensity would make the photocurrent smaller. This is shown in the diagram below.

FIGURE 10.19



10.3.4 A photon model for the photoelectric effect

Almost thirty years after the first observation of the photoelectric effect, experimental measurements confirmed the need for a photon model for light. The wave model for light was incapable of explaining the observations of the photoelectric effect.

TABLE 10.1 Timeline of key discoveries about the photoelectric effect.

Date	Event
1887	It all started with Hertz carefully noting the unusual behaviour of sparks across the gaps in his radio wave detector circuit. This was the first observation of the photoelectric effect.
1901	Max Planck solves the black body radiation problem theoretically, paving the way for light to be modelled not only as a wave but also as a localised particle with energy proportional to the frequency of the light, f.
1902	Philipp Lenard carried out experiments to accumulate knowledge about the behaviour of electrons emitted by light. There were several puzzling aspects to his results — electron energies did not depend on the light intensity and there was a unique cut-off frequency for each material.
1905	The flash of insight was Albert Einstein's, when he realised that all of Lenard's observations could be explained if he changed the way he thought about light — if light energy travelled as particles not waves. He used the particle model to predict that the graph of stopping voltage versus frequency would be straight, with a slope that was the same for all electron emitters.
1915	Robert Millikan sealed the success of Einstein's theory with plots of V_0 versus f for the alkali metals that were straight and parallel to one another. He used the plots to measure Planck's constant. The photon energy was hf.

TABLE 10.2 Observations made from the photoelectric effect and model predictions.

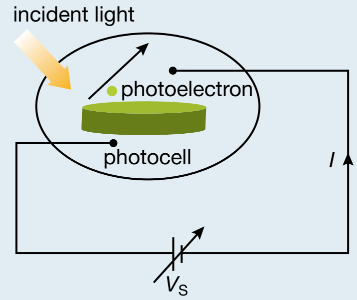
Observation	Wave model prediction	Photon model prediction
For a given frequency of light, the photocurrent is dependent in a linear fashion on the brightness or intensity of light.	The wave model makes no significant predication other than that brighter light should produce electrons with greater energy, which is not the case.	Intensity of light relates to the number of photons per second striking the photocell. We would expect the photocurrent to be dependent on the intensity of light.
The energy of photoelectrons is independent of intensity of light and only linearly dependent on frequency.	The energy of electrons is dependent on the intensity of light: the bigger the amplitude of the wave, the larger the energy transferred to electrons.	The energy of photoelectrons is linearly dependent on the frequency of light, provided we interpret the energy of a single photon of light as equal to hf.
There is no significant time delay between incident light striking a photocell and subsequent emission of electrons, and this observation is independent of intensity.	Time delay to be shorter with increasing intensity.	No time delay expected as individual photons of light strike photocell and transfer energy to individual electrons.
There exists a threshold frequency below which the photoelectric effect does not occur, and this threshold is independent of intensity.	No threshold effect should exist, as energy transfer to electrons from light source is accumulative and eventually emission will occur.	A threshold frequency is predicted, as photons with energy less than the work function are incapable of freeing electrons from the photocell.

10.3 SAMPLE PROBLEM 3

Light with a wavelength of 425 nm strikes a clean metallic surface and photoelectrons are emitted. A voltage of 1.25 V is required to stop the most energetic electrons emitted from the photocell.

(a) Calculate the frequency of a photon of light whose wavelength is 425 nm.
(b) Calculate the energy in joules and also in electron volts of a photon of light whose wavelength is 425 nm.
(c) State the energy of the emitted electron in both electron volts and joules.
(d) Calculate the work function W of the metal in eV and J.
(e) Determine threshold frequency f_0 and consequently the maximum wavelength of a photon that will just free a surface electron from the metal.
(f) Light of wavelength 390 nm strikes the same metal surface. Calculate the stopping voltage.

FIGURE 10.20 Schematic of a simple photocell.



SOLUTION:

(a) $f = \dfrac{c}{\lambda}$

$= \dfrac{3.0 \times 10^8}{4.25 \times 10^{-7}}$

$= 7.06 \times 10^{14}$ Hz

$= 7.1 \times 10^{14}$ Hz

(b) $E = \text{h}f$

$= 6.63 \times 10^{-34} \times 7.06 \times 10^{14}$

$= 4.68 \times 10^{-19}$ J

To convert energy in joules into energy in electron volts, divide by 1.6×10^{-19} joules eV^{-1}.

$E = \dfrac{4.68 \times 10^{-19}}{1.6 \times 10^{-19}}$

$= 2.92$ eV

$= 2.9$ eV

(c) Since the stopping voltage is 1.25 V, the energy of the emitted electron is 1.25 eV. The energy in joules can be found by multiplying by 1.6×10^{-19}. Thus the energy is:

$1.25 \times 1.6 \times 10^{-19} = 2.00 \times 10^{-19}$ J.

(d) Using the equation $E_{K_{max}} = \text{h}f - W$, the work function can be found. We know that when the photon energy hf equals 2.92 eV, the electrons have an energy of 1.25 eV. Thus $1.25 = 2.92 - W$. Thus:

$W = 2.92 - 1.25 = 1.67 \text{ eV} = 2.67 \times 10^{-19} \text{ J} = 2.7 \times 10^{-19}$ J.

(e) Again use the equation $E_{K_{max}} = \text{h}f - W$, The threshold frequency f_0 is the frequency below which the photoelectric effect does not occur. At this frequency, electrons are just not able to leave the surface. This model implies $0 = \text{h}f_0 - W$. Rearrange this equation to give the useful result:

$f_0 = \dfrac{W}{\text{h}}$

$= \dfrac{2.67 \times 10^{-19}}{6.63 \times 10^{-34}}$

$= 4.03 \times 10^{14}$ Hz.

The maximum wavelength is thus:

$$\lambda = \frac{c}{f_0}$$
$$= \frac{3.0 \times 10^8}{4.03 \times 10^{14}}$$
$$= 7.4 \times 10^{-7} \text{ m or } 740 \text{ nm.}$$

(f) Use the equation $E_{k_{max}} = h\frac{c}{\lambda} - W$ to find the energy of the emitted elecrons. When this is known, the stopping voltage can be readily found. It is convenient to use eV here.

$$E_{k_{max}} = \frac{4.15 \times 10^{-15} \times 3.0 \times 10^8}{3.90 \times 10^{-7}} - 1.67$$
$$= 3.19 - 1.67$$
$$= 1.52 \text{ eV}$$
$$= 1.5 \text{ eV}$$

A stopping voltage of 1.5 V is required to stop the emitted electrons

10.3 SAMPLE PROBLEM 4

The table below gives some data collected by students investigating the photoelectric effect using a photocell with a lithium cathode. This cell is illustrated in the schematic diagram in Figure 10.21.

TABLE 10.3

Wavelength of light used (nm)	Frequency of light used $\times$ 10^{14} (Hz)	Photon energy of light used, E_{photon} (eV)	Stopping voltage readings (V)	Maximum photo-electron energy E_e (J)
663			0.45	
	6.14			1.84×10^{-19}

(a) Complete Table 10.3.

(b) Using only the two data points supplied in the table, plot a graph of maximum photoelectron energy in joules versus photon frequency in hertz for the lithium photocell.

(c) Using only your graph, state your values for the following quantities. In each case, state what aspect of the graph you have used.

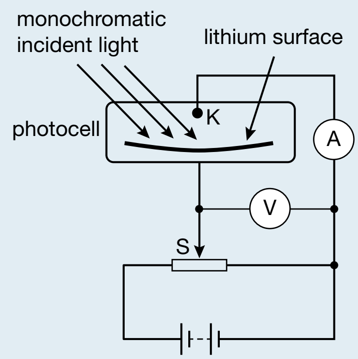
i. Planck's constant, h, in the units J s and eV s as determined from the graph

ii. The threshold frequency, f_0, for the metal surface in Hz as determined from the graph

iii. The work function, W, for the metal surface as determined from the graph, in the units J s and eVs.

(d) On the same axes, draw and label the graph you would expect to get when using a different photocell, given that it has a work function slightly larger than the one used to collect the data in the table above.

FIGURE 10.21



A new photocell is now investigated. When light of frequency 9.12×10^{14} Hz is used, a stopping voltage of 1.70 V is required to stop the most energetic electrons.

(e) Calculate the work function of the new photocell, giving your answer in both joules and electron volts.

(f) When the battery voltage of the new photocell is set, the photocurrent is measured to be 48 μA. The intensity of the light is now doubled. Describe what happens in the electric circuit with the power supply voltage set to 0 V when the light intensity is doubled.

(g) With the intensity still doubled, the voltage is now slowly increased from 0 and the photocurrent slowly reduces to 0 A. State the stopping voltage when the current first equals 0 A with the light intensity still doubled.

SOLUTION:

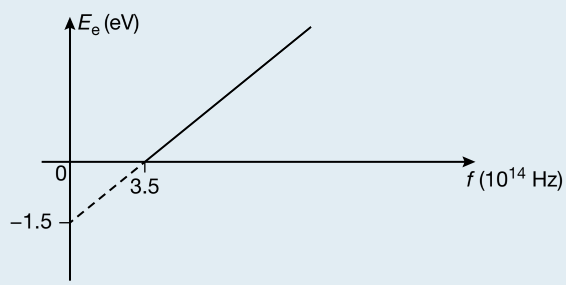
(a) Use $c = f\lambda$ to complete columns 1 and 2. Use $E = hf$ to complete column 3, and use the conversion factor for joules to eV to complete columns 4 and 5.

TABLE 10.4

Wavelength of light used (nm)	Frequency of light used $\times 10^{14}$ (Hz)	Photon energy of light used, E_{photon} (eV)	Stopping voltage readings (V)	Maximum photoelectron energy E_e (J)
663	4.52	1.88	0.45	7.20×10^{-20}
488	6.14	2.55	1.15	1.84×10^{-19}

(b) The graph will contain two points representing the fact that light of frequency 4.52×10^{14} Hz will produce electrons of energy 0.45 eV and light of frequency 6.52×10^{14} Hz will produce electrons of energy 1.15 eV. A line drawn containing these two data points will give a work function of 1.5 eV and a threshold frequency of 3.5×10^{14} Hz.

FIGURE 10.22



(c) i. Planck's constant = gradient of graph

$$= \frac{1.84 \times 10^{-19} - 7.20 \times 10^{-20}}{(6.14 - 4.52) \times 10^{14}} = 6.9 \times 10^{-34} \text{ J s},$$

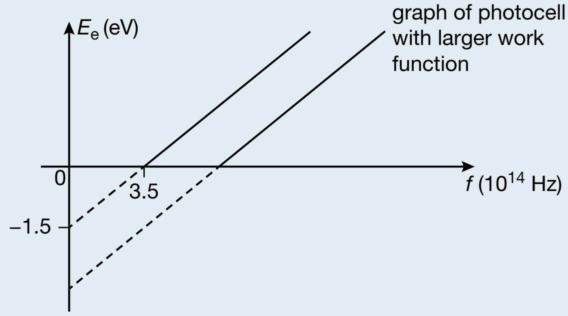
which is close to the accepted value. It also has the value 4.3×10^{-15} eVs.

ii. From the line of best fit in graph (b), the threshold frequency = x-axis intercept = 3.5×10^{14} Hz.

iii. From the line of best fit in the graph (b), the work function = y-axis intercept = 2.4×10^{-19} J = 1.5 eV.

(d)

FIGURE 10.23



(e) Use $E_e = E_{photon} - W$ to calculate the work function, W.

$$1.7 \times 1.6 \times 10^{-19} = 6.6 \times 10^{-34} \times 9.12 \times 10^{14} - W$$
$$W = 6.02 \times 10^{-19} - 2.72 \times 10^{-19}$$
$$= 3.3 \times 10^{-19} \text{ J}$$
$$= 2.1 \text{ eV}$$

(f) With the light intensity doubled, the photocurrent would also double.

(g) The stopping voltage would remain the same, 1.7 V, as the colour and hence the frequency of the light source is unchanged.

10.3 Exercise 1

1 Table 10.5 gives some data collected by students investigating the photoelectric effect using a photocell with a clean metallic cathode.

TABLE 10.5

Wavelength of light used (nm)	Frequency of light used $\times 10^{14}$ (Hz)	Photon energy of light used, E_{photon} (eV)	Stopping voltage readings (V)	Maximum photoelectron energy E_e (J)
		3.19		3.78×10^{-19}
524			1.54	

(a) Complete the table.

(b) Using only the two data points supplied in the table, plot a graph of maximum photoelectron energy in joules versus photon frequency in hertz for the photocell.

(c) Using only your graph, state your values for the following quantities. In each case, state what aspect of the graph you have used.

i. Planck's constant, h, in the units J s and eV s as determined from the graph

ii. The threshold frequency, f_0, for the metal surface in Hz as determined from the graph

iii. The work function, W, for the metal surface as determined from the graph, in the units J s and eV s

(d) On the same axes, draw and label the graph you would expect to get when using a different photocell, given that it has a work function slightly larger than the one used to collect the data in the table above.

A new photecell is now investigated. When light of frequency 8.25×10^{14} Hz is used, a stopping voltage of 1.59 V is required to stop the most energetic electrons. In addition, when the battery voltage is set to 0 V, the photocurrent is measured to be 38 μA.

(e) Calculate the work function of the new photocell.
(f) Describe what happens in the electric circuit with the power supply voltage set to 0 V when the light intensity is halved.
(g) With the intensity still halved, the stopping voltage is now slowly increased from 0 V and the photocurrent slowly reduces to 0 A. State the stopping voltage when the current first equals 0 A with the light intensity still halved.

eBook*plus* RESOURCES

Model the photoelectric effect with this spreadsheet: Modelling the photoelectric effect
Searchlight ID: doc-0042

Explore more with these weblinks:
Explaining the photoelectric effect
The photoelectric effect

Watch this eLesson: Graphing the photoelectric effect
Searchlight ID: med-0421

Watch this eLesson: Photocurrent and light intensity
Searchlight ID: med-0422

Try out this interactivity: Using the photoelectric effect
Searchlight ID: int-0121

Complete this digital doc: Investigation: Producing and transmitting radio waves
Searchlight ID: doc-26320

10.4 Review

10.4.1 Summary

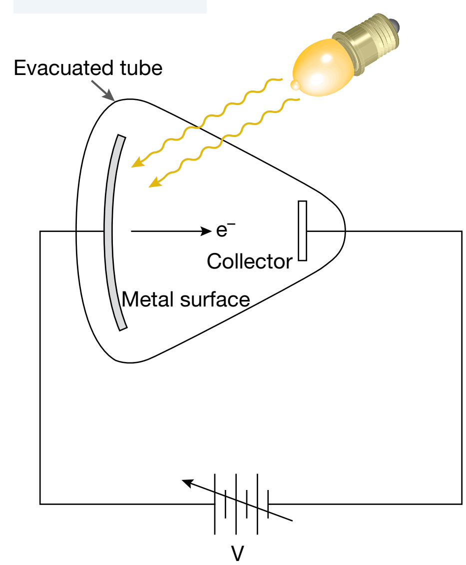
- In 1887, Hertz showed that both radio and light waves are electromagnetic waves — involving varying or changing electric fields, coupled with a changing magnetic field that was perpendicular to the electric field.
- The relationship $v = f\lambda$, relates the frequency (Hz) and the wavelength (m) to the velocity (m s^{-1}). In a vacuum, all electromagnetic waves travel at the speed of light (usually referred to as c) and thus c $= f\lambda$.
- Wein's law relates the wavelength of the most intense light λ_{max} and the temperature of a black body radiator T. This law assumes the Stefan-Boltzmann black body emission spectrum law, which is in excellent agreement with experimental data. Thus: $\lambda_{max}T = \text{constant} = 2.9 \times 10^{-3}$ K m. The wavelength is in metres and the temperature is in kelvin.
- An oscillating or vibrating atom can emit electromagnetic energy only in discrete packets, or quanta. For light, in particular, the basic quantum of energy is a photon.
- The electron volt is a useful unit of energy. An electron will acquire 1 eV when it passes through a potential difference of 1 V. Thus $1 \text{ eV} = 1.6 \times 10^{-19}$ J.
- The photoelectric effect is the release of electrons from materials, usually metals, by the action of light or some other electromagnetic radiation such as X-rays or gamma radiation.
- Quantum mechanics states that the amount of energy emitted from a black body (or any other atomic-level energy transformation) is quantised so that it can increase only in certain small steps.
- Albert Einstein explained the photoelectric effect, based on Planck's work on black body radiation, and included the development of the idea of a quantum of energy or photon being the basic unit of energy.
- As scientists, Einstein and Planck both experienced pressure from social and political forces and coped in different ways.

- When a photon hits electrons in a metal, it will release either all or none of its energy. An individual electron cannot accept energy from more than one photon.
- The amount of energy required to overcome the attractive forces of an electron in the electron 'sea' is called the work function of the electron.
- The maximum kinetic energy of the electron after it has been ejected from the metal's surface can be determined by measuring the stopping voltage, v_{stop}.

10.4.2 Questions

1. State two changes in observable quantities that occur when the temperature of a black body radiator increases.
2. If a black body was to gain sufficient energy to raise its temperature from 2000 K to 4000 K, describe how a plot of its radiant intensity versus radiant wavelength of the electromagnetic radiation would change.
3. Research the ultraviolet catastrophe and summarise why the fact this doesn't occur is evidence for quantised light emitted by a black body radiator (and therefore the light is not wavelike).
4. The sun has a surface temperature of approximately 5800 K.
   (a) Determine the wavelength of the most intense light emitted from its surface.
   (b) What is the colour seen by humans for light of this wavelength?
5. The colour red has a wavelength of approximately 650 nm.
   (a) What is this wavelength in metres?
   (b) What is the surface temperature of a black body radiator having λ_{max} = 650 nm?
   (c) If the temperature of the black body radiator increased by a factor of 2, what would the new λ_{max} now equal?
6. Examine figures 10.7 and 10.8. Explain why Hertz came to the conclusion that the waves produced by the sparks were polarised.
7. Demonstrate the dependence of colour on the temperature of a black body. In Figure 10.9, what is the significance of the peak of the curve?
8. Maxwell, Michelson and Hertz carried out experiments on electromagnetic waves of different frequencies. Compare their observations and discuss how an understanding of the electromagnetic spectrum was developed.
9. A beam of UV light of frequency 7.0×10^{15} Hz is incident on the apparatus shown in Figure 10.24. If the maximum kinetic energy of an emitted electron is 9.0×10^{-19} J, calculate:
   (a) the potential required to stop electrons reaching the collector
   (b) the work function of the material on which the light is shining
   (c) the threshold frequency of the material.

FIGURE 10.24



10. A scientist is investigating the effect of different types of radiation on the surface of a piece of sodium metal. A method is devised to cut a new surface across the sodium plate while in vacuo, since sodium is highly reactive and oxidises quickly. The apparatus is finally set up as shown in Figure 10.25.
The two variables under investigation will be the frequency, f, of the radiation and the kinetic energy, E_k, of the photoelectrons.
(a) Should the sodium plate be positively or negatively charged in order to make the proper investigations?
(b) Results for the experiment are as follows:

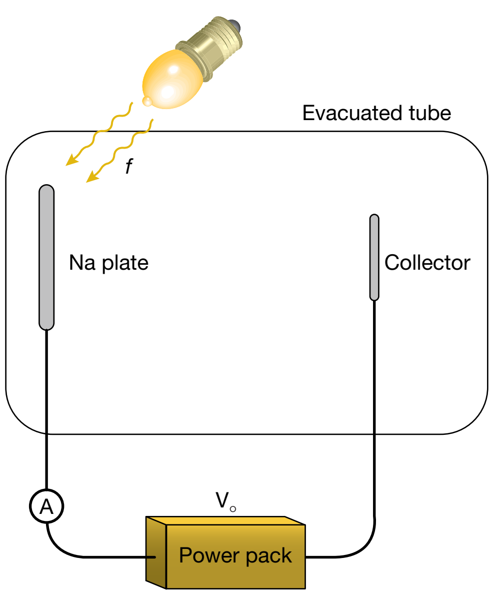
TABLE 10.6

Frequency of incident light ($\times 10^{14}$ Hz)	Stopping potential (volts)
5.4	0.45
6.8	1.00
7.3	1.15
8.1	1.59
9.4	2.15
11.9	2.91

Record these results on an $E_{k\ max}$ versus frequency graph.

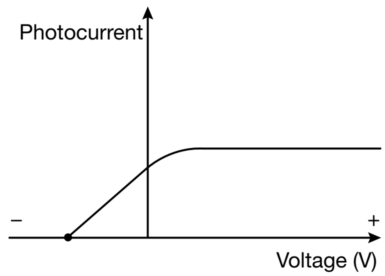
(c) Determine the threshold frequency for the sodium metal.
(d) Determine a value for Planck's constant, h, from your graph.
(e) What is the work function for sodium?

FIGURE 10.25



11. Above a particular (and specific to each material) threshold frequency of electromagnetic radiation, electrons are ejected immediately. Below this threshold frequency, electrons are never ejected. Explain how the photon model for light, rather than the wave model, explains this behaviour.
12. A photon collides with an electron and is scattered backwards so that it travels back along its original path. Describe and explain the expected wavelength of the scattered light.
13. The light from a red light-emitting diode (LED) has a frequency of 4.63×10^{14} Hz. What is the energy change of electrons that produce this light?
14. What is the stopping voltage when UV radiation having a wavelength of 200 nm is shone onto a clean gold surface? The work function of gold is 5.1 eV.
15. In Figure 10.26, the curve shows how the current measured in a photoelectric effect experiment depends on the potential difference between the anode and cathode.

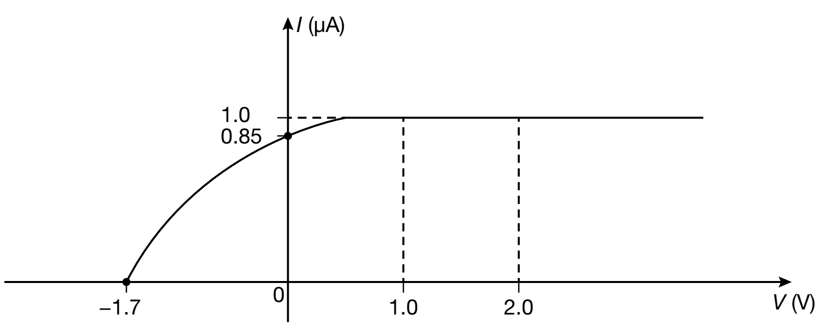
FIGURE 10.26



(a) Explain the curve. Why does it reach a constant maximum value at a certain positive voltage, and why does it drop to zero at a certain negative voltage?
(b) If the intensity of the light was increased without changing its frequency, sketch the curve that would be obtained. Explain your reasoning.
(c) If the frequency of the light was increased without changing its intensity, sketch the curve that would be obtained. Explain your reasoning.
(d) If the material of the cathode was changed, but the light was not changed in any way, sketch the curve that would be obtained. Explain your reasoning.

16. Figure 10.27 shows the current in a photoelectric cell versus the potential difference between the anode and the cathode when blue light is shone onto the anode.

FIGURE 10.27



(a) State the current when the voltage is 0 V.
(b) State the current when the voltage is +1.0 V.
(c) State the current when the voltage is increased to +2.0 V.
(d) Why does increasing the voltage have no effect on the current in the circuit?

(e) The polarity is now reversed and the voltage increased until the current drops to 0 A. State the stopping voltage and hence the maximum energy of electrons emitted from the anode.
(f) The light source is now made brighter without changing the frequency. Copy the figure and sketch a second curve that illustrates the effect of increasing the intensity of the blue light.
(g) The light source is now returned to its original brightness and green light is used. A current is still detected. Sketch a third curve to illustrate the effect of using light of a lower frequency.
(h) The apparatus is altered so that the anode consists of a metal with a smaller work function. Again blue light is used. Sketch a fourth curve to illustrate the effect of changing the anode without changing either the brightness or colour of the light.

17. In a photoelectric effect experiment, the threshold frequency is measured to be 6.2×10^{14} Hz.
    (a) Calculate the work function of the metal surface used.
    (b) If electrons of maximum kinetic energy 3.4×10^{-19} J are detected when light of a particular frequency is shone onto the apparatus, what is the stopping voltage?
    (c) With the same source of light, what is the wavelength and hence the momentum of the photons?
18. We can detect light when our eye receives as little as 2.00×10^{-17} J. How many photons of light, with a wavelength of 5.50×10^{-7} m, is this?
19. A red laser emitting 600 nm ($\lambda = 6.0 \times 10^{-7}$ m) wavelength light and a blue laser emitting 450 nm light emit the same power. Compare their rate of emitting photons.
20. One electron ejected from a clean zinc plate by ultraviolet light has kinetic energy of 4.0×10^{-19} J.
    (a) What would be the kinetic energy of this electron when it reached the anode, if a retarding voltage of 0.90 V was applied between anode and cathode?
    (b) What is the minimum retarding voltage that would prevent this electron reaching the anode?
    (c) All electrons ejected from the zinc plate are prevented from reaching the anode by a retarding voltage of 4.3 V. What is the maximum kinetic energy of the electrons ejected from the zinc?
    (d) Sketch a graph of photocurrent versus voltage for this metal's surface. Use an arbitrary photocurrent scale.
21. The waves of the electromagnetic spectrum share some similarities and have some differences. What are their similarities? What are their differences?
22. (a) What is the wavelength of the radio waves of broadcasting station 2 MMM in Sydney if the frequency of the broadcast is 104.9 MHz?
    (b) What is the energy of a photon of that wave?
23. Arrange the following electromagnetic waves in order of increasing energy levels:
    long-wave radio waves, gamma rays, blue light, red light, infra-red light, microwave radio waves, X-rays.
24. What were the features of radio waves that were demonstrated by Hertz's experiments?
25. What is the energy of an X-ray of wavelength 2.5×10^{-11} m?
26. Give four reasons why a particle model for light better explains the observations made for the photoelectric effect. In particular, explain why a wave model is inadequate for each reason.

PRACTICAL INVESTIGATIONS

Investigation 10.1: Producing and transmitting radio waves

Aim

To demonstrate the production and transmission of radio waves.

Apparatus

induction coil
Transformer rectifier with leads
Small transistor radio

Method

1. Adjust the gap on the induction coil to about 5 mm and adjust the transformer to 6 V DC.
   WARNING: The spark across the gap of an induction coil generates long-wavelength X-rays and short-wavelength ultraviolet radiation. These are potentially dangerous, and students should not stand closer than one metre.
2. Adjust the tuner of the radio, so that it does not receive a station.
3. Move around the room and try to estimate where the radio can receive the static noise from the spark.
4. Adjust the gap to 10 mm, and repeat the exercise.
5. Change the tuner of the radio and scan across the range of wavelengths.

Questions

1. What is the maximum distance from the induction coil at which the radio receives static noise from the 5 mm spark?
2. Is the distance different when the spark is produced by a 10 cm gap?
3. Can you detect any pattern in the static received at different wavelengths?
4. An induction coil is an example of a transformer. What can you infer about the voltage across the gap and the resulting charge movement observed as the spark?
5. In what form is the energy transferred from the spark to the radio? In what manner must charges move to produce this energy?

TOPIC 11 Light and special relativity

11.1 Overview

11.1.1 Module 7: The nature of light

Light and special relativity

Inquiry question: How does the behaviour of light affect concepts of time, space and matter?

Students:

- analyse and evaluate the evidence confirming or denying Einstein's two postulates:
  - the speed of light in a vacuum is an absolute constant
  - all inertial frames of reference are equivalent (ACSPH131)
- investigate the evidence, from Einstein's thought experiments and subsequent experimental validation, for time dilation, $t = \frac{t_0}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$, and length contraction, $l = l_0\sqrt{\left(1 - \frac{v^2}{c^2}\right)}$, and analyse quantitatively situations in which these are observed, for example:
  - observations of cosmic-origin muons at the Earth's surface
  - atomic clocks (Hafele–Keating experiment)
  - evidence from particle accelerators
  - evidence from cosmological studies
- describe the consequences and applications of relativistic momentum with reference to:
  - $p_v = \frac{m_0 v}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$
  - the limitation on the maximum velocity of a particle imposed by special relativity (ACSPH133)
- Use Einstein's mass–energy equivalence relationship ($E = mc^2$) to calculate the energy released by processes in which mass is converted to energy, for example: (ACSPH134)
  - production of energy by the Sun
  - particle–antiparticle interactions, eg positron–electron annihilation
  - combustion of conventional fuel

FIGURE 11.1 The velocity of a yacht can be measured relative to wind, land, water or other yachts, and all of these measurements can be different.

11.2 What is relativity?

11.2.1 Changes in understanding what is relative

The speed of an object depends on the **relative** motion of the observer. So do the object's time, kinetic energy, length and mass; that is, these properties are relative rather than fixed. Albert Einstein discovered that some of the physical properties that people assumed to be fixed for all observers actually depend on the observers' motions. But not everything is relative. The laws of physics and the speed of light are the same for all observers. Major developments in physics have come about at times when physicists such as Galileo and Einstein developed a clearer understanding of what is relative and what is not.

Albert Einstein (1879–1955) is one of the most famous figures in history, largely due to his work on relativity. Einstein did not invent the idea of relativity — it dates back to Galileo — but he brought it into line with nineteenth-century developments in the understanding of light and electricity, leading to some striking changes in how physicists viewed the world. Einstein established his ideas on the basis of two postulates:

- the speed of light in a vacuum is an absolute constant
- all inertial frames of reference are equivalent.

FIGURE 11.2 Albert Einstein (1879–1955)

The next section of this topic explores what is meant by relativity before returning to explain the meaning and some of the implications of these two postulates.

11.2.2 There is no rest

Let's start with a down-to-earth scenario. Consider a police officer pointing her radar gun at an approaching sports car from her car parked on the roadside. She measures the sports car's speed to be 90 km h^{-1}. This agrees with the speed measured by the driver of the sports car on his car's speedometer. However, another police car drives towards the sports car in the opposite direction at 60 km h^{-1}. A speed radar is also operating in this car, and it measures the speed of the sports car to be 150 km h^{-1}. So each police officer has a different measurement for the speed of the sports car. Which measurement is correct? The answer is that they are both correct — the speed measured for the car is relative to the velocity of the observer — but only the speed measured by the officer at rest on the roadside is relevant when receiving a speeding ticket.

The sports car is approaching the oncoming police car at the same rate as if the police car was parked and the sports car had a reading of 150 km h^{-1} on its speedometer. We say that the speed of the car is relative to the observer rather than being an invariant quantity, agreed on by all observers. The significance of relative speed becomes all too clear in head-on collisions. For example, you might be driving at only 60 km h^{-1}, but if you collide head-on with someone doing the same speed in the opposite direction, the impact occurs for both cars at 120 km h^{-1}!

FIGURE 11.3 A speed limit is the maximum allowed speed relative to the road.

Relativity is about the laws of physics being meaningful for all observers. Newton's First Law of Motion states that an object will continue at constant velocity unless acted on by an unbalanced net force. The speed itself does not matter. In the example above, this law works for both of the police officers, as do the other laws of motion.

FIGURE 11.4 The radar gun would measure a different speed if it was in a moving vehicle.

The Italian scientist Galileo Galilei (1564–1642) did not know about police cars and speed limits. His examples featured sailing ships and cannon balls, but the physics ideas were the same. In Galileo's time, much of physics was still based on ancient ideas recorded by the Greek philosopher Aristotle (384–322 BC). Aristotle taught that the Earth was stationary in the centre of the universe. Motion relative to the centre of Earth was a basis for Aristotelian physics, so a form of relativity was key to physics even before Galileo. But Galileo had to establish a new understanding of relativity before it became widely accepted that the Earth moved around the Sun.

Galileo's insight helped provide the platform for physics as we know it today, but the idea of a fixed frame of reference persisted. Following on from Galileo, Isaac Newton considered the centre of mass of the solar system to be at absolute rest. James Clerk Maxwell (1831–1879), who put forward the theory of electromagnetism, regarded the medium for electromagnetic waves (light) to be at rest. It was Einstein who let go of the concept of absolute rest, declaring that it was impossible to detect a place at absolute rest and therefore the idea had no consequence. Once again, relativity was updated to take into account the latest discoveries and enable physics to make enormous leaps of progress.

The speed (velocity) of bodies in motion is truly relative to whoever is measuring it. We will return to Einstein's advances shortly, but let's look at some more examples from Galilean relativity.

FIGURE 11.5 Two different measurements of the speed of a car.

FIGURE 11.6 Galileo Galilei (1564–1642), from a nineteenth-century engraving

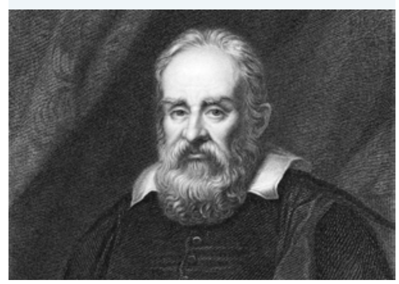
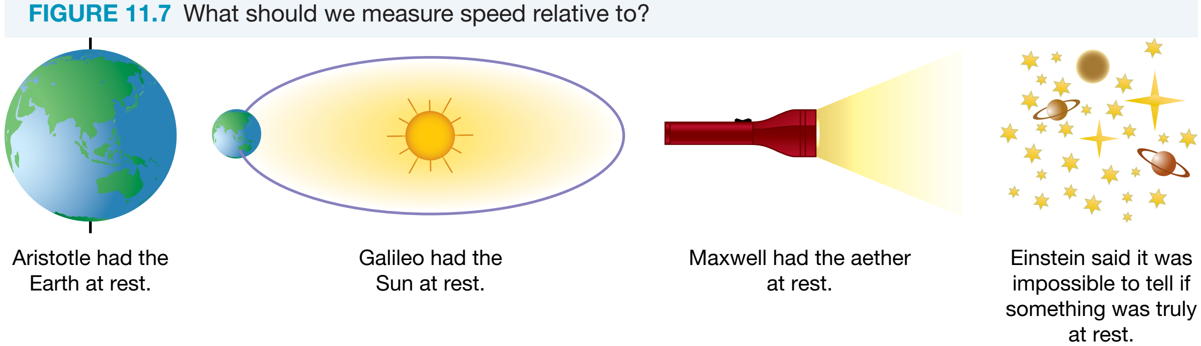


FIGURE 11.7 What should we measure speed relative to?



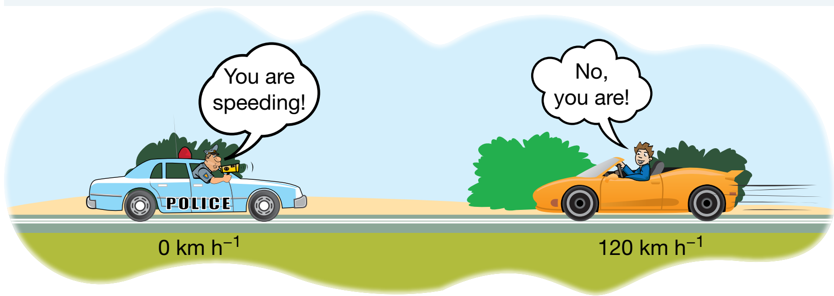
11.2.3 The principle of relativity

Consider the driver of the sports car discussed earlier. His position relative to features of the landscape he drives through is continuously changing, but inside the car life goes on as normal. He has the same position, weight, mass and height; everything inside the car behaves just as he remembers it from when the car was parked. On a smooth road at constant speed, his passenger could pour a drink without difficulty. The effect of the bumps in the road would be indistinguishable from a situation in which the car was stationary and someone outside was rocking it.

Nothing inside a vehicle moving with constant velocity can be affected by the magnitude of the velocity. If it was, we would need to ask: which velocity? If a velocity of 90 km h^{-1} caused a passenger to have a mass of 50 kg, but a velocity of 150 km h^{-1} caused the passenger to have a mass of 60 kg, we would have a problem. The driver cannot simultaneously observe his passenger to have two different masses.

The principle of relativity is the name that physicists give to this realisation. This states that the laws of physics do not depend on the velocity of the observer. Galileo played a major role in the development of the principle of relativity, and Newton's laws of motion are fully consistent with it. Another way of describing the principle of relativity is that there is no way that anyone in the car can measure its velocity without making reference to something external to the car. The sports car driver can measure his speed relative to the two police officers mentioned above. He would measure that he is moving relative to each of them at different speeds, but he would not feel any difference. As long as the road is straight and smooth and the car is travelling at a constant speed, there is no way to detect that the car is moving at all! He could be stationary while one police car is approaching him at 90 km h^{-1} and the other at 150 km h^{-1}.

FIGURE 11.8 How can we tell who is actually speeding?



Even on an aeroplane travelling smoothly at 700 km h^{-1}, we feel essentially the same as we do at rest. The only giveaway is the turbulence the aircraft experiences and the change in air pressure in our ears. Neither of these effects is dependent on the forward velocity of the plane. The laws of physics are the same: you can pour your can of drink safely, walk down the aisle, and drop a pencil and notice it fall vertically to the floor just as it would if you were on the ground.

By introducing the principle of relativity, Galileo provided the necessary framework for important developments in physics. Physics builds on the premise that the universe follows some order that can be expressed as a set of physical laws. The Aristotelian ideas that were held at the time of Galileo suggested that a force is necessary to keep objects moving. This led to one of the major arguments against Earth's motion: everyone would be hurled off the Earth's surface as it hurtled through space, and the Moon would be left behind rather than remaining in orbit around Earth. Galileo's physics, including the principle of relativity, helped to explain why this argument was wrong. Forces are not required to keep objects moving, only to change their motion.

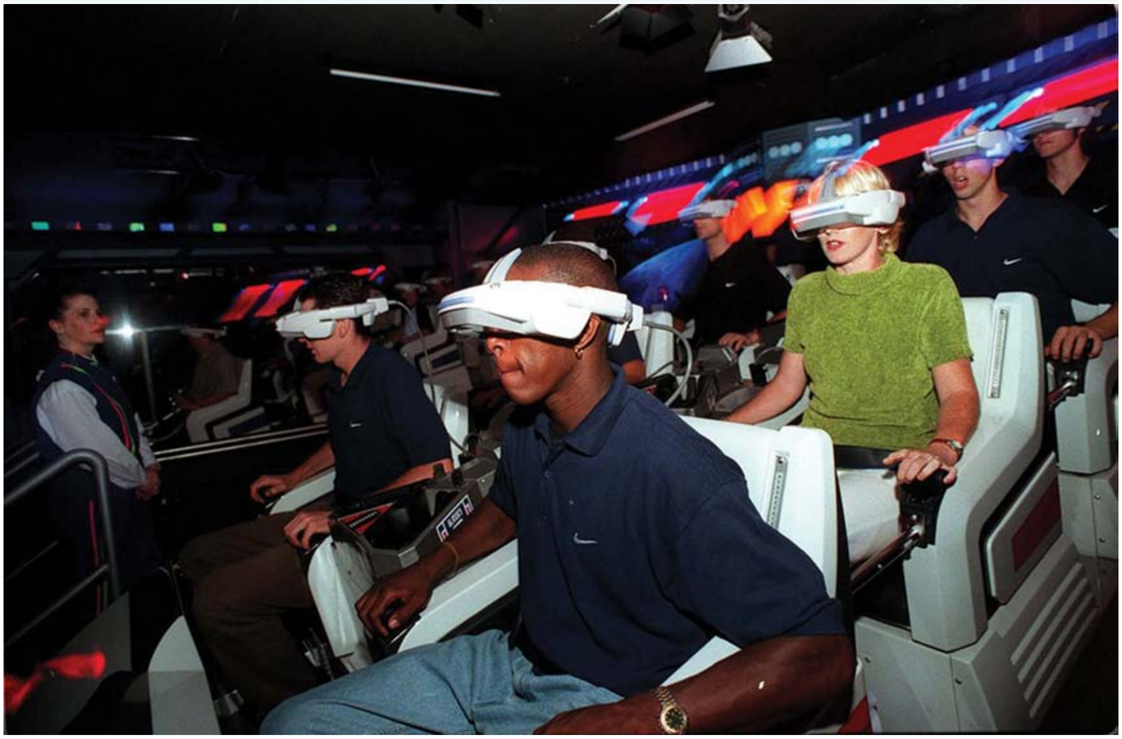
The science of Galileo and Newton was spectacularly successful: it explained the motion of everything from cannon balls to planets. Later, however, as new theories of physics developed in the nineteenth century, physicists faced the challenge of how to make everything fit together. It was not until the early twentieth century that Einstein found a way to make sense of it all.

11.2.4 Examples of Galilean relativity

Here are some examples that support the Galilean principle of relativity.

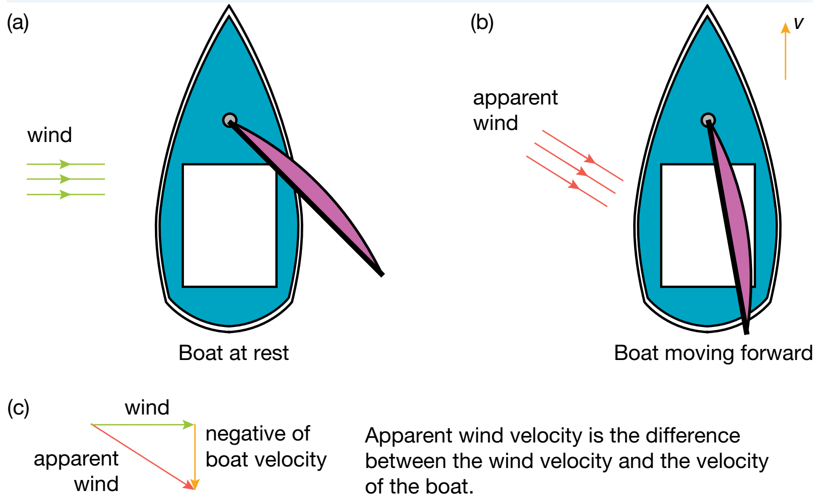
1. If you are in a car stopped at the lights and another car next to yours slowly rolls past, it is difficult to tell whether you or the other car is moving if nothing but the other car is in view.
2. In IMAX and similar films, viewers can feel as though they are going on a thrilling ride, even though they are actually sitting on a fixed seat in a cinema. Theme parks enhance this effect in virtual reality rides by jolting the chairs in a way that mimics movements you would feel on a real ride. Virtual reality rides are very convincing because what you see and feel corresponds with an expected movement, and your senses do not tell you otherwise. As long as the jolts correspond with the visual effects, there is no way of telling the difference. The motion or lack of motion of the seat is irrelevant.

FIGURE 11.9 A virtual reality ride.



3. Acceleration does not depend on the velocity of the observer. An astronaut in a spacecraft travelling through deep space with constant velocity feels weightless, regardless of the magnitude of the velocity. She moves along with the same velocity as the spacecraft, as Newton's first law would suggest. When the spacecraft accelerates due to the force of its rocket engines, the astronaut feels pushed against the back wall of the spacecraft by a force that depends on the magnitude of the acceleration. The effect of the acceleration on the astronaut is noticeable, and may even cause the astronaut to lose consciousness if it is too great.
4. When you are riding in a car with the window down, most of the wind you feel on your face is due to the motion of the car through the air. It is present even on a still day. Only very severe winds exceed $60\,\text{km}\,\text{h}^{-1}$; whenever you drive at greater than $60\,\text{km}\,\text{h}^{-1}$, your windscreen is saving you from gale-force winds! Similarly, it is always windy on moving boats. This is because on deck you are not as well protected from the apparent wind as you are in a car.
5. Apparent wind becomes especially significant when sailing. As the boat increases its speed, the sailor notices that he is heading more into the wind, even though neither he nor the wind has changed direction relative to the shore. This leads the sailor to change the sail setting to suit the new wind direction.

FIGURE 11.10 The faster the boat moves, the more the wind appears to blow from in front.



11.2 SAMPLE PROBLEM 1

Compare the following two scenarios in terms of velocity.

1. A car travelling down the highway at 80 km h^{-1} collides with a stationary car.
2. A car travelling down the highway at 100 km h^{-1} collides with a car travelling at 20 km h^{-1} in the same direction.

SOLUTION:

In the first scenario, the first car is travelling at 80 km h^{-1} relative to the second car.

In the second scenario, the first car is travelling at $100 - 20 = 80$ km h^{-1} relative to the second car. Although the speeds relative to the road in each case are different, the relative speeds of the cars are the same and will cause similar effects on collision.

11.2 Exercise 1

1 The key to Galilean relativity is that:
   (A) **acceleration**
   (B) **velocity**
   (C) **time**
   (D) **mass**
   is relative.

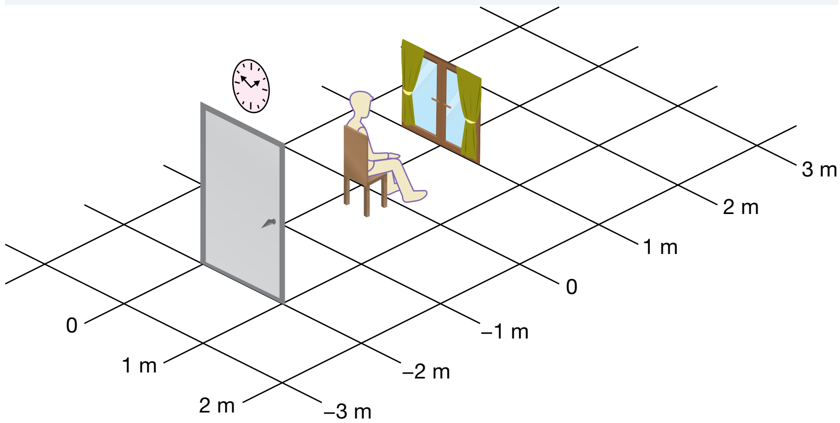
2 Two swimmers are training in a pool. One is 10 metres behind the other and they are both swimming at 2.0 m s^{-1} relative to the end of the pool. What are their velocities relative to each other when:
   (a) they are swimming in the same direction
   (b) they are swimming in opposite directions?

3 Two cars are travelling along the freeway, one at 100 km h^{-1} and the other at 95 km h^{-1}. They are travelling for 10 km.
   (a) What is the velocity of the faster one relative to the slower one?
   (b) How much longer will the slower car take to reach the destination?

11.2.5 Frames of reference

To help make sense of all the possible velocities, physicists consider frames of reference. A frame of reference involves a system of coordinates. For example, where you are sitting reading this book, you view the world through your frame of reference. You can map the position of things around you by choosing an origin (probably the point where you are), then noting where everything else is in reference to that: the window is one metre in front of you, the door is two metres behind you, and so on. Your reference frame also includes time, so you can see that the position of the window in front of you is not changing and you can therefore say its velocity is zero.

FIGURE 11.11 A reference frame is a set of space and time coordinates that are stationary relative to an observer.



When we say something is 'at rest', we mean it is at rest in the reference frame in which we view the world. In everyday life we have a tendency to take a somewhat Aristotelian point of view and regard everything from the perspective that the Earth is at rest. For example, another student walking behind you has her own reference frame. As she walks, your position in her frame of reference is moving. However, she would probably say that she is moving past you while you are stationary, rather than saying that she is stationary while you and the rest of the room are on the move!

In many situations, considering the Earth to be at rest is a convenient assumption. In more complex examples of motion, such as sports events, car accidents involving two moving vehicles, or the motions of the solar system, it can be useful to choose alternative frames of reference.

In classical physics, the differences between frames of reference are their motion and position. ('Classical physics', simply put, is the physics that predated Einstein's discoveries leading to the laws of relativity and quantum mechanics.) In other words, position and speed are relative in classical physics. For example, I might record an object to have a different position than you would (it might be 3 metres in front of me but 4 metres behind you), and I might also record it as having a different speed (maybe it is stationary in my frame of reference but approaching you at $2\,\text{m}\,\text{s}^{-1}$). The position and speed are dependent on the observer. However, in classical physics all observers can agree on what 3 metres and $2\,\text{m}\,\text{s}^{-1}$ are. The rulers in my frame of reference are the same as the ones I see in yours, and the clocks in my frame of reference tick at the same rate as I measure those ticking in yours. Time and space are seen as absolute in the classical physics established by Galileo, Newton and the other early physicists.

Frames of reference that are not accelerating are called **inertial reference frames**. An inertial reference frame moves in a straight line at a constant speed relative to other inertial reference frames.

11.2 SAMPLE PROBLEM 2

Consider the reference frame in which a spacecraft is initially at rest (reference frame A). Astronaut Axel is in the spacecraft and he fires its rockets for 10 s, achieving a final velocity of $100\,\text{m}\,\text{s}^{-1}$. Show that the acceleration of the rocket does not depend on the reference frame.

SOLUTION:

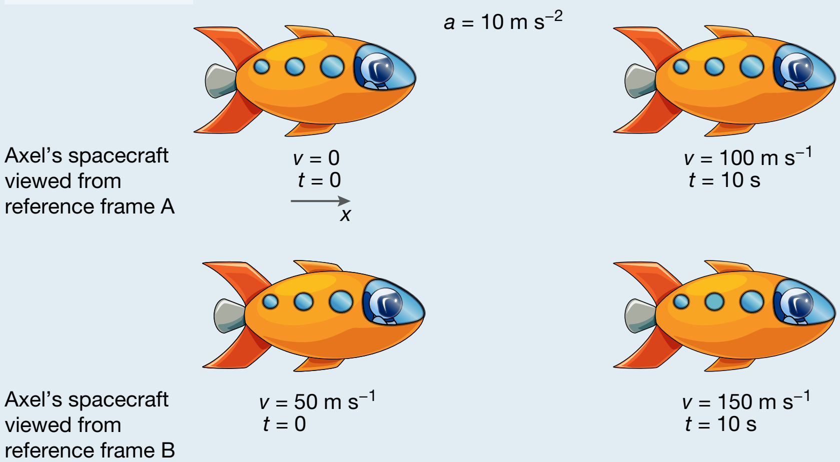
We will show this by determining what the acceleration of the spacecraft is in reference frame A and randomly choosing another inertial reference frame, B, to see if the acceleration is the same.

According to the measurements made in A, the rocket accelerated for 10 s at:

$$\begin{aligned} a &= \frac{\Delta v}{t} \\ &= \frac{100 \text{ m s}^{-1} - 0 \text{ m s}^{-1}}{10 \text{ s}} \\ &= 10 \text{ m s}^{-2} \end{aligned}$$

Axel would feel a force towards the rear of the spacecraft similar in magnitude to his weight on Earth.

FIGURE 11.12



Now we choose a different reference frame. Effie is in reference frame B in another spacecraft, moving at 50 m s^{-1} relative to A. She also measures the acceleration of Axel's spacecraft from her reference frame. Effie measures the velocity of Axel's spacecraft to change from 50 m s^{-1} to $(50 + 100)$ m s^{-1} in 10 s. From B:

$$\begin{aligned} a &= \frac{\Delta v}{t} \\ &= \frac{150 \text{ m s}^{-1} - 50 \text{ m s}^{-1}}{10 \text{ s}} \\ &= 10 \text{ m s}^{-2} \end{aligned}$$

The acceleration is the same whether it is measured from frame A or frame B. We observe that it will still be 10 m s^{-2} regardless of the speed of the reference frame.

An **invariant** quantity is a quantity that has the same value in all reference frames. In classical physics, mass is the same in all reference frames, so all observers will observe that Newton's second law holds. In sample problem 2, all observers would agree on the forces acting on the astronauts. Unlike velocity, acceleration in Galilean relativity does not depend on the motion of the frame of reference; it is also invariant.

It is interesting to consider the motion of Axel's spacecraft as viewed by Effie in reference frame B. Reference frame B is in an inertial reference frame as it is not accelerating. Axel, however, looks back at Effie and sees her falling behind at an increasing rate. Is it Axel or Effie that is accelerating? The answer is clear to them: the force experienced by Axel is not felt by Effie. The acceleration can be measured by this force without any reference to the relative motions of other objects; an object's velocity cannot.

11.2 Exercise 2

1 Explain what is meant by the statement 'speed is relative to the frame of reference'.
2 By referring to Newton's laws of motion, explain why it is important for acceleration to be invariant, but velocity can be relative.
3 Explain why the principle of relativity is so important to physics.
4 A car accelerates from zero to 10 m s^{-1} in 2.0 s according to a stationary observer beside the road. Show that an observer in a car passing at a constant speed of 10 m s^{-1} also observes that car to accelerate at the same rate as the stationary observer.

eBook*plus* RESOURCES

Watch this eLesson: Relative velocity
Searchlight ID: med-2918

11.3 Electromagnetism brings new challenges

11.3.1 Light as an electromagnetic wave

Galilean relativity seemed to work well for the motion of massive bodies, but by the nineteenth century physicists were learning much more about other physical phenomena.

As described in Topic 8, James Clerk Maxwell's theory of electromagnetism drew together the key findings of electricity and magnetism to completely describe the behaviour of electric and magnetic fields in a set of four equations. One of the outcomes of this was an understanding of electromagnetic waves. The equations dictated the speed of these waves, and Maxwell noticed that the speed was the same as what had been measured for light. He suggested that light was an electromagnetic wave and predicted the existence of waves with other wavelengths that were soon discovered, such as radio waves. A medium for these fields and waves was proposed, called the luminiferous aether. The speed of light, c, was the speed of light relative to this aether.

Understanding electromagnetic phenomena was the foundation for Einstein's special relativity. In particular, the physicists of the nineteenth century, such as Michael Faraday, knew that they could induce a current in a wire by moving a magnet near the wire. They also knew that if they moved a wire through a magnetic field, a current would be induced in the wire. They saw these as two separate phenomena.

Imagine this: two students are in different physics classes. Annabel has learned in her class that electrons moving in a magnetic field experience a force perpendicular to their direction of motion and in proportion to their speed. Her friend Nicky has learned in her class that a current is induced in a loop of wire when the magnetic flux through the wire changes. Are these two different phenomena? Because they have also learned about the principle of relativity, Annabel and Nicky have doubts. They get together after class to perform experiments. The force depends on the speed. Annabel holds a stationary loop of conducting wire. Nicky moves the north pole of a magnet towards the loop, and they notice that a current is present in the wire as she does this. Nicky says that this is consistent with what she has learned. The conclusion is that a current is induced by a changing magnetic field. Then Nicky holds the magnet still so that the magnetic field is not changing. Annabel moves her loop of wire towards Nicky's magnet. Annabel states that the result agrees with what she learned in class — that electrons and other charged particles experience a force when moving in a magnetic field.

Einstein realised that there was only one phenomenon at work here. Both experiments are doing exactly the same thing, and it is only the relative speeds of the coil and the magnet that are important. This may seem obvious, but to make this jump it was necessary to discard the idea that the electric and magnetic fields depended on the luminiferous aether. It was the relative motion that was important, not whether the magnet or charge was moving through the aether.

Before Einstein's realisation, the understanding was that if light moves through the aether, then the Earth must also be moving through the aether. Changes in the speed of light as the Earth orbits the Sun should be detectable. Maxwell predicted that electromagnetic waves would behave like sound and water waves, in that the speed of electromagnetic waves in the medium would not depend on the motion of the source or the detector through the medium.

To understand the significance of this aether, consider the sound produced by a jet plane. When the plane is stationary on the runway preparing for takeoff, the sound travels away from the plane at the speed of sound in air, about 340 m s^{-1}. When the plane is flying at a constant speed, say 200 m s^{-1}, the speed of sound is still 340 m s^{-1} in the air. However, to find the speed relative to the reference frame of the plane, we must subtract the speed of the plane relative to the air. From this we find that the sound is travelling at:

$340 - 200 = 140$ m s^{-1} in the forward direction relative to the plane

$340 - -200 = 540$ m s^{-1} in the backward direction relative to the plane.

FIGURE 11.13 An experiment in electromagnetism.

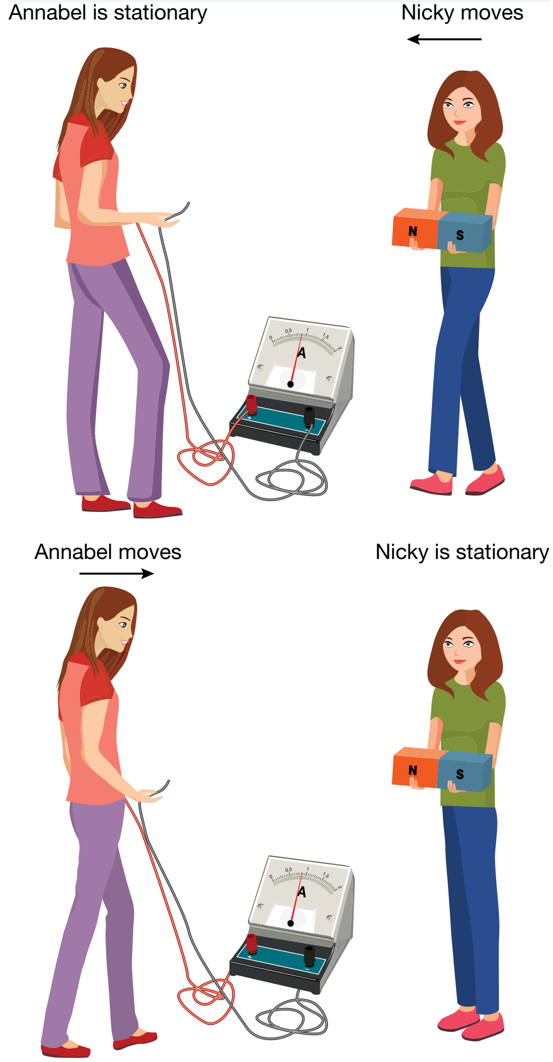
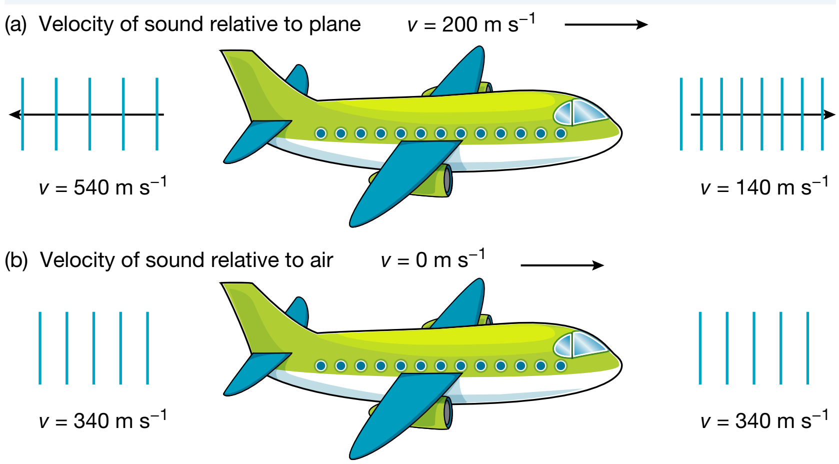


FIGURE 11.14 Sound moving away from a plane.



In this example we could measure the speed of the plane through the air by knowing the speed of sound in air (340 m s^{-1}) and measuring the speed of a sound sent from the back of the plane to the front (140 m s^{-1}) in the reference frame of the plane. As long as the plane is flying straight, we could infer the speed of the plane relative to the air by setting the forward direction as positive and subtracting the velocities:

$$340 - 140 = 200\,\mathrm{m\,s^{-1}}.$$

The speed of the plane has been measured relative to an external reference frame, that of the air, and therefore this example has not violated Galilean relativity. As light had been shown to travel in waves, scientists felt they should be able to measure Earth's speed through the aether in the same way.

11.3 SAMPLE PROBLEM 1

Explain how Maxwell's concept of electromagnetic waves such as light challenged the Galilean principle of relativity.

SOLUTION:

The principle of relativity states that the laws of physics hold true in all inertial reference frames. Maxwell predicted that the speed of light was constant relative to the aether but would be different in other frames of reference. The result was that different explanations were required for electromagnetic phenomena depending on whether the magnets or charges were in motion through the aether.

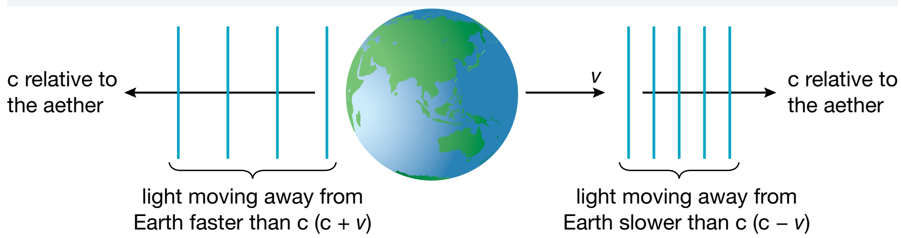
11.3 Exercise 1

1 Assuming that electromagnetic waves travel at c relative to the aether, determine the speed of light shining from the rear of a spacecraft moving at half the speed of light relative to the aether according to Kirsten, who is onboard the spacecraft.

11.3.2 The Michelson–Morley experiment

In 1887, Albert Michelson and Edward Morley devised a method of using interference effects to detect slight changes in time taken for light to travel through different paths in their apparatus. As with sound travelling from the front and rear of a plane through the air, the light was expected to take different amounts of time to travel in different directions through the luminiferous aether as the Earth moved through it. Much to their astonishment, the predicted change in the interference pattern was not observed. It was as though the speed of light was unaffected by the motion of the reference frame of its observer or its source!

FIGURE 11.15 The idea behind the Michelson–Morley experiment.



11.3.3 Einstein's two postulates of special relativity

Physicists tried all sorts of experiments to detect the motion of Earth through the luminiferous aether, and they attempted to interpret the data in ways that would match the behaviour of light with what they expected would happen. Their attempts were unsuccessful.

Einstein managed to restore order to our understanding of the universe. While others suspected the new theory of electromagnetism to be wrong, Einstein took apart the established theory of Newtonian mechanics, even though its success had given physicists reason to believe in relativity in the first place. Einstein dared to see what would happen if he embraced the results of the Maxwell equations and the experiments with light, and accepted that the speed of light was invariant. The results were surprising and shocking, but this bold insight helped usher in the modern understanding of physics.

Einstein agreed with Galileo that the laws of physics must be the same for all observers, but he added a second requirement: that the speed of light in a vacuum is the same for all observers. The speed of light is not relative, as had been expected by those who went before him, but invariant. He set these two principles down as requirements for the development of theoretical physics. They are known as Einstein's two postulates of special relativity:

1. The speed of light in a vacuum is an absolute constant.
2. All inertial frames of reference are equivalent.

The physics based on these postulates has become known as special relativity. It is 'special' because it deals with the special case where there is no gravity. To deal with gravity, Einstein went on to formulate his theory of general relativity, but that is beyond the scope of this course.

Einstein's postulates were radical. The consequence of his insistence that physics be based on these two postulates was that ideas that had been taken for granted for centuries were thrown out. As well as the removal of the luminiferous aether, the intuitive notions that time passed at the same rate for everyone, that two simultaneous events would be simultaneous for all observers, and that distance and mass are the same for all observers had to be discarded.

Einstein's work explained why the velocity of Earth could not be detected. The second postulate implied that there is no experiment that can be done on Earth to measure the speed of Earth. We must take an external reference point and measure the speed of Earth relative to that point in order for the speed of Earth to have any meaning. With his first postulate, Einstein also declared that it does not matter which direction the Michelson–Morley apparatus was pointing in; the light would still travel at the same speed. No change in the interference pattern should be detected when the apparatus was rotated.

11.3 SAMPLE PROBLEM 2

How do Einstein's postulates differ from the physics that preceded him?

SOLUTION:

Firstly, the principle of relativity is applied to all laws of physics, not just the mechanics of Galileo and Newton.

Secondly, the speed of light is constant for all observers. Before Einstein, the speed of light was assumed to be relative to its medium, the luminiferous aether.

11.3 Exercise 2

1 Einstein realised that something that had been regarded as relative was actually invariant. As a result of this, quantities that had been regarded as invariant now had to be regarded as relative. What did he find to be invariant and what relative?

2 Einstein's postulates stated that the speed of light is the same for all observers. How was this different from the understanding of those who preceded him?

11.3.4 Broadening our horizons

Why did scientists before Einstein (and most of us after Einstein) not notice the effects of light speed being invariant? Newton's laws provided a very good approximation for the world experienced by people before the twentieth century. By the beginning of the twentieth century, however, physicists were able to take measurements with much greater precision. They were also discovering new particles, such as electrons, that could travel at incredible speeds. Indeed, these speeds were completely outside the realm of human experience. Light travels at $c = 3 \times 10^8\ \mathrm{m\,s^{-1}}$ or 300 000 km per second. (To be precise, $c = 299\,792\,458\ \mathrm{m\,s^{-1}}$.) At this speed, light covers the distance to the Moon in roughly 1.3 seconds!

11.3 SAMPLE PROBLEM 3

To get a sense of how fast light travels, Andrei considers how long it would take to accelerate from rest to a tenth of this great speed at the familiar rate of $9.8\ \mathrm{m\,s^{-2}}$ — the acceleration of an object in free fall near the surface of Earth.

SOLUTION:

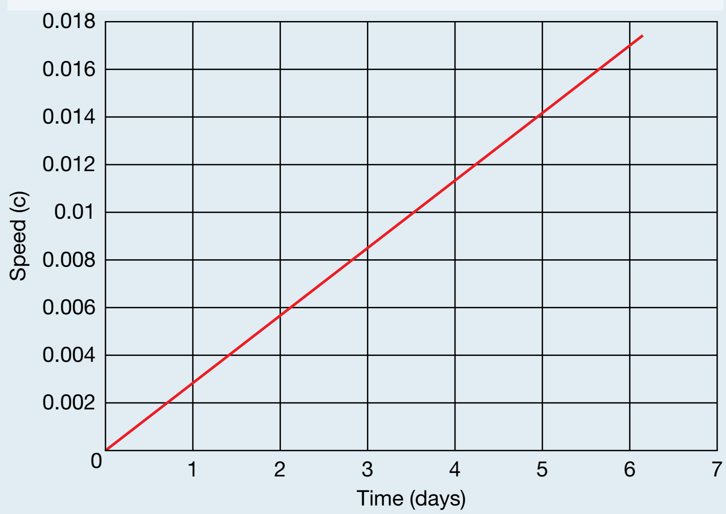
$u = 0\ \mathrm{m\,s^{-1}}$, $v = 0.1\,c = 3 \times 10^7\ \mathrm{m\,s^{-1}}$, $a = 9.8\ \mathrm{m\,s^{-2}}$, $t = ?$

$$v = u + at$$

$$\begin{aligned} t &= \frac{v-u}{a} \\ &= \frac{3 \times 10^7}{9.8} \\ &= 3.06 \times 10^6 \text{ seconds} \\ &= 35.4 \text{ days} \end{aligned}$$

It would take more than 35 days to achieve a speed of 0.1 c! (This is the fastest speed for which use of Newtonian kinematics still gives a reasonable approximation.)

FIGURE 11.16 This graph shows how speed as a fraction of c increases over time at an acceleration of $9.8\ \mathrm{m\,s^{-2}}$.



11.3 Exercise 3

1. With an acceleration of $9.8\ \mathrm{m\,s^{-2}}$, occupants of a spacecraft in deep space would reassuringly feel the same weight they feel on Earth. What would happen to the astronauts if the acceleration of the spacecraft was much greater to enable faster space travel?
2. How far would a spacecraft accelerating at $9.8\ \mathrm{m\,s^{-2}}$ travel in 35.4 days if it started from rest?

Light speed really is beyond our normal experience! Maybe Einstein's predictions would not be so surprising if we had more direct experience of objects travelling at great speeds, but as it is they seem very strange.

AS A MATTER OF FACT

The distance light travels in a year is known as a light-year. Even on Earth, we now measure distance in terms of the speed of light. One metre is defined as the distance light travels in exactly $\frac{1}{c} = \frac{1}{299\,792\,458}$ of a second.

eBook*plus* RESOURCES

Watch this eLesson: Michelson–Morley experiment
Searchlight ID: eles-2561

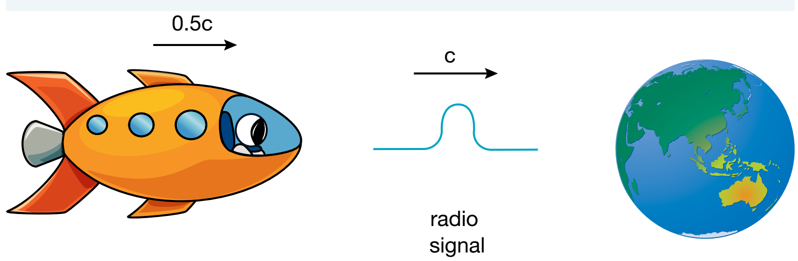
11.4 Understanding the speed of light as an absolute constant

11.4.1 The speed of light is constant

This simple statement of Einstein's may not seem remarkable. To highlight what it means, we will again compare light with sound. In the nineteenth century, sound and light were thought to have a lot in common, because they both exhibited similar wavelike behaviours, such as diffraction and interference. However, sound is a disturbance of a medium, whereas light does not require any medium at all. Sound has a speed that is relative to its medium. If the source of the sound is moving through the medium, the speed of the sound relative to the source is different to the speed of sound relative to the medium. Its speed can be different again from the reference frame of the observer.

Einstein was saying there is no medium for light, so the concept of the speed of light relative to its medium is not meaningful. Light always moves away from its source at 299 792 458 m s^{-1} and always meets its observer at 299 792 458 m s^{-1}, no matter what the relative speeds of the observer and the source. Even if the Earth were hurtling along its orbit at 0.9 c, the result of the Michelson–Morley experiment would have been the same.

FIGURE 11.17 A spacecraft approaching Earth at 0.5c. The radio signal is travelling at c relative to both Earth and the spacecraft!



As an example, consider a spacecraft in the distant future hurtling towards Earth at 0.5 c. The astronaut sends out a radio message to alert Earth of his impending visit. (Radio waves, as part of the electromagnetic spectrum, have the same speed as visible light.) He notices that, in agreement with the Michelson–Morley measurements of centuries before, the radio waves move away from the spacecraft at c. With what speed do they hit the Earth? Relative velocity, as treated by Galileo, insists that, as the spacecraft already has a speed of 0.5 c relative to the Earth, the radio waves must strike the Earth at 1.5 c. However, this does not happen. The radio waves travel at c regardless of the motion of the source and the receiver.

This concept was very difficult for physicists to deal with, and many resisted Einstein's ideas. But the evidence is irrefutable. Newtonian physics works as a very good approximation only for velocities much less than c. The faster something moves, the more obvious it is that the Newtonian world view does not match reality. It was not until the twentieth century that scientists dealt with objects (such as cosmic rays) moving at great speeds. Satellites in orbit need to be programmed to follow Einstein, rather than Newton, if they are to provide accurate data.

11.4.2 Space–time diagrams

In 1908, Hermann Minkowski invented a useful method of depicting situations similar to the spacecraft scenario described above. His diagrams are like distance–time graphs with the axes switched around. However, they differ from time–distance graphs in an important way. When reading these diagrams, the markings on the scales for time and position are only correct for the reference frame in which the axes are stationary.

11.4 SAMPLE PROBLEM 1

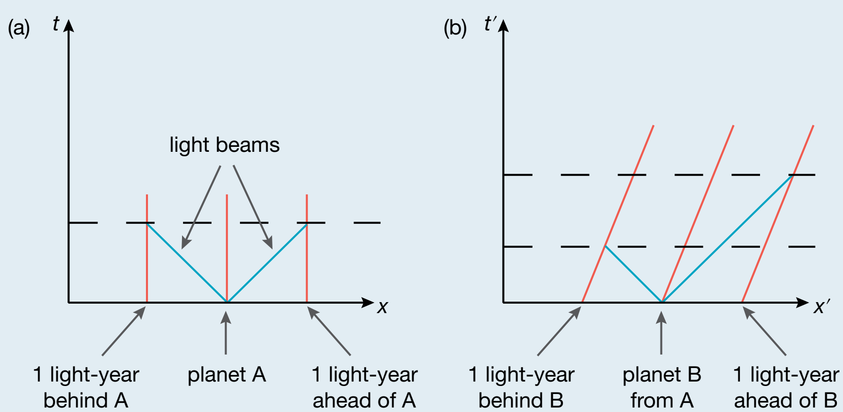
Light reflecting off planet A radiates in all directions at c so that, after one year, the light that left the planet forms a circle one light-year in radius. Another planet, B, passes planet A at great speed, just missing it. Light from B's surface also leaves at c, according to the second postulate, forming a circle around it. How can both planets be at the centre of their light circles as the postulates demand?

SOLUTION:

Draw Minkowski diagrams for each planet. Diagram (a) shows the situation for planets A and B from the reference frame of the planet, with the planet at the centre — the labels refer to planet A, but the diagram is the same for both planets. The light radiates in all directions at the same rate, and the diagram shows where the light in one direction and the opposite direction would be after one year.

Diagram (b) shows what is happening on B according to observers on A. The light moving out behind the moving planet reaches the one-light-year distance sooner than the light moving out from the front! But we know that planet B is at the centre of this light circle. The way to achieve this is to move away from absolute space and time and understand that these are relative to the observer. When we do this, we see that it is possible for planet B to be at the centre of the light circle. However, this requires that A and B disagree about when two events occur. According to planet A, the different sides of the light circle reach the light-year radius at different times, but from planet B this must occur simultaneously.

FIGURE 11.18 Events that are simultaneous in one reference frame are not simultaneous in another.



11.4 Exercise 1

1 State whether the simultaneity of events is invariant or relative in:
(a) classical physics
(b) special relativity.

11.5 The evidence for Einstein's two postulates

11.5.1 Summarising the evidence

Einstein's two postulates of special relativity are justified in many ways by the implications that will be outlined in the remainder of this topic. However, from what we have seen already, there are several reasons Einstein might have seen them as worth exploring.

Postulate 1: The speed of light in a vacuum is an absolute constant.

The evidence for this includes the fact that no motion of the Earth through a luminiferous aether was determined by the Michelson–Morley experiment although it was sensitive enough to detect the expected motion. Improved experiments since that time have continued to show the same result. Also, the effects that come from this postulate are generally only noticed at extremely high speeds, speeds humans do not experience in everyday life. If you consider, for example, the space–time diagrams for the two planets A and B in Figure 11.18, the effect demonstrated is due to the incredible speed of planet B relative to planet A. If planet B was moving at the speed of the Earth around the Sun (about 11 000 m s^{-1}), the angle of the red lines in diagram (b) would be imperceptibly different from those in diagram (a). This postulate, while it produces some strange outcomes, does agree with the world of our experience.

Postulate 2: All inertial frames of reference are equivalent.

This postulate was largely understood back in Galileo's time. As evidence, he explained the experience of being below deck on a ship while racing through calm seas. Fish swam in their bowls, flies buzzed around the room and coins dropped to the floor in the same way as if the ship was at rest at a dock. We see this today in more extreme cases of high-speed trains, planes and spacecraft where fast travel at constant velocity is experienced much the same as at rest in the station or on the ground. The motion of the Earth around the Sun and around the galaxy at vast velocities (with acceleration too small to notice) has no impact on our experience of physical effects. It also agrees with the observations of Earth-based experiments involving electromagnetism where equal currents are induced by a charge in a changing magnetic field, as by charges moving into a constant magnetic field. The reference frames are equivalent; only the relative motion is important.

The passing of time can be measured in many ways, including using the position of the Sun in the sky, the position of hands on a watch, the changing of the seasons, and the signs of a person ageing. Galileo is known to have made use of the beat of his pulse, the swinging of a pendulum and the dripping of water. As already stated, Newtonian physics assumed that each of these clocks ticked at the same rate regardless of who was observing them. However, the theory of relativity shows that this assumption that time is absolute is actually wrong. This error becomes apparent when the motion of the clock relative to the observer approaches the speed of light.

Consider a simple clock consisting of two mirrors, A and B, with light reflecting back and forth between them. This is an unusual clock, but it is very useful for illustrating how time is affected by relativity. Experiments that involve pursuing an idea on paper without actually performing the experiment are common in explanations of relativity. They are known as **thought experiments**.

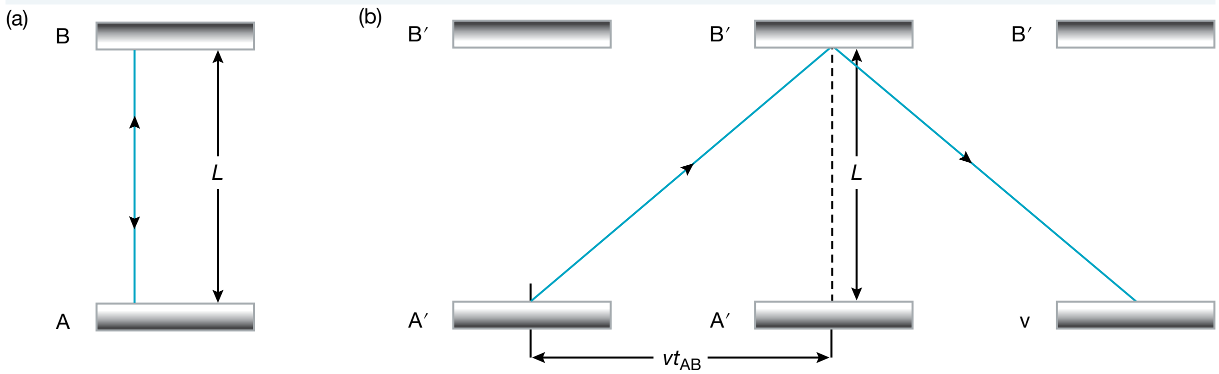
Let the separation of the mirrors be L. The time for the pulse of light to pass from mirror A to mirror B and back is calculated in the conventional way:

$$c = \frac{2L}{t_0}$$

$$t_0 = \frac{2L}{c}$$

where t_0 is the time for light to travel from A to B and back, as measured in the frame of reference in which the clock is at rest. We will define this time, t_0, to be one tick of the clock. In this case, the position of the clock does not change in the frame of reference. The passing of time can be indicated by two events separated by time but not by space — the event of the photon of light first being at A and the event of the photon being back at A.

FIGURE 11.19 A light clock (a) at rest relative to the observer, and (b) in motion relative to the observer.



Imagine an identical clock, with mirrors A′ and B′, moving past this light clock at speed v. At what rate does time pass on this moving clock according to the observer? Label the time interval measured by this clock t to distinguish it from t_0. The light leaves A′ and moves towards B′ at speed c. The speed is still c even though the clock is moving, as stated by Einstein's second postulate. In the time the light makes this journey, the clock moves a distance $d = vt_{AB}$, where t_{AB} is the time the light takes to travel from A′to B′. Diagram (b) depicts this situation and shows that the light in the moving frame of reference has further to travel than the light in the rest frame. Using Pythagoras's theorem, the light has travelled a distance of $2\sqrt{L^2 + (vt_{AB})^2}$ from A′ to B′ and back to A′. This is a greater distance than $2L$, given $v \neq 0$ and c is constant. Therefore, the time the light takes to complete the tick must be greater than for the rest clock.

The speed of the light relative to the observer is:

$$c = \frac{d}{t}$$

$$c = \frac{2\sqrt{L^2 + (vt_{AB})^2}}{2t_{AB}}$$

Transpose the equation to make a formula for t:

$$2ct_{AB} = 2\sqrt{L^2 + v^2\,(t_{AB})^2}$$

$$c^2\,(t_{AB})^2 = L^2 + v^2\,(t_{AB})^2$$

But $t_{\mathrm{AB}} = \dfrac{t}{2}$.

$$\frac{c^2t^2}{4} - \frac{v^2t^2}{4} = L^2$$
$$t^2\left(c^2 - v^2\right) = 4L^2$$
$$t = \frac{2L}{\sqrt{c^2 - v^2}}$$
$$= \frac{2L}{c\sqrt{1 - \frac{v^2}{c^2}}}$$

We have already determined that $t_0 = \dfrac{2L}{c}$, so

$$t = \frac{t_0}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

The expression $\dfrac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$ appears frequently in special relativity. So that we do not have to write it all the time, it is simply called gamma, γ. It is also known as the Lorentz factor.

We can now write the equation as $t = t_0\gamma$.

The equation $t = t_0\gamma$ is the known as the **time dilation** formula. This formula enables us to determine the time interval between two events in a reference frame moving relative to an observer.

Note that gamma is always greater than 1. As a result, t will always be greater than t_0, hence the term 'time dilation'. In a reference frame moving relative to the observer like this, the two events that we are using to mark the time interval, the time between the light being at A, occur at different points in space. The time t_0 is the time measured in a frame of reference where the events occur at the same points in space. It is known as the **proper time**. This is not proper in the sense of correct, but in the sense of property. It is the time in the clock's own reference frame, whatever that clock might be.

Examples:

1. A mechanical clock's large hand moves from the 12 to the 3, showing that 15 minutes have passed. Fifteen minutes is the proper time between the two events of the clock showing the hour and the clock showing quarter past the hour. However, if that clock was moving relative to us at great speed, we would notice that the time between these two events was longer than 15 minutes. The time is dilated.
2. A candle burns 2 centimetres in 1 hour. One hour is the proper time between the events of the candle being at a particular length and the candle being 2 centimetres shorter. If the candle was moving relative to the observer, she would notice that it took longer than 1 hour for the candle to burn down 2 centimetres.
3. A man dies at 89 years of age. His life of 89 years is the time between the events of his birth and his death in his reference frame. To an observer moving past at great speed, the man appears to live longer than 89 years. He does not fit any more into his life; everything he does appears to the observer as if it was slowed down.

11.5 SAMPLE PROBLEM 1

James observes a clock held by his friend Mabry moving past at 0.5 c. He notices the hands change from 12 pm to 12.05 pm, indicating that 5 minutes have passed for the clock. How much time has passed for James?

SOLUTION:

The proper time t_0 is the time interval between the two events of the clock showing 12 pm and the clock showing 12.05 pm, which is 5 minutes.

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = 1.155 \text{ when } v = 0.5c.$$

So $t = t_0\gamma = 5 \times 1.155 = 5.775$ minutes.

James notices that the moving clock takes 5.775 minutes (or 5 minutes 46.5 seconds) for its hands to move from 12 pm to 12.05 pm.

11.5 Exercise 1

1 In another measurement, James looks at his own clock and waits the 5 minutes it takes for the clock to change from 1 pm to 1.05 pm. He then looks at Mabry's clock as she moves past at 0.5 c. How much time has passed on her clock?

Unlike in Newtonian physics, time intervals in special relativity are not invariant. Rather, they are relative to the observer.

11.5 SAMPLE PROBLEM 2

Mabry is travelling past James at 0.5 c. She looks at James and sees his clock ticking. How long does she observe it to take for his clock to indicate the passing of 5 minutes?

SOLUTION:

In this case it is James's clock that is showing the proper time. Mabry notices that 5.775 minutes pass when James's clock shows 5 minutes passing. These situations are symmetrical. Mabry sees James as moving at 0.5 c, and James sees Mabry moving at 0.5 c, so her measurement of time passing is the same as his.

11.5 Exercise 2

1 Aixi listens to a 3-minute song on her phone. As soon as she starts the song she sees her friend Xiaobo start wrestling with his brother on a spaceship moving by at 0.8c. When the song finishes, she sees Xiaobo stop wrestling. How long were the two boys wrestling for?

11.5 SAMPLE PROBLEM 3

A car passes Eleanor at $20\,\mathrm{m\,s^{-1}}$. She compares the rate that a clock in the car ticks with the rate the clock in her hand ticks.

SOLUTION:

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = 1.000\,000\,000\,000\,0022 \text{ when } v = 20\,\mathrm{m\,s^{-1}}.$$

The difference between the rates of time in the two perspectives is so small that it is difficult to calculate, much less to notice it.

11.5 Exercise 3

1 Jonathan observes a clock on a passing spaceship to be ticking at half the rate of his identical clock. What is the relative speed of Jonathan and the passing spaceship?

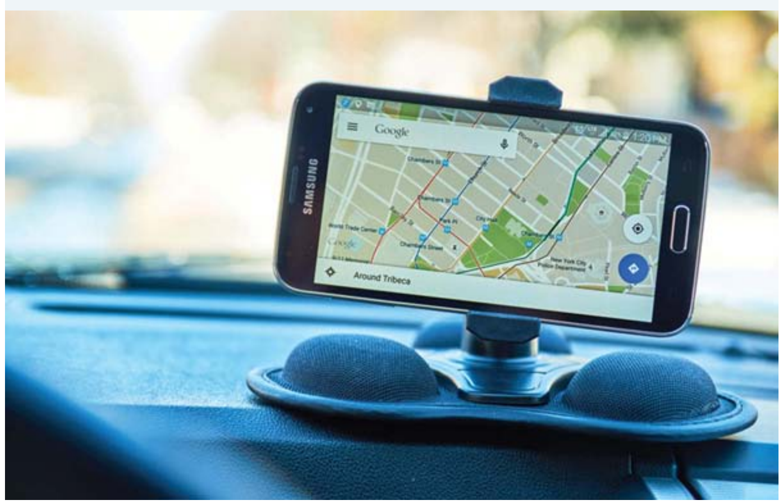
Newton’s assumption that all clocks tick at the same rate, regardless of their inertial reference frame, was very reasonable. Learning the very good approximation of Newton’s laws is well justified. They are simpler than Einstein’s laws, and they work for all but the highest speeds. A good theory in science has to fit the facts, and Newton’s physics fit the data very successfully for 200 years. It was a great theory, but Einstein’s is even better.

If Newton knew then what we know now, he would realise that his theories were in trouble. At speeds humans normally experience, time dilation is negligible, but the dilation increases dramatically as objects approach the speed of light. If you passed a planet at $2.9 \times 10^8\ \mathrm{m\,s^{-1}}$, you would measure the aliens’ usual school lessons of 50 minutes as taking 195 minutes. An increase in speed to $2.99 \times 10^8\ \mathrm{m\,s^{-1}}$ would dilate the period to 613 minutes. If you could achieve the speed of light, the period would last forever — time would stop. Photons do not age, as they do not experience time passing!

11.5.2 Time dilation and modern technology

Time dilation has great practical significance. A global positioning system (GPS) is able to tell you where you are, anywhere on Earth, in terms of longitude, latitude and altitude, to within a few metres. Einstein’s general relativity also shows that the difference in gravity acting on a satellite in orbit affects the time significantly. Nanosecond accuracy is required for a GPS, but if Newtonian physics was used, the timing would be out by more than 30 microseconds. GPSs are widely used in satellite navigation, and ships, planes, car drivers and bushwalkers can find their bearings far more accurately than they ever could using a compass.

FIGURE 11.20 With a GPS device you can know your position to within a few metres.



The Hafele-Keating experiment

Numerous tests of time dilation had been conducted prior to the deployment of the global positioning system. Famously, in 1971 the Hafele-Keating experiment involved flying atomic clocks on commercial airliners to compare the time elapsed with one that remained on the ground. Although the reference frames were not strictly inertial as the clocks were all travelling in circles about the centre of the Earth, and they were under the influence of gravity, the differences in time measured by the clocks was sufficiently large to detect the time dilation effects and they agreed with Einstein's prediction.

PHYSICS FACT

Atomic clocks were used because of their high levels of precise time keeping. For example, one type of clock utilised the regular oscillations between energy states of caesium atoms. The second is defined to be the time taken for 9 192 631 770 oscillations of the caesium-133 atom.

Three sets of clocks were used: one that remained on the ground; one that had a relatively higher speed as it went in planes that travelled around the world in an easterly direction; and one that had a relatively lower speed as it was in planes that circled the globe in a westerly direction. When the effects of gravity and orbital accelerations were taken into account, these very accurate clocks indicated different amounts of time passing, in keeping with the predictions of Einstein's ideas of time dilation.

More recent experiments have tested time dilation predictions to high precision. For example, in 2014 a team of physicists published their work using particle accelerators. They measured the rates of transitions between energy states of atoms moving at one-third the speed of light in the particle accelerator. They were then able to compare these rates with the rates of transitions of the same atoms at rest in the laboratory. The difference in rates matched Einstein's predictions to the highest levels of precision achieved to date.

There are significant implications for cosmology in time dilation. For example, cosmic rays made mostly of protons and alpha particles hurtle across the universe at near light speeds. From the point of view of these particles, very little time passes on their journey, which we would time as lasting millions of years. If we too could travel at such speeds, we would travel on journeys across galaxies in minutes but in which time Earth would have aged millions of years. Also, due to the relative speeds of some objects in the universe, the time it takes for a star to explode, for example, is observed by astronomers to take longer than it would to an observer in its reference frame.

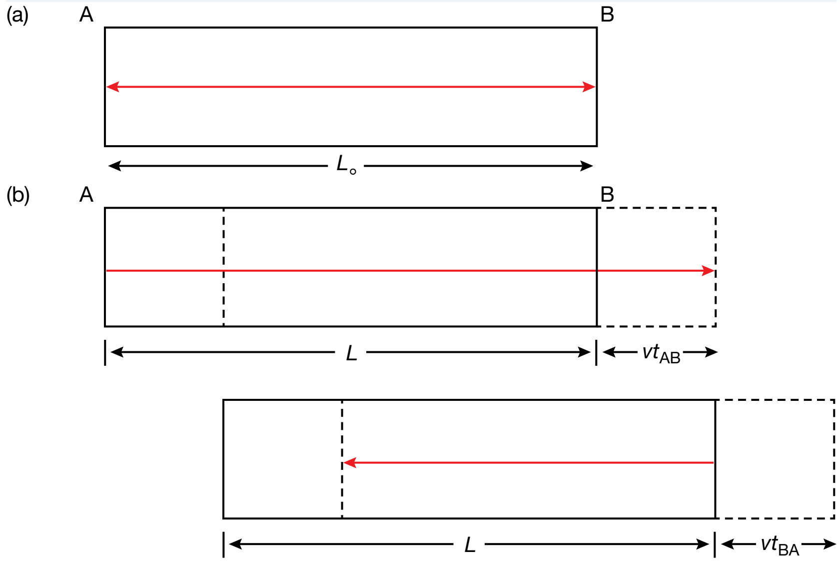
11.6 Length contraction

11.6.1 Measuring length

Once we accept that simultaneity of events and the rate that time passes are relative, we have to accept that length must be relative as well. The length of an object is simply the distance between the two ends of the object. To find that distance, the position of both ends must be noted at the same time. If they were measured at different times, a moving object would have changed position, so the distance between the end that was measured second and the end that was measured first would have changed. The fact that any two inertial reference frames do not agree on which events are simultaneous is going to cause the measurement of length to be different in different reference frames. The speed of light is invariant and time is relative, so we have even more reason to doubt that lengths will be the same for all observers.

A clever thought experiment of Einstein's enables us to determine the effect the speed of an observer has on a length to be measured. It is essentially the same as the thought experiment used to derive the time dilation equation, but with the light clock tipped on its side so that its length is aligned with the direction of its motion.

FIGURE 11.21 Light journeys in (a) a clock at rest and (b) a clock moving to the right at speed *v*.



From the reference frame of the clock, again $t_0 = \frac{2L_o}{c}$. What about the reference frame of an observer with a speed of v relative to the clock? We can measure the distance between the ends of the clock using the time for light to travel from one end to the other and back.

From A to B:

$$L + vt_{AB} = ct_{AB}$$

where
L = the length of the clock as observed by the moving observer
vt_{AB} = the distance the clock has moved in the time the light passes from A to B
ct_{AB} = the distance the light has travelled passing from A to B.

Transposing the equation to make t_{AB} the subject:

$$t_{AB} = \frac{L}{c - v}$$

From B to A:

$$L - vt_{BA} = ct_{BA}$$

where vt_{BA} = the distance the clock has moved in the time the light passes from B back to A
ct_{BA} = the distance the light has travelled passing from B back to A.

Transposing the equation to make t_{BA} the subject:

$$t_{AB} = \frac{L}{c + v}$$

As A moves to meet the light, the time t_{BA} is less than t_{AB}. The total time is:

$$t = t_{AB} + t_{BA}$$

$$= \frac{L}{\mathrm{c} - v} + \frac{L}{\mathrm{c} + v}$$

$$= \frac{2Lc}{\mathrm{c}^2 - v^2}$$

$$= \frac{2L}{\mathrm{c}\left(1 - \frac{v^2}{\mathrm{c}^2}\right)}$$

According to the time dilation formula:

$$t = \frac{t_0}{\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}}$$

Substituting this for our time in the moving clock gives:

$$\frac{t_0}{\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}} = \frac{2L}{\mathrm{c}\left(1 - \frac{v^2}{\mathrm{c}^2}\right)}$$

$$t_0 = \frac{2L}{\mathrm{c}\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}}$$

Substituting $t_0 = \frac{2L_0}{\mathrm{c}}$ gives:

$$\frac{2L_0}{\mathrm{c}} = \frac{2L}{\mathrm{c}\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}}$$

$$\Rightarrow L_0 = \frac{L}{\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}} \quad \text{or} \quad L = \frac{L_0}{\gamma}$$

The formula $L = L_0\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}$ is known as the Lorentz contraction formula after one of the early pioneers of relativity theory, Hendrik Antoon Lorentz (1853–1928). The Lorentz contraction is the shortening of an object in its direction of motion when measured from a reference frame in motion relative to the object.

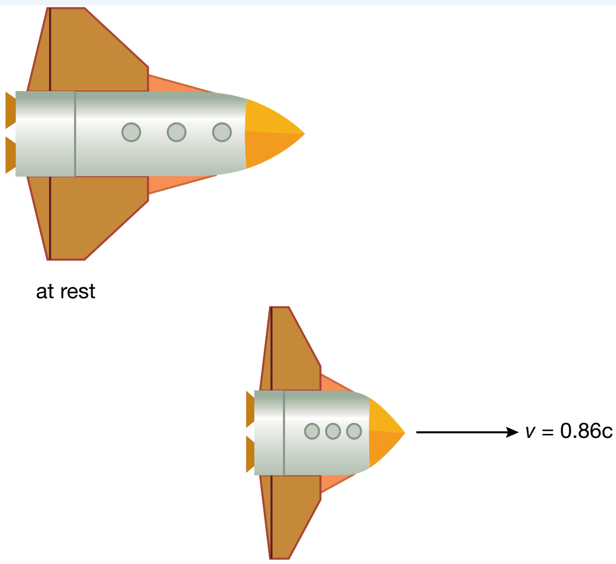
The **proper length** of an object, L_0, is the length measured in the rest frame of the object. L is the length as measured from an inertial reference frame travelling at a velocity v relative to the object. This change in length applies only to the length along the direction of motion. The other dimensions are not affected by this contraction.

AS A MATTER OF FACT

George Fitzgerald and Hendrik Lorentz independently proposed an explanation for the result of the Michelson–Morley experiment (in 1889 and 1892 respectively). If the length of the apparatus contracted in the direction of Earth's movement, the light would take the same time to travel the two paths. This explanation assumed that the aether existed and that light would travel at constant speed through it; therefore, light would travel at different speeds relative to Earth as Earth moved through the aether. This explanation was not completely satisfying as there was no known force that would cause the contraction, and the aether had never been directly detected. The contraction would be measured by those in the reference frame at rest with respect to the aether.

In special relativity, any observer in motion relative to an object measures a contraction. As the contraction is simply a feature of observation from different reference frames, no force is required to cause the contraction. Nothing actually happens to the object in its reference frame.

FIGURE 11.22 A spaceship travelling at high speed has its length contracted. The contraction is only in the direction of motion of the spaceship.



The Lorentz contraction is negligible at velocities we commonly experience. Even at a relative speed of 10% of the speed of light, the contraction is less than 1%. As speed increases beyond 0.1c, however, the contraction increases until at relative speed c, the length becomes zero.

11.6 SAMPLE PROBLEM 1

Observers on Earth observe the length of a spacecraft travelling at 0.5c to have contracted. By what percentage of its proper length is the spacecraft contracted according to the observers?

SOLUTION:

$$L = \frac{L_0}{\gamma}$$

$$\frac{L}{L_0} = \frac{1}{\gamma} = \frac{1}{1.155} = 0.866$$

The spacecraft appears to be only 0.866 or 86.6% of its proper length. This is a contraction of 13.4%.

11.6 Exercise 1

1 Rebecca and Madeline take measurements of the journey from Melbourne to Sydney. Rebecca stays in Melbourne and stretches a hypothetical tape measure between the two cities. Madeline travels towards Sydney at great speed and measures the distance with her own measuring tape that is in her own reference frame.
  (a) How would the two measurements compare, assuming that perfect precision could be achieved?
  (b) Which measurement could be considered to be the proper length of the journey? Explain.

2 What is the length contraction of a galaxy, 100 000 light years across, according to a cosmic ray proton hurtling by the galaxy at 99% of the speed of light?

AS A MATTER OF FACT

The twins paradox

A paradox is a seemingly absurd or contradictory statement. Relativity provides a few paradoxes that are useful in teaching the implications of relativity. The 'twins paradox' is probably the best known. Despite its name, the twins paradox is explained fully by the logic of relativity.

Imagine a spacecraft that starts its journey from Earth. After 3 years in Earth time it will turn around and come back, so that those on Earth measure the total time between the events of the launch and the return to take 6 years. The astronaut, Peter, leaves his twin brother, Mark, on Earth. During this time, Peter and Mark agree that Earth has not moved from its path through space, it is Peter in his spaceship who has gone on a journey and has experienced the effects of acceleration that Mark has not. Mark measures the length of Peter's journey from Earth. His measurement is longer than Peter's due to length contraction, but the speed of Peter is measured relative to Earth. They disagree on distance travelled but not speed, so they must disagree on time taken. This is not just an intellectual dispute — the difference in time will show in their ageing, with Peter actually being younger than Mark on his return to Earth.

We all go on a journey into the future; we cannot stop time. Relativity shows us that the rate that time progresses depends on the movements we make through space on the journey. Coasting along in an inertial reference frame is the longest path to take. Zipping through different reference frames then returning home enables objects to reach the future in a shorter time: they take a longer journey through space but a shorter journey through time.

AS A MATTER OF FACT

The parking spot paradox

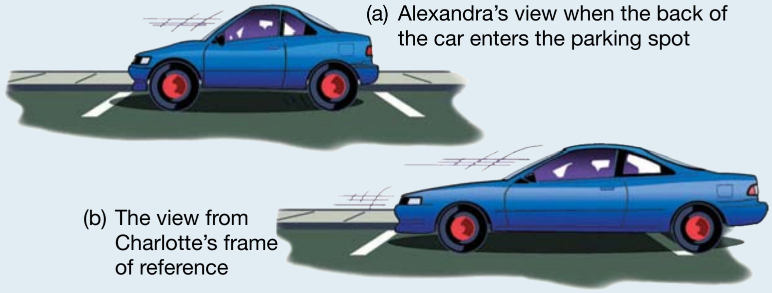
Can a long car enter a parking spot that is too short for it by making use of length contraction? The answer is yes and no. To explain, consider another famous paradox of relativity.

Charlotte's car is 8 m long and she proudly drives it at a speed of 0.8c. She observes her friend Alexandra, who is stationary on the roadside, and asks her to measure the length of her car. (For the sake of argument, we will ignore the issues of where a car could go at such a huge speed, and how Alexandra communicates with Charlotte and measures the car.)

Alexandra says that Charlotte must be dreaming if she thinks her car is 8 m long, because she measures it to be only 4.8 m long. She believes her measurements to be accurate.

To prove her point, Alexandra marks out a parking spot 4.8 m long. She says that if Charlotte can park her car in the spot, then the car is not as long as she thinks. Charlotte argues that her car will not fit in a 4.8 m parking spot, but she agrees to the test.

FIGURE 11.23 The parking spot paradox.



From Charlotte's frame of reference, the parking spot would be merely 2.9 m long. This is because it has a length contraction due to the car's relative motion of 0.8c. Alexandra's measuring equipment detects that the front of the car reaches the front of the parking spot at the same instant as the back of the car fits in the back. However, much to Alexandra's amazement, the stopped car is 8 m long. Charlotte and Alexandra now agree that the stopped car does not fit the 4.8 m parking spot, and that it has a length of 8 m. This may at first seem impossible, which is why it is sometimes called a paradox. Once we consider that Charlotte and Alexandra do not agree on which events are simultaneous, the paradox is resolved. Alexandra measured the front and the back of the car to be within the parking spot at the same time but did not check that the front and back had stopped.

11.6.2 A note on seeing relativistic effects

In this topic, we use the term observer frequently. Much of the imagery used in teaching relativity is in principle true but in practicality fantasy. Seeing anything in detail that is moving at close to the speed of light is not feasible. However, measuring distances and times associated with these objects is reasonable. Images formed of objects moving at speeds approaching c will be the result of time dilation, length contraction and other effects including the relativistic Doppler effect and the aberration of light.

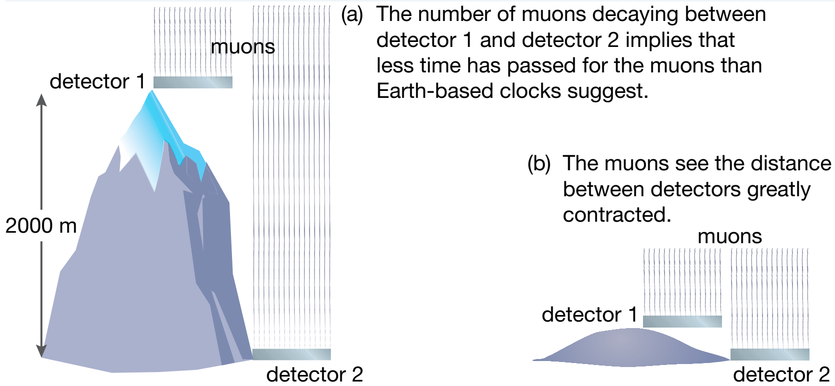
Imagine speeding through space in a very fast spacecraft. When you planned your trip on Earth, you forgot to take relativity into account. Everything on board would appear normal throughout the trip, but when you looked out the front window, the effects of relative speed would be obvious. Some examples of what you would see include: aberration of light causing the stars to group closer together, so that your forward field of vision would be increased; the Doppler effect causing the colours of stars to change; and the voyage taking much less time than you expected.

11.6.3 The journey of muons

Bruno Rossi and David Hall performed a beautiful experiment in 1941, the results of which are consistent with both time dilation and length contraction. Earth is constantly bombarded by energetic radiation from space, known as cosmic radiation. These rays collide with the upper atmosphere, producing particles known as muons. Muons are known to have a very short half-life, measured in the laboratory to be 1.56 microseconds. Given the speed at which they travel and the distance they travel through the atmosphere, the vast majority of muons would decay before they hit the ground.

The Rossi–Hall experiment involved measuring the number of muons colliding with a detector on top of a tall mountain and comparing this number with how many muons were detected at a lower point. They found that far more muons survived the journey through the atmosphere than would be predicted without time dilation. The muons were travelling so fast relative to Earth that the muons decayed at a much slower rate for observers on Earth than they would at rest in the laboratory. The journey between the detectors took about 6.5 microseconds according to Earth-based clocks, but the muons decayed as though only 0.7 microseconds had passed. Due to length contraction, the muons did not see the tall mountain but, rather, a small hill. Rossi and Hall were not surprised that the muons survived the journey at all.

FIGURE 11.24 Muons are a measurable example of special relativistic effects.



11.6 SAMPLE PROBLEM 2

Use the description of the Rossi–Hall experiment to answer the following questions.

(a) What is the proper time for the half-life of muons?

(b) What is the value of gamma as determined from the journey times from the different reference frames?

(c) How fast were the muons travelling though the atmosphere according to the value for gamma?

(d) Calculate the half-life of the muons from the reference frame of the Earth.

SOLUTION:

(a) The proper time for the half-life is in the reference frame of the muon and is 1.56 microseconds.

(b) $t = t_0\gamma$

$$\gamma = \frac{t}{t_0} = \frac{6.5}{0.7} = 9.29$$

(c) $$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$v = c\sqrt{1 - \frac{1}{\gamma^2}} = c\sqrt{1 - \frac{1}{9.29^2}} = 0.994c$$

(d) $t = t_0\gamma$

$$t = 1.56 \times 10^{-6} \times 9.29 = 14\,\mu\text{s}$$

11.6 Exercise 2

1 Use the description of the Rossi–Hall experiment above to answer the following questions.

a. Use the travel time from the Earth reference frame and the speed of the muons to calculate the height of the mountain.

b. Use the travel time of the muons to determine how high the mountain appeared to the muons.

11.7 Relativistic momentum

In Newtonian physics we understand that momentum is given by $p = mv$ and that momentum is increased by applying a force described by $\Delta p = F\Delta t$. However, this implies that if we apply a force long enough, we could accelerate an object beyond the speed of light. This would create negative values under our square root sign in γ, so something else must be taken into account.

To calculate momentum in special relativity, we simply replace m in the momentum formula with this relativistic mass:

$$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Why does mass need to be relative if we use Einstein's postulates? Consider the change of momentum expression of Newton's Second Law: $\Delta p = F\Delta t$. This expression treats Δt as the proper time, Δt_0. Given that this describes an object in relative motion, Δt is affected by time dilation and m as the rest mass m_0. The time dilation formula is $\Delta t = \gamma \Delta t_0$, so $\Delta p = F\frac{\Delta t}{\gamma}$ or $\gamma m_0 \Delta v = F\Delta t$. We now have $\Delta p = \gamma m_0 \Delta v$, which is consistent with the formula given above. The momentum is consistent with a mass increase by a factor of gamma, $m = \gamma m_0$.

It turns out that not only time and space are relative but that mass is too. The faster a body is going relative to the observer's reference frame, the more massive it becomes, $m = m_0\gamma$, where m_0 is the rest mass, the mass of the object observed from a frame of reference not moving relative to the object.

Relativistic momentum is important in particle accelerators where subatomic particles are accelerated to speeds approaching the speed of light. The force required to accelerate the particles at a constant rate increases with speed, unlike what we experience at low velocities where force is proportional to acceleration through $F = ma$. Newton's second law turns out to be just a very good approximation at low velocities. Once the velocity exceeds 0.1c, the difference between Newton's predictions and Einstein's become large. The behaviour of particles in particle accelerators conforms with the predictions of special relativity.

11.7 SAMPLE PROBLEM 1

Calculate the momentum of a proton moving at 0.6c relative to an observer.

SOLUTION:

$$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$p = \frac{1.67 \times 10^{-27} \times 0.6 \times 3 \times 10^8}{\sqrt{1 - 0.6^2}} = 3.76 \times 10^{-19}\ \text{kg m s}^{-1}$$

11.7 Exercise 1

1 Calculate the momentum of an electron moving at 0.9 c relative to a target it is about to hit.

11.8 The most famous equation: $E = mc^2$

11.8.1 Another thought experiment

The result of special relativity that people are most familiar with is the equation $E = mc^2$. In fact, it is probably the most well known equation of all. This formula expresses an equivalence of mass and energy. If we do work, ΔE, on an object, that is we increase its energy, its mass will increase. Usually, however, we do not notice this increase in mass because of the factor $c^2 = 9 \times 10^{16}\ \text{m}^2\ \text{s}^{-2}$. According to $\Delta E = \Delta mc^2$, it would take 9×10^{16} J of energy to increase the mass by 1 kg. This is similar to the amount of electrical energy produced in Victoria every year. Conversely, if we could convert every gram of a 1 kg mass into electricity, we would supply Victoria's electricity needs for a year. Nuclear fission reactors produce electricity from the small loss of mass that occurs when large nuclei such as those of uranium-235 undergo fission. The Sun and other stars generate their energy by losing mass to nuclear fusion.

A simplified derivation of this equation can help us gain a sense of the physics involved. Consider a box suspended in space, with no external forces acting on it, as shown in Figure 11.25. Maxwell found that electromagnetic radiation carries momentum $p = \dfrac{E}{c}$ where E is the energy transmitted and c is the speed of light. In the context of photons, each photon carries a momentum $p = \dfrac{E}{c}$. As a result, light exerts pressure on surfaces. This effect can nudge satellites out of orbit over time.

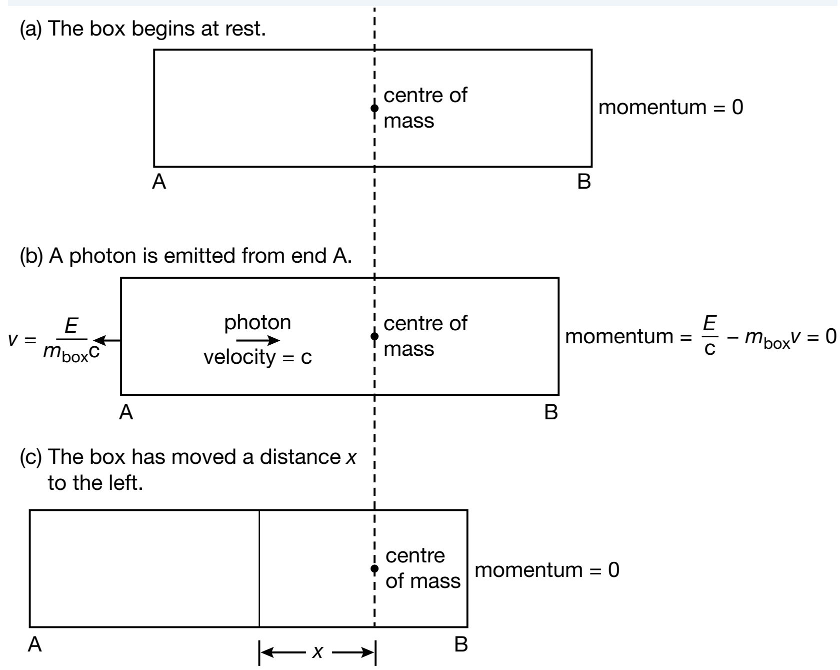
In Figure 11.25a, the box begins at rest. The total momentum is zero and its centre of mass is in the centre.

In Figure 11.25b, a photon of energy E is emitted from end A, carrying momentum with it. To conserve momentum, the box moves in the direction opposite to the movement of the photon.

$$p_{\text{photon}} + p_{\text{box}} = 0$$

$$\Rightarrow \frac{E}{c} - m_{\text{box}} v = 0$$

FIGURE 11.25 Einstein's box suspended in space.



where
m_{box} = the mass of the box
v = the velocity of the box.

Rearranging gives us the velocity of the box in the leftward direction, $v = \dfrac{E}{m_{box}c}$, a very small number!

In Figure 11.25c, after time Δt, the light pulse strikes the other end of the box and is absorbed. The momentum of the photon is also absorbed into the box, bringing the box to a stop. In this process, the box has moved a distance x where:

$$x = v\Delta t$$

Substituting $v = \dfrac{E}{m_{box}c}$ from (b) gives

$$x = \frac{E\Delta t}{m_{box}c}$$

As v is very small (almost non-existent), we can assume that the photon travels the full length of the box and put $\Delta t = \dfrac{L}{c}$. Substituting this into $x = \dfrac{E\Delta t}{m_{box}c}$ gives:

$$x = \frac{EL}{m_{box}c^2}$$

or $E = \dfrac{xm_{box}c^2}{L}$

There are no external forces acting on the box, so the position of the centre of mass must remain unchanged (see the dotted line in the diagram). The box moved to the left as a result of the transfer of the energy of the photon to the right. Therefore, the transfer of the photon must be the equivalent of a transfer of mass. If we can show that $\dfrac{xm_{box}}{L}$ is the same as the mass equivalent of the transferred energy, we have our answer. To show this, we will pay attention to the shift in the box relative to the centre of mass of the system.

The centre of mass is the point where the box would balance if suspended. This can be determined by balancing moments — the mass times the distance from a reference point. We choose the centre of the box as the reference point to ensure that the distance x is in our calculations. The moment for the box is $m_{box}x$ anticlockwise, because the mass of the box can be considered to be acting through a point at distance x to the left of the reference point. The photon's equivalent mass is acting at distance $\dfrac{L}{2}$ to the right of the reference point, so its moment is $m\dfrac{L}{2}$ clockwise. However, this moment was acting on the other end of the box before the photon was emitted, so we can consider its absence from that end of the box as an equal moment in the same direction. We then have:

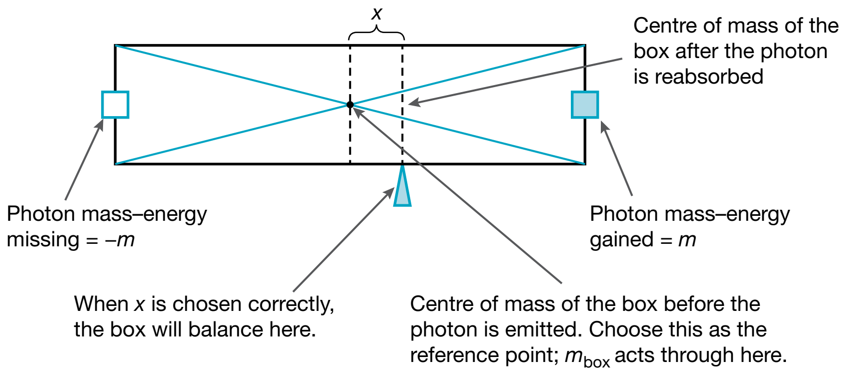
$$m_{box}x = m\frac{L}{2} + m\frac{L}{2}$$

or $m = \dfrac{m_{box}x}{L}$ as required.

Substitute this into $E = \dfrac{xm_{box}c^2}{L}$ and we have $E = mc^2$.

In other words, when the photon carried energy to the other end of the box, it had the same effect as if it had carried mass. In fact, Einstein concluded that energy and mass are equivalent. If we say that some energy has passed from one end of the box to the other, we are equally justified in saying that mass has passed as well. Note the distinction: the photon carries an amount of energy that is equivalent to an amount of mass, but the photon itself does not have mass.

FIGURE 11.26 Balancing Einstein's box.



One implication of this is that the measurement of mass depends on the relative motion of the observer. The kinetic energy of a body depends on the inertial reference frame from which it is measured. The faster the motion, the greater the kinetic energy. So kinetic energy is relative, and so is mass! Energy is equivalent to mass, so the mass of an object increases as its velocity relative to an observer increases.

The mass of an object that is in the same inertial frame as the observer is called its **rest mass** (m_0). When measured from other reference frames, the mass is given by $m = m_0\gamma$. The derivation of this is complex, so it will not be addressed here.

11.8 SAMPLE PROBLEM 1

Use $m = m_0\gamma$ to show that it is not possible for a mass to exceed the speed of light.

SOLUTION:

If $v = c$, γ becomes infinitely large. As $m = m_0\gamma$, an object travelling at c would have infinite mass. Speeds larger than c would produce a negative under the square root sign, so these speeds are not possible

11.8 Exercise 1

1 The Earth ($m = 6 \times 10^{24}$ kg) moves around the Sun at close to 30 000 m s^{-1}. From the Sun's frame of reference, how much additional mass does the Earth have?

11.8 SAMPLE PROBLEM 2

Calculate the mass increase of a proton that is accelerated from rest using 11 GeV of energy, an energy that can be achieved in particle accelerators.

SOLUTION:

$$\begin{aligned}\Delta E &= 11\ \text{GeV} \\ &= 11 \times 10^{9} \times 1.6 \times 10^{-19}\,\text{J} \\ &= 1.76 \times 10^{-9}\ \text{J}\end{aligned}$$

$$\begin{aligned}\Delta m &= \frac{\Delta E}{c^2} \\ &= \frac{1.76 \times 10^{-9}\ \text{J}}{(3 \times 10^{8}\ \text{m s}^{-1})^2} \\ &= 1.96 \times 10^{-26}\ \text{kg}\end{aligned}$$

Note that the rest mass of a proton is 1.67×10^{-27} kg, so the accelerated proton behaves as though its mass is nearly 13 times its rest mass.

11.8 SAMPLE PROBLEM 3

In Newtonian physics, if we gave a proton 11 GeV of kinetic energy, what would be its speed?

SOLUTION:

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

$$\begin{aligned} v &= \sqrt{\frac{2E}{m}} \\ &= \sqrt{\frac{2 \times 11 \times 10^9 \times 1.6 \times 10^{-19}}{1.67 \times 10^{-27}}} \\ &= 1.45 \times 10^9 \text{ m s}^{-1} \end{aligned}$$

This speed is not possible as the maximum speed attainable is 3×10^8 m s^{-1}.

11.8 Exercise 2

1. Calculate the mass of a proton hurtling through space with 50 GeV of kinetic energy.
2. Calculate the rest mass of a particle that has 20 GeV of kinetic energy with a relative speed of 0.5c.

11.9 Relativity and momentum

11.9.1 Relativity and inertia

The solution to 11.8 sample problem 3 is well in excess of the speed of light, and is an example of the limitations of Newtonian physics. In relativity, when more energy is given to a particle that is approaching the speed of light, the energy causes a large change in mass and a small change in speed. By doing work on the particle, the particle gains inertia, so the increase in energy has an ever-decreasing effect on the speed. The speed cannot increase beyond the speed of light, no matter how much energy the particle is given.

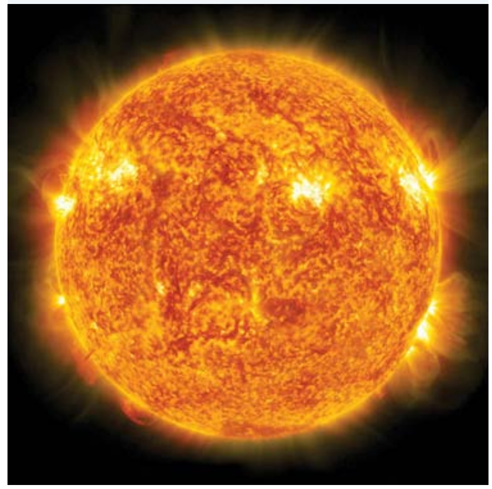
FIGURE 11.27 Particle accelerators such as the Australian Synchrotron in Melbourne accelerate subatomic particles to near-light speeds, where special relativity is essential for understanding the behaviour of the particles. Electrons in the Australian Synchrotron have kinetic energies up to 3 GeV.

In particle accelerators, where particles are accelerated to near the speed of light, every tiny increase in the speed of the particles requires huge amounts of energy. Physicists working in this field rely on ever-higher energies to make new discoveries. This costs huge amounts of money. Nonetheless, a number of accelerators have been built that are used by scientists from around the world. This area of research is often called high-energy physics. At these high energies, Newtonian mechanics is hopelessly inadequate and Einstein's relativity is essential.

11.9.2 Mass conversion in the Sun

One of the consequences of Einstein's great contribution to our understanding of relativity is that we understand now a great deal about how energy is generated by the Sun. At the centre of it all is the equation $E = mc^2$. The Sun continuously converts mass–energy stored as mass into radiant light and heat. Each second the Sun radiates enough energy to meet current human requirements for billions of years. It takes the energy generated in the core about 100 000 years to reach the surface. Even if the fusion in the Sun stopped today, it would take tens of thousands of years before there was a significant impact on Earth.

FIGURE 11.28 The Sun's energy comes from nuclear fusion converting mass into energy.



The Sun is a ball made up mostly of hydrogen plasma and some ionised atoms of lighter elements. The temperatures in the Sun ensure that virtually all the atoms are ionised. The composition of the Sun is shown in Table 11.1.

At this stage of the Sun's life cycle, it is ionised hydrogen atoms (i.e. protons) that provide the energy. The abundance of protons and the temperatures and pressures in the core of the Sun are sufficient to fuse hydrogen, but not heavier nuclei. We will learn more about this process in the next topic.

The energy is released mainly through the gamma photons and the annihilation of the positrons when they meet free electrons in the Sun. Positrons are antimatter versions of electrons. They have a positive charge. When positrons and electrons come together, they annihilate each other, resulting in a release of energy equal to their masses times the speed of light squared. The net result is an enormous release of energy and a corresponding loss of mass. The mass loss has been measured to be 4.4 Tg (4.4×10^9 kg) per second. As the mass of the Sun is around 2.0×10^{30} kg, even at this incredible rate, there is plenty of hydrogen to sustain it for about twice its age of four and a half billion years.

TABLE 11.1 The composition of the Sun.

Element	Percentage of total number of nuclei in the Sun	Percentage of total mass of the Sun
Hydrogen	91.2	71.0
Helium	8.7	27.1
Oxygen	0.078	0.97
Carbon	0.043	0.40
Nitrogen	0.0088	0.096
Silicon	0.0045	0.099
Magnesium	0.0038	0.076
Neon	0.0035	0.058
Iron	0.030	0.014
Sulfur	0.015	0.040

11.9 SAMPLE PROBLEM 1

A nucleus of hydrogen-2 made of one proton and one neutron has a smaller mass than the total of an individual proton and an individual neutron. Account for this mass difference.

SOLUTION:

The mass of the nucleus is different to the mass of the individual particles, but when the binding energy of the hydrogen-2 nucleus is included, we find that the mass–energy of both is the same. The separate particles have their mass and zero potential energy. The particles bound in the nucleus have a reduced mass and the binding energy of the nucleus. (The binding energy is the energy required to separate the particles. It is released as a combination of increased kinetic energy of the particles and gamma rays.)

11.9 SAMPLE PROBLEM 2

What is the power output of the Sun?

SOLUTION:

$$\begin{aligned} E &= mc^2 \\ &= 4.4 \times 10^9 \times (3.0 \times 10^8)^2 \text{ J} \\ &= 4.0 \times 10^{26} \text{ J} \\ P &= \frac{E}{t} \\ &= 4.0 \times 10^{26} \text{ W} \end{aligned}$$

The mass loss of 4.4×10^9 kg s^{-1} equates to a power output of 4.0×10^{26} W.

11.9 Exercise 1

1. An electron ($m = 9.11 \times 10^{-31}$ kg) and a positron (also $m = 9.11 \times 10^{-31}$ kg) collide and are annihilated. Calculate the energy released.
2. One tonne of coal can produce about 20 GJ of energy by combustion in 1.6 tonne of air. Calculate the mass change that results.

11.10 Review

11.10.1 Summary

- There is no frame of reference that is at absolute rest. Velocity is always relative to a chosen reference frame.
- Classical physics is the physics established by Galileo, Newton and other scientists before the twentieth century. It does not include twentieth-century developments in physics, such as special relativity and quantum mechanics.
- In classical physics, velocity is relative, but time, distance and mass measurements are invariant — they are the same for all observers. Classical physics provides a good approximation at low velocities, but it does not provide accurate values as relative speeds approach the speed of light.
- In special relativity, velocities of masses are still relative but the speed of light is invariant. As a result, it is recognised that the measurement of time intervals, lengths and masses is relative to the reference frame of the observer.

- Einstein's two postulates of special relativity are:
  - 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.
- Proper time is the time interval between two events in a reference frame where the two events occur at the same point in space, that is, the reference frame in which the clock is stationary.
- Proper length is the length that is measured in the frame of reference in which objects are at rest.
- In reference frames in motion relative to the observer, time is dilated according to $t = t_0\gamma$, where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

- In reference frames in motion relative to the observer, length is contracted along the line of motion according to $L = \frac{L_0}{\gamma}$.
- In reference frames in motion relative to the observer, mass increases according to $m = m_0\gamma$.
- An example of where the effects of special relativity can be observed is muons formed in the upper atmosphere. They travel to Earth at nearly the speed of light, so that even though most would decay in the time it takes them to reach the surface according to classical physics, many survive the journey as they see the distance contracted. From the perspective of the Earth, the time is dilated so that the muons have time to reach the surface.
- Momentum is relative and is given by the formula $\rho = \gamma m_0 v$.
- $E = m\mathrm{c}^2$ expresses the equivalence of mass and energy.
- Fusion is the source of the Sun's energy. The Sun is constantly losing mass as it radiates energy in accordance with mass–energy equivalence.

11.10.2 Questions

1. According to Maxwell, who would see light travelling the fastest?
   (A) Someone moving towards a light source that is stationary in the aether
   (B) Someone who is stationary in the aether with the light source moving away
   (C) Someone who is stationary in the aether with the light source moving towards her
   (D) Someone who is moving away from a light source that is stationary in the aether.
2. What is a frame of reference?
3. What do physicists mean when they say that velocity is relative?
4. What is the difference between an inertial and a non-inertial reference frame?
5. How can you determine whether your car is accelerating or moving with constant velocity?
6. Two cars drive in opposite directions along a suburban street at $50\,\mathrm{km\,h^{-1}}$. What is the velocity of one car relative to the other?
7. Explain, using the concept of velocity, why head-on collisions are particularly dangerous. Use an example.
8. Earth varies from motion in a straight line by less than 1° each day due to its motion around the Sun.
   (a) Explain, with the help of the principle of relativity, why we do not feel Earth moving, even though it is travelling around the Sun at great speed.
   (b) What are the other motions Earth undergoes that we cannot feel?
   (c) Earth is not an inertial reference frame. Explain why we often refer to it as though it is.
9. A car accelerates from 0 to $100\,\mathrm{km\,h^{-1}}$ in 10 s.
   (a) What is its acceleration relative to the road?
   (b) What is its acceleration relative to a car travelling at $100\,\mathrm{km\,h^{-1}}$ in the opposite direction?
   (c) Would you describe the acceleration as absolute, relative, invariant or arbitrary?

10. (a) If Earth is moving at 100 km s^{-1} relative to the supposed aether, what speed would Michelson have measured for light emitted in the same direction that Earth is travelling?
    (b) What speed would Michelson have expected given the aether theory? (Take the speed of light to be 2.9979×10^8 m s^{-1}.)
11. (a) What are Einstein's two postulates of special relativity?
    (b) What is in these postulates that was not present in previous physics?
12. What place did the luminiferous aether take in Einstein's theory?
13. (a) Why did Newton's laws seem correct for so long?
    (b) Why do we often still use Newton's laws today?
14. Why is Einstein's second postulate surprising? Give an example to show why Newtonian physicists would think it wrong.
15. A star emits light at speed c. A second star is hurtling towards it with speed 0.3c. What is the speed of the light when it hits the second star relative to this second star?
16. Explain how Einstein's first postulate makes sense of the results of the Michelson–Morley experiment.
17. What is time dilation? In your explanation, give an example of where time dilation would occur.
18. If a box was moving away from you at nearly light speed, which dimensions of the box would undergo length contraction from your perspective: width, height or depth?
19. Which clock runs slow: yours or one in motion relative to you?
20. You observe that an astronaut moving very quickly away from you ages at a slower rate than you. The astronaut views you as ageing faster than she ages. True or false? Explain.
21. The twins paradox shows that less time passes for the travelling twin. Does this also mean that the twin will return shortened due to length contraction? Explain.
22. Draw diagrams of a light clock in motion and at rest to explain why time dilation occurs for moving clocks.
23. Explain why time dilation must occur for all clocks, not just the light clock.
24. Explain the difference between t_o and t in the time dilation formula.
25. Two spacecraft pass each other with a relative speed of 0.3c.
    (a) Calculate γ.
    (b) A drummer pounds a drum at 100 beats per minute on one of the spacecraft. How many beats per minute would those on the other spacecraft measure as a result of time dilation?
26. An alien spacecraft speeds through the solar system at 0.8c.
    (a) What is the effect of its speed on the length of the spacecraft from the perspective of an alien on board?
    (b) What is the effect of its speed on the length of the spacecraft from the perspective of the Sun?
    (c) At what speed does light from the Sun reach it?
27. A high-energy physicist detects a particle in a particle accelerator that has a half-life of 20 s when travelling at 0.99c.
    (a) Calculate the particle's half-life in its rest frame.
    (b) The detector is 5 m long. How long would it be in the rest frame of the particle?
28. It takes 5 min for an astronaut to eat his breakfast, according to the clock on his spacecraft. The clock on a passing spacecraft records that 8 min passed while he ate his breakfast.
    (a) Which time is proper time?
    (b) What is the relative speed of the two spacecraft?
29. The nearest star, apart from the Sun, is 4.2 light-years distant.
    (a) How far is it to that star according to astronauts in a spacecraft travelling at 0.7c?
    (b) How long would it take to get there in this spacecraft?
    (c) How long will the journey take, based on measurements from Earth? (Assume that Earth is stationary relative to the star.)

30. A spacecraft ($L_o = 80$ m) travels past a space station at speed 0.7c. Its radio receiver is on the tip of its nose. The space station sends a radio signal the instant the tail of the spacecraft passes the space station.
    (a) What is the length of the spacecraft in the reference frame of the space station?
    (b) How far from the space station is the nose of the spacecraft when it receives the radio signal from the reference frame of the space station?
    (c) What is the time taken for the radio signal to reach the nose of the spacecraft, according to those on the space station?
    (d) What is the time taken for the radio signal to reach the nose, according to those on the spacecraft?
31. An astronaut on a space walk sees a spacecraft passing at 0.9c. The spacecraft has a proper length of 100 m. What is the length of the spacecraft L due to length contraction according to the astronaut?
32. Explain why muons reach the surface of the Earth in greater numbers than would be predicted by classical physics given their speed, their half-lives and the distance they need to travel through the atmosphere.
33. A muon forms 30 km above the Earth's surface and travels straight down at 0.98c. From its frame of reference, what is the distance it has to travel through the atmosphere?
34. The proper time for the half-life of a muon is 1.56 microseconds. If the muon moves at 0.98c relative to an observer, what does the observer measure its half-life as?
35. Explain how muons produced by cosmic rays became an early confirmation of special relativity.
36. Use your knowledge of relativity to argue that matter cannot travel at the speed of light.
37. How much energy would be required to accelerate 1000 kg to:
    (a) 0.1 c (b) 0.5 c (c) 0.8 c (d) 0.9 c?
38. Sketch a graph of energy versus speed using your answers to the previous question.
39. Travelling at near light speed would enable astronauts to cover enormous distances. Explain the difficulties in terms of energy of achieving space travel at near light speed.
40. Which of the following would be a consequence of the relativistic mass increase of a person travelling past you at near light speed?
    (A) They would appear physically larger.
    (B) They would weigh more on a balance.
    (C) They would require more force to accelerate.
41. Explain in words what $E = mc^2$ tells us about energy and mass.
42. An astronaut in a spacecraft moves past Earth at 0.8c and measures his mass. (He has no weight in his inertial reference frame.) According to him, his mass is 70 kg.
    (a) What is his mass according to an observer on Earth?
    (b) How much energy was required to give him the extra mass?
43. Calculate the rest energy of Earth, which has a rest mass of 6.0×10^{24} kg.
44. Consider Earth to be a mass moving at 30 km s^{-1} relative to a stationary observer. Given that the rest mass of Earth is 5.98×10^{24} kg, what would be the difference between this rest mass and the mass from the point of view of the stationary observer?
45. Calculate the momentum of a 10 000 kg asteroid travelling at 0.6c according to Earth-based observers.
46. Calculate the speed of a 10 kg meteorite that has 3.0×10^9 kg m s^{-1} of momentum.
47. If a 250 g apple could be converted into electricity with 100% efficiency, how many joules of electricity would be produced?
48. Much of Victoria's electricity is produced by burning coal. What can you say about the mass of the coal and its chemical combustion products as a result of burning it?
49. What would have greater rest mass, the Moon in orbit about Earth, or the Moon separated from Earth?
50. What is happening to the mass of the Sun over time? Why?
51. Part of the fusion process in the Sun involves the fusion of two protons into a deuteron. This results in the release of 0.42 MeV of energy. What is the mass equivalent of this energy release?

TOPIC 12
Elemental origins

12.1 Overview

12.1.1 Module 8: From the universe to the atom

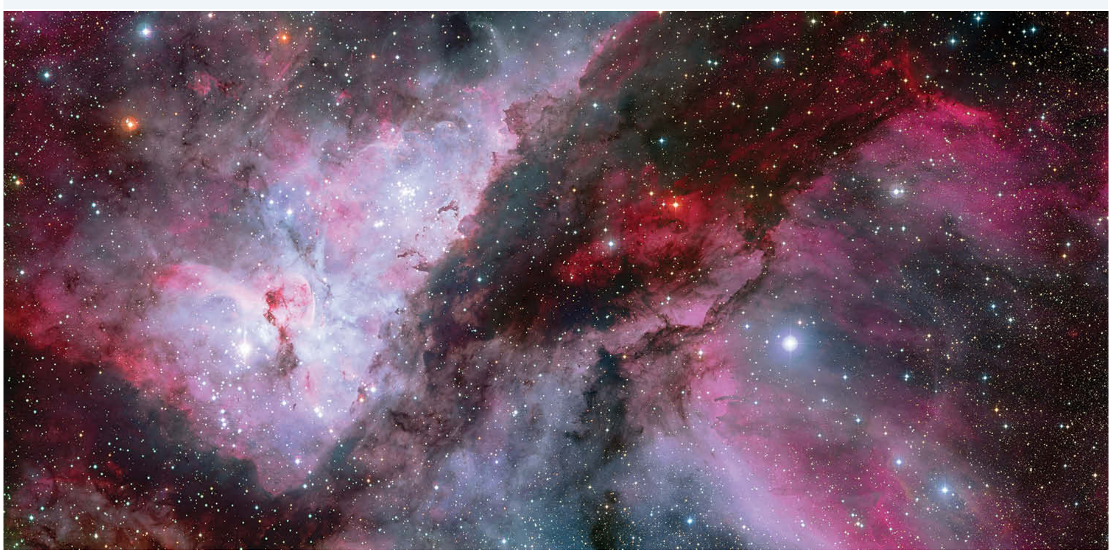
Elemental origins

Inquiry question: What evidence is there for the origins of the elements?

Students:

- investigate the processes that led to the transformation of radiation into matter that followed the 'Big Bang'
- investigate the evidence that led to the discovery of the expansion of the universe by Hubble (ACSPH138)
- analyse and apply Einstein's description of the equivalence of energy and mass and relate this to the nuclear reactions that occur in stars (ACSPH031)
- account for the production of emission and absorption spectra and compare these with a continuous black body spectrum (ACSPH137)
- investigate the key features of stellar spectra and describe how these are used to classify stars
- investigate the Hertzsprung–Russell diagram and how it can be used to determine the following about a star:
  - characteristics and evolutionary stage
  - surface temperature
  - colour
  - luminosity
- investigate the types of nucleosynthesis reactions involved in Main Sequence and Post-Main Sequence stars, including but not limited to:
  - proton–proton chain
  - CNO (carbon-nitrogen-oxygen) cycle

FIGURE 12.1 Spectacular panoramic view of the Carina Nebula and the unique star Eta Carinae in the heart of the nebula.



12.2 The earliest atoms

12.2.1 A long story!

Scientists find it useful to analyse matter in terms of the atoms that make it up. This is the basis of chemistry, and our understanding of biology and geology. But atoms did not always exist; they only exist under the right conditions. Through an amazing journey of exploration, physicists are now confident that the first atoms formed about 13.82 billion years ago, a mere 380 000 years after the universe itself came into being. This topic explores how physicists have come to this conclusion and the later developments that led to the atoms that make up you and me.

To understand the origin of atoms, we will need to understand the early universe. The big bang is the name given to the theory that scientists use to explain why the universe is as it is. The big bang model of the universe is a triumph of decades of observation, measurement, theory and scientific exploration. However, there is still a lot that is not known, and there are alternative interpretations. The universe is a very active research field with new and surprising discoveries being made and new questions being asked.

The big bang model of the universe describes the universe as beginning from a point, or singularity, that has grown through expansion for the past nearly 14 billion years. Subsequent sections of this topic will explore how physicists came to this conclusion. The big bang was not an explosion in the normal sense of the word, where material located at one point in space at one time is suddenly pushed into the space around it over a short period of time. With the big bang, there was no matter at the start, no space and no time. All of these emerged from the big bang itself. This is very challenging to understand because many questions arise:

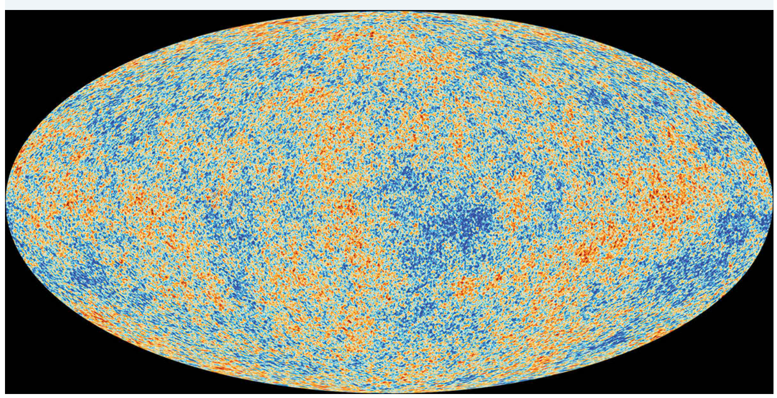
- 'What does it mean for time to begin?'
- 'Where did the big bang occur, if there was no space before it?'
- 'Where did all of the energy come from in the first place?'

Science often has this effect; we learn about relations between things that we see and find that these lead to further questions. These questions take us to the edge of the universe in space and time but may also take us to the edge of what it is possible to understand.

12.2.2 The early universe

The very early universe contained no stars and galaxies; in fact, there were no atoms. The earliest data that scientists have for the universe is from what is known as the Cosmic Microwave Background (CMB). This is the very low energy radiation left over from when the first atoms formed. The CMB is one of the key pieces of evidence for the big bang model. It provides an image of the universe as it was about 380 000 years after its beginning.

FIGURE 12.2 The cosmic microwave background as measured by the Planck satellite. The differences in colour indicate temperature differences that correspond to differences in the density of the universe when it was 380 000 years old.



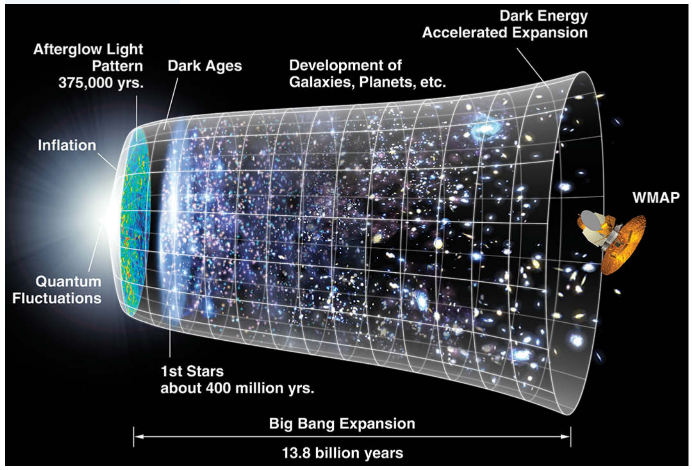
To understand what happened prior to this event, physicists rely on particle physics and the Theory of General Relativity to help them make sense of their observations of the CMB and the universe that followed. Particle physics experiments, such as those in the Large Hadron Collider that discovered the Higgs boson at CERN, create conditions that existed for particles in the early universe, so physicists do have experimental evidence for much of what happened prior to the release of energy that formed the cosmic microwave background.

Let's piece the story together from the beginning until the formation of the very first atoms, the time when the CMB was formed.

The first 10^{-43} seconds

This is known as the Planck era and current theories of physics cannot explain what happened in the conditions that were present. The universe was too small, too dense, too hot and existed for too short a time for current physics to say anything precise about it. Physicists refer to what existed prior to when the universe was 10^{-43} seconds old as a singularity. With this period, space and time began. Gravity became a distinct force at the end of the Planck era. The temperature was 10^{32} degrees Celsius and the universe was 10^{-35} cm across.

FIGURE 12.3



Source: NASA WMAP Science Team.

10^{-43} seconds to 10^{-36} seconds

This tiny interval of time in the early universe is known as the grand unified era. During this period, physicists believe that the strong nuclear force, the weak nuclear force and the electromagnetic force did not yet exist as separate forces. The first matter begins to form, but for each particle of matter, there was a particle of antimatter. As soon as a particle was formed, it would meet an antiparticle and be annihilated.

10^{-36} seconds to 10^{-32} seconds

This next period of time is known as the inflation era. During this time, a period of exponential expansion has been proposed to explain a number of features of the universe. During this brief period, the universe is thought to have expanded in size by a factor of 10^{26}, so that it was about 10 cm across. This rapid expansion explains, among other things, why the cosmic microwave background radiation is so uniform, being close to 2.7 degrees above absolute zero in all directions.

Alan Guth proposed this radical idea in 1980. It was a response to some of the limitations of the standard big bang model that worked well for most of the evidence, but fell short when it came to explaining the relative uniformity of the CMB in all directions, and what is known as the 'flatness' of the universe. At first, inflation may look like a crazy idea invented just to explain away problems. However, in all the time since, no one has come up with a more successful theory to explain why the visible universe is so uniform in temperature or why the universe appears so 'flat'. In a flat universe, parallel lines remain an equal distance apart. In a curved universe, they could be more like north–south lines on the surface of the Earth, which meet at the poles. It is thought that rapid expansion during this era smoothed out deviations in the flatness of the universe that would otherwise have grown with time.

The small variations in the cosmic microwave background detected by the Planck orbiting observatory are consistent with a universe that underwent a period of inflation and resulted in the clumping of matter into clusters of galaxies.

10^{-32} seconds to 10^{-12} seconds

The electroweak era was the period when the strong nuclear force came into play. The Higgs boson formed, enabling particles to have mass.

10^{-12} seconds to 10^{-6} seconds

In the quark era, particles began to appear in large numbers. These included quarks, electrons and neutrinos. Most particles still formed in pairs with an antiparticle, but a slight bias towards particles resulted in matter that was not annihilated through contact with antimatter.

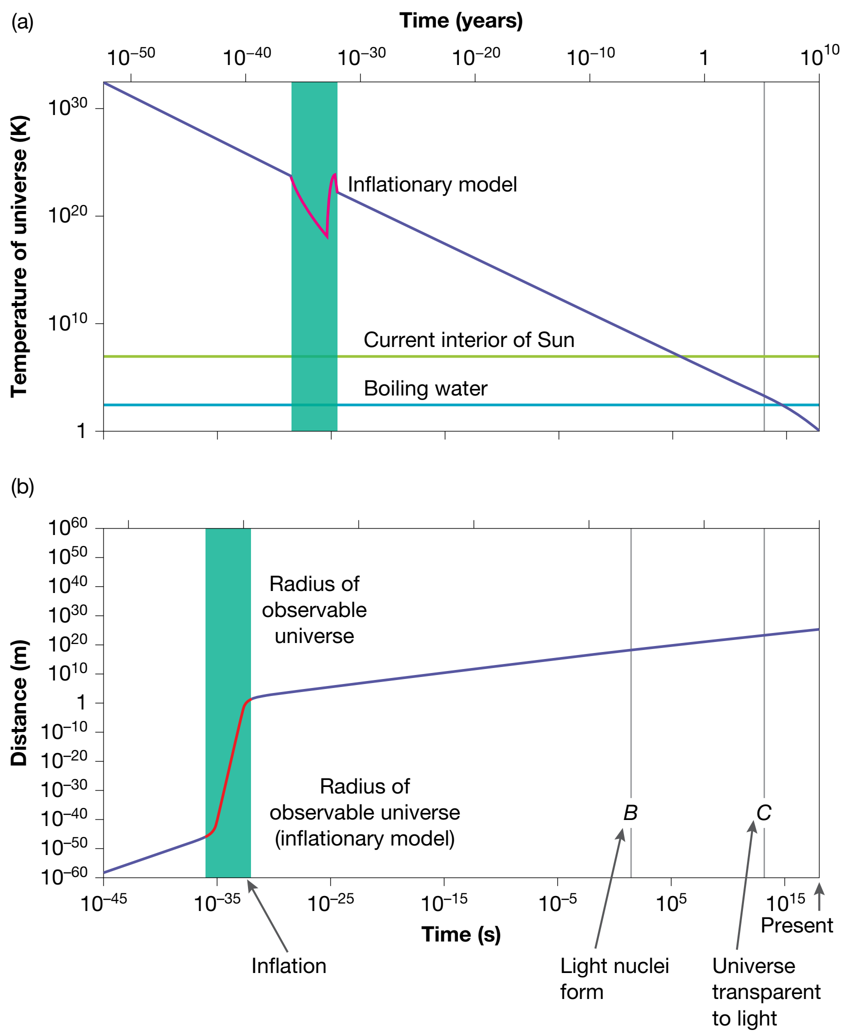
10^{-6} seconds to 3 minutes

The hadron era was the period when the temperature of the universe had dropped to the point where three quarks could form protons and neutrons (particles like neutrons and protons made from three quarks are called hadrons). Most were annihilated by contact with their antiparticles, and leptons such as electrons dominated.

3 minutes to 20 minutes

Nucleogenesis occurs. During this era, annihilation with antimatter became less significant and the critical phase of fusion occurred. During these few minutes, most of the nuclei in the universe formed. The following section outlines this in detail.

FIGURE 12.4



12.2.3 The birth of atoms

In 1948, George Gamow (1904–1968) and Ralph Alpher (1921–2007) proposed a model that explained how over 99% of the atoms found in the universe could have formed. The protons and neutrons in the early period of the universe readily interacted with the abundance of electrons and neutrinos present, causing them to change from proton to neutron and vice versa. This produced equal numbers of each, but as the universe cooled, fewer protons interacted with electrons with sufficient energy to form neutrons. This resulted in more protons being formed than neutrons in a ratio of about 7 to 1. This formed a universe of hydrogen. Under the right temperature and pressure, protons can fuse to form helium-4. These conditions were present when the early universe was about 400 seconds old, allowing nuclei up to a mass number of 4 to form.

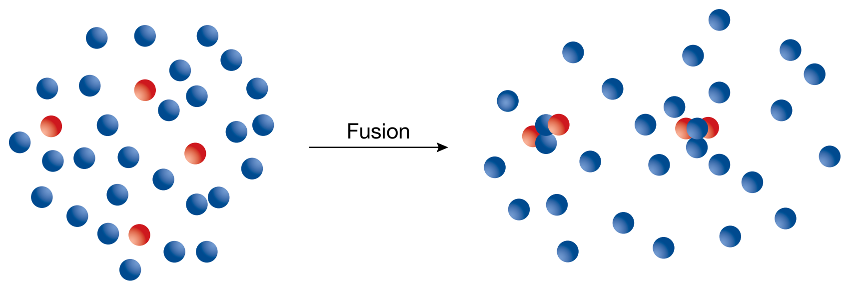
This produced the following nuclei:

- hydrogen (1 proton)
- deuterium (an isotope of hydrogen with 1 proton and 1 neutron)
- tritium (an isotope of hydrogen with 1 proton and 2 neutrons)
- helium-3 (2 protons and 1 neutron)
- helium-4 (2 protons and 2 neutrons).

No stable isotope with a mass number of 5 exists (this has been tested in the laboratory), so fusion of nuclei beyond helium-4 was very limited. Tiny quantities of lithium-7 and beryllium-7 were produced through fusion of helium nuclei, but the step to heavier elements involving the fusion of three helium nuclei to form carbon takes too long. The universe rapidly cooled through its expansion to the point where the conditions no longer supported fusion. This provided the young universe with its composition of about 75% hydrogen and 25% helium by mass.

To understand the prediction of the proportion of hydrogen to helium, consider the ratio of protons to neutrons, which is 7:1. To form a helium-4 nucleus, 2 protons and 2 neutrons are required. According to the 7:1 ratio, for every 2 neutrons in the universe, there were 14 protons, so the formation of helium-4 would take 2 neutrons and 2 protons, leaving 12 protons in the mix. That leaves 4 nucleons in helium and 12 in hydrogen, or 25% of the mass in helium and 75% in hydrogen. The fusion in these first moments of the universe was so rapid and complete that virtually all of the available neutrons went into helium-4. Only a tiny proportion remained as deuterium or tritium (which decays to helium-3) so this simple calculation gives a very good prediction of the composition of the universe prior to star formation.

FIGURE 12.5



One of the strongest pieces of measured evidence for the big bang model of the universe (along with the observed expansion and cosmic microwave background) is this predicted proportion of hydrogen and helium. As astrophysicists measure the proportions of the elements in regions of the universe not greatly affected by later fusion in stars, the elements are found in this predicted abundance.

So we have the formation of nuclei in the early universe, but there are no atoms. That will take more time because although the universe has cooled sufficiently for hydrogen and helium nuclei to form, it is still way too hot for electrons to stay bound to those nuclei. Before the necessary cooling required to enable atoms to form, 380 000 years would pass.

An overview

It was not until 800 million years into the universe's life that the story of the atom resumed, when the first stars formed and new elements began to form via nuclear fusion in their interiors. In the centres of these enormous stars, the temperature and pressure was sufficient for long enough for fusion to continue beyond the formation of helium 4, resulting ultimately in the genesis of all of the elements in nature.

12.2 SAMPLE PROBLEM 1

1. The first atoms formed about 380 000 years after the universe began. What were the elements of these atoms and why did it take so long for them to form?
2. What determined the ratio of hydrogen to helium in the universe?

SOLUTION:

1. The first atoms were mostly hydrogen and helium with a trace of lithium. The nuclei for these atoms formed in the first minutes of the universe but the temperatures were too hot for the next 380 000 years for electrons to bind with the nuclei. Before that time, the electrons had so much kinetic energy that they could escape any nucleus they came near.

2. The ratio of protons to neutrons at the time was 1 neutron to every 7 protons. This came about because neutrons were formed when protons and electrons came together with sufficient energy. The energy of impact was not sufficient after a while, stopping the production of further neutrons through the interaction of a proton and an electron. Neutrons could still decay into a proton and an electron and the result was 7 protons for every neutron. When nucleosynthesis occurred over a period of about 17 minutes, all of the available neutrons underwent fusion with protons to form helium-4 nuclei. What remained was twelve protons (hydrogen nuclei) for every helium-4 nucleus, that is, twelve particles of hydrogen to four of helium or a universe of 75% hydrogen by mass and 25% helium.

The following table outlines significant events in the early universe.

TABLE 12.1 Significant events in the early universe.

Time since beginning of universe (seconds)	Temperature (K)	Event
0		Universe is born
10^{-36} to 10^{-32}		Inflation occurs
10^{-12} to 10^{-6}	10^{16}	Elementary particles including quarks and leptons form
10^{-6} to 10^{0}	10^{12}	Annihilation of antimatter and matter leave relatively small amount of matter
10^{2}	10^{9}	Commencement of nuclear fusion
10^{3}		Cessation of fusion
10^{13} (380 000 years)	3000	The formation of atoms (recombination), CMB produced
10^{13} to 10^{16}		The Dark Ages (stars yet to form)
10^{16} (800 000 000 years)		The first stars and galaxies form; most atoms re-ionised
10^{17} (9.3 billion years)		The Earth and solar system form
10^{18} (13.82 billion years)	2.7	Today

12.2 Exercise 1

1 When did the first nuclei form and why did they stop forming about 20 minutes after the universe began?
2 In the early universe, both matter and antimatter formed, which quickly annihilated each other. How did the universe end up being dominated by matter?
3 According to the big bang model, particle physics predicts that there would have been 75% hydrogen by mass and 25% helium by mass formed in the conditions present in the first 20 minutes of the universe. How does this prediction provide convincing evidence that the big bang model is accurate?

eBook*plus* RESOURCES

Watch this eLesson: From hydrogen to helium
Searchlight ID: eles-2459

12.3 The expanding universe

12.3.1 Discovering the universe of galaxies

In the early twentieth century, significant progress was made in our understanding of atoms and subatomic particles with the discovery of the nuclear atom made up of protons, neutrons and electrons. At the same time, our understanding of the universe was advancing at a tremendous pace. At the turn of the twentieth century, there was little grasp of how large the universe was. Astronomers knew that there were a very large number of stars and that these stars seemed to be clustered into a region of space known as the Milky Way. It was not yet clear whether the Milky Way was the extent of the universe, or just one population of stars among many.

FIGURE 12.6 The Milky Way.

In 1912, Henrietta Leavitt (1868–1921) made a discovery that was about to shine light on this puzzle. She investigated a type of star called a Cepheid variable. These stars vary in brightness over time in a regular way. Leavitt studied the two clouds of stars called the Large Magellanic Cloud and the Small Magellanic Cloud. She reasoned that the stars in each of these ‘clouds’ would all be approximately the same distance from us, meaning that any variation in their brightness could be attributed to an actual difference in luminosity, rather than just being the result of varying distance. As with all light sources, stars become dimmer the further away they are. Working with the stars from the Magellanic Clouds proved to be a very powerful technique. Leavitt plotted a graph of the maximum luminosity of the stars versus the period of their variation in brightness. She discovered a clear relationship between the luminosity and period: the brighter the Cepheid variable, the longer its period.

The period is the time it takes for one occurrence of a repeating event. For example, the period of the Earth rotating on its axis is 24 hours; the period of the motion of the second hand on a clock is 60 seconds.

FIGURE 12.7 Henrietta Leavitt.

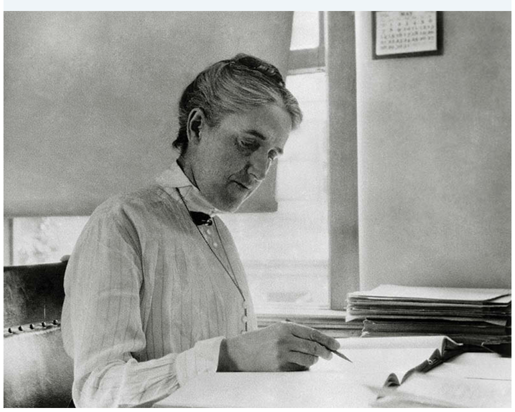


FIGURE 12.8 The vertical axis measures the brightness of the star. The horizontal axis shows increasing period plotted on a logarithmic scale; the scale is 10 to the power of the numbers on the axis.

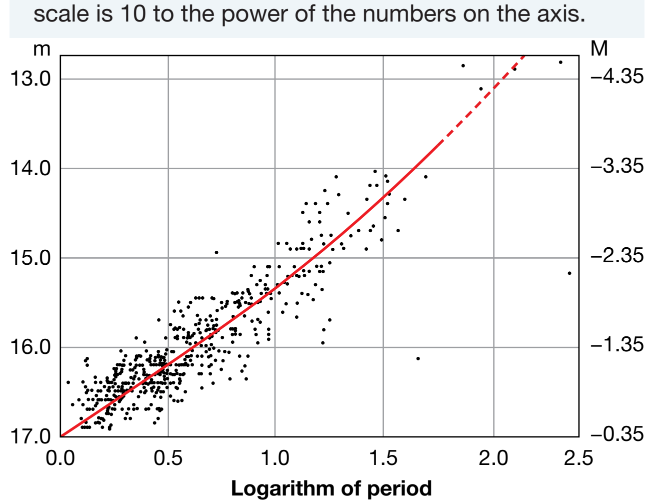


FIGURE 12.9 Globular cluster.

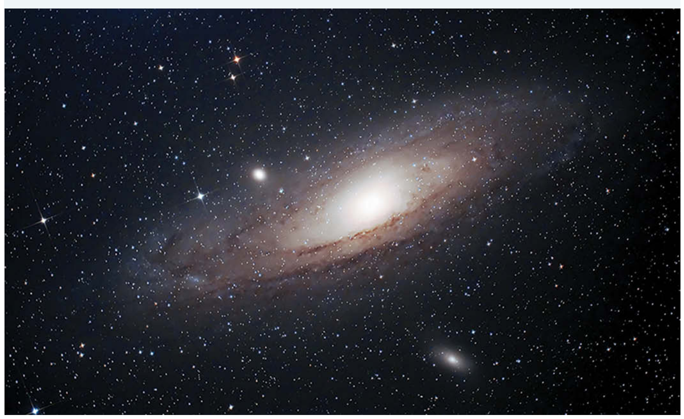
FIGURE 12.10 Edwin Hubble.

By 1919, Harlow Shapley (1885–1972) had used this relationship to determine how far the Earth is from groups of stars called globular clusters and the Large Magellanic Cloud. All he needed to do was measure the period of the variation in brightness of the Cepheid variable stars in these clusters and then use Leavitt's relationship to determine the luminosity of the star. Comparing this luminosity with the brightness he could measure revealed the distance to the stars, and hence the distance to the clusters they were in. Most of the stars in the Milky Way lie in a plane in the shape of a spiral. Shapley found that the globular clusters were not confined to the plane of the Milky Way, but rather cluster around its centre in a sphere.

Globular clusters are spherical clusters of thousands, and sometimes millions, of stars. Shapley used his mapping of globular clusters to determine the general shape and size of the Milky Way galaxy and found that the clusters formed a spherical shell whose centre lay in the direction of the constellation of Sagittarius. This suggested that the centre of the galaxy was in this direction and that our solar system was located towards the edge. Until Shapley's measurements, the solar system was thought to be near the centre of the galaxy.

Shapley's estimate of the size of the galaxy included the Magellanic Clouds and also assumed that the Andromeda nebula was within the Milky Way. He had made quite an error in his estimate. By this time thousands of nebulae in the shape of spirals and ellipses had been catalogued. Some of them seemed to be very small. How could they all be in our galaxy?

FIGURE 12.11 The Andromeda galaxy contains most of its stars in a flat disk similar to the Milky Way. The globular clusters are not restricted to this disc.



From 1919 to 1926, Edwin Hubble (1889–1953) examined the Andromeda nebula in great detail using the large telescopes at Mount Wilson, California. He took long exposure photographs of the spiral arms and counted numerous novae (singular nova; bright stars that were not in previous photographs) and Cepheid variables. Using these, Hubble established the size and distance of the Andromeda nebula from Earth, and found that it lay well beyond our galaxy and was comparable in size to the Milky Way. It seemed that our galaxy was only a tiny portion of the universe after all.

12.3 SAMPLE PROBLEM 1

Henrietta Leavitt discovered a feature of some types of distant stars that enabled the distance to those stars to be measured. How did this discovery change scientists' understanding of the universe?

SOLUTION:

The discovery of the relationship between the luminosity and the period of Cepheid–variable stars enabled astronomers to measure the distance to groupings of stars such as globular clusters, the Magellanic clouds and spiral nebulae to see whether they were part of our group of stars or so far away that they were part of separate groups altogether. This was how the Magellanic clouds and spiral nebulae were discovered to be separate galaxies from our own.

FIGURE 12.12 This view of the southern skies shows the two Magellanic Clouds, our near neighbour galaxies.

12.3.2 The expansion of space

One of the most significant results in the history of **cosmology** came from the work of Shapley and Hubble. Hubble correctly interpreted data that had earlier been used by Shapley. In 1919, Shapley noticed that the velocities of nebulae, later found to be galaxies, indicated that they were nearly all moving away from us. Hubble explored further and, in 1929, showed that the more distant the galaxies were from us, the faster they were moving away. The speeds were enormous; one measurement suggested that a galaxy was moving away from us at $42\,000\,\text{km s}^{-1}$ more than one-tenth of the speed of light!

This information was astonishing and very unexpected. Everywhere in the sky that astrophysicists looked, they found galaxies moving away from us. All galaxies are experiencing the same thing. If people in another galaxy looked at us, they would see us racing off in the opposite direction. In fact, it is not the galaxies themselves that are moving away; rather, space is expanding and taking the galaxies with it! Galaxies do also move through space due to gravitational interactions with other galaxies; we know that the Milky Way is currently colliding with the relatively tiny Sagittarius Dwarf Elliptical Galaxy. But these motions do not explain why galaxies are moving away from one another with a speed that increases with distance.

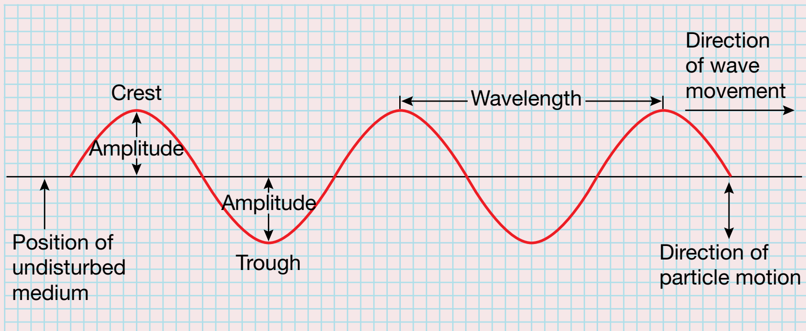
Galaxies and tightly bound clusters of galaxies do not expand, because they are held together by gravity. The Milky Way is part of a cluster of galaxies known as the Local Group. The Local Group includes over 30 galaxies, the two largest being Andromeda and the Milky Way. The Large and Small Magellanic Clouds are also members of the Local Group. The Andromeda galaxy is actually approaching us at enormous speed due to gravitational attraction, but as we look at more distant galaxies beyond the Local Group, the expansion of space dominates gravitational forces.

Hubble jumped to the obvious conclusion. All the galaxies are racing away from one another, so if we imagine running time backwards, we would see that they had all come from the same spot. It was as though the universe was born from some form of explosion that threw space and matter out in all directions. This was the first scientific evidence that the universe had a beginning. Prior to this, scientists assumed that the universe had existed forever.

REMEMBER THIS

Waves can be described in terms of their period, frequency, speed and wavelength. The wavelength is the distance between successive corresponding parts of a periodic wave — the length of one cycle. The period of a periodic wave is the time for one part of the wave to travel one wavelength. The frequency of a periodic wave is the number of cycles that occur every second.

FIGURE 12.13 Transverse periodic waves in a piece of string.



12.3.3 The Doppler effect

How did Hubble measure the expansion of space? He used a phenomenon called the Doppler effect. We are familiar with its effects on sound. When fast moving objects go by, the sound they make drops in pitch. You can hear this when trains, fire engines or racing cars speed past you. This change in pitch (or frequency) is known as the Doppler effect, after Christian Doppler, who predicted its existence in 1842, before it had been observed.

Doppler realised that, as light has a frequency like sound, it is also changed by this effect. Light emitted by a star or galaxy that is moving away from us is shifted towards the red end of the spectrum. This is commonly called the **red-shift**. The faster the galaxy moves away, the greater the shift. Stars or galaxies moving towards us have their light shifted towards the blue end of the spectrum, which is called blue-shift.

FIGURE 12.14 The Doppler effect: as the source moves to the right, the wavelengths in front of it are smaller than those behind it.

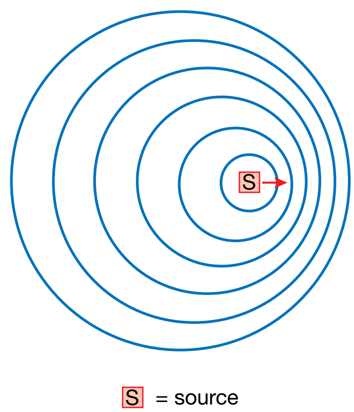
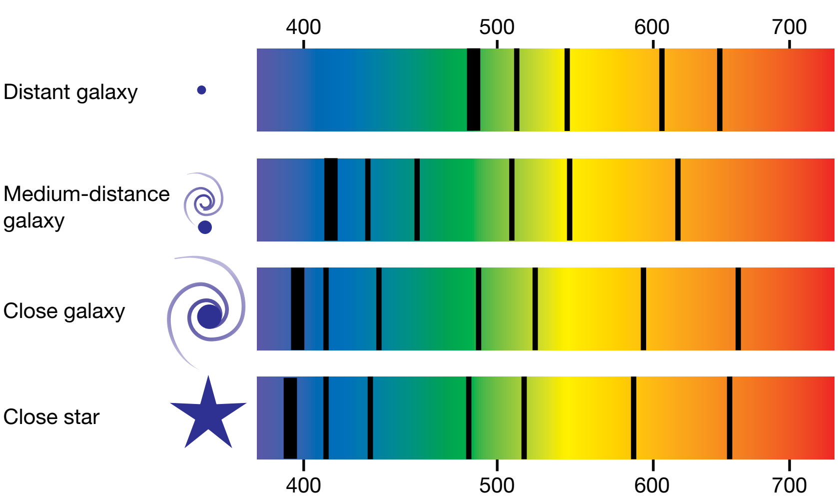


FIGURE 12.15 Spectra showing red-shift. The scale indicates the length in nanometres.



REMEMBER THIS

The visible spectrum of light contains red, orange, yellow, green, blue, indigo and violet. The spectrum continues into invisible forms of radiation, including infra-red at lower frequencies than red and ultraviolet at higher frequencies than violet.

Using the Doppler effect on stars

Spectra from stars contain numerous lines, which are characteristic of the gases they are made from. These lines can be identified by comparing them with the spectra of known gases (found by passing light through each gas at rest in a laboratory). When the light from a star is passed through a spectroscope, the characteristic pattern of lines observed may be shifted. A shift towards the red end of the spectrum indicates the star is moving away from us; a shift towards the blue end indicates that it is moving towards us.

12.3 SAMPLE PROBLEM 2

How was Hubble able to measure the motions of galaxies relative to us?

SOLUTION:

Hubble analysed the light from galaxies using a spectroscope. He observed that the light was red-shifted and, the greater the distance to the galaxy, the greater the red-shift.

Red-shift revisited

The Doppler effect is the same no matter how far the observer is from the source. The speed of the source determines the red-shift or blue-shift. What Hubble observed was that the red-shift was greater for more distant galaxies. According to the theory behind the Doppler effect, this must mean that these galaxies were moving away at much greater speeds.

But space is expanding. While the light from a distant source travels through space en route to Earth, the space it passes through stretches, increasing the light's wavelength. The light from distant galaxies takes many millions, if not billions, of years to reach Earth, and during this time the wavelength increases. The longer the light travels through space (that is, the further away the galaxy), the greater the increase in wavelength due to expanding space, and so the greater the red-shift.

Locally, the Doppler effect is more significant. Some galaxies are moving towards us and some away, under gravitational influences, but in the far reaches of the universe, the Doppler effect due to these local interactions is drowned out by the expansion of the universe.

This expansion effect can be quickly demonstrated using a rubber band to represent space. Mark a rubber band every 2 mm along a 2 cm length. The distance between neighbouring marks represents the wavelength of light. As you stretch the rubber band a little, you will see each mark move away from all the others (see Figure 12.17). As time goes on, the universe keeps expanding, like stretching the rubber band further. As a consequence, the wavelength of light from distant galaxies gets longer time.

FIGURE 12.16 A star with only tangential motion would not change its distance from the Earth. The radial component of its motion will move the star towards or away from the Earth.

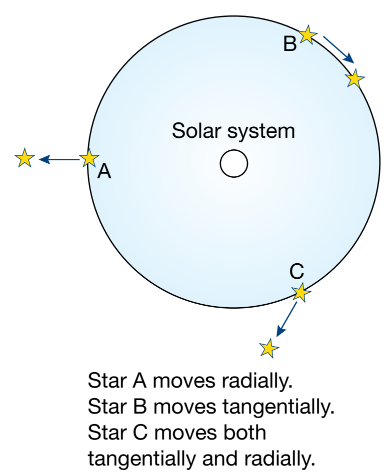
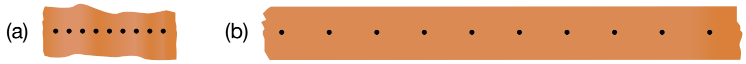


FIGURE 12.17 (a) The rubber band before stretching (b) the stretched rubber band; each dot has moved away from all the other dots.



12.3 SAMPLE PROBLEM 3

Explain why the observation that the red-shift is greater for more distant galaxies is consistent with red-shift being due to an expanding universe.

SOLUTION:

An expanding universe stretches the wavelength of light on its journey. The longer the light travels, the more it is stretched, so the more distant galaxies would show greater red-shift.

12.3 Exercise 1

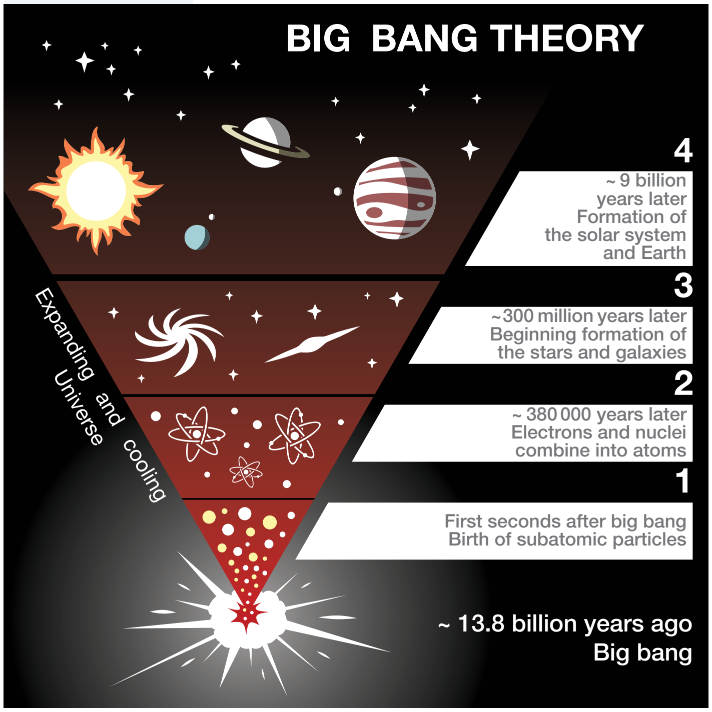
1 It is difficult to determine the luminosity of stars because their brightness depends on their distance from us, which is usually not otherwise known. How did Henrietta Leavitt control for distance when determining the relationship between the luminosity and the period of Cepheid–variable stars?
2 What evidence did Hubble have that the universe was expanding?
3 Does the greater red-shift observed for more distant galaxies mean that they are moving more rapidly away from us?

12.3.4 The big bang

By 1929 there was both a theoretical basis for this new cosmology from Einstein's general relativity and the observational evidence of the red-shift of galaxies to support it; however, it was by no means generally accepted.

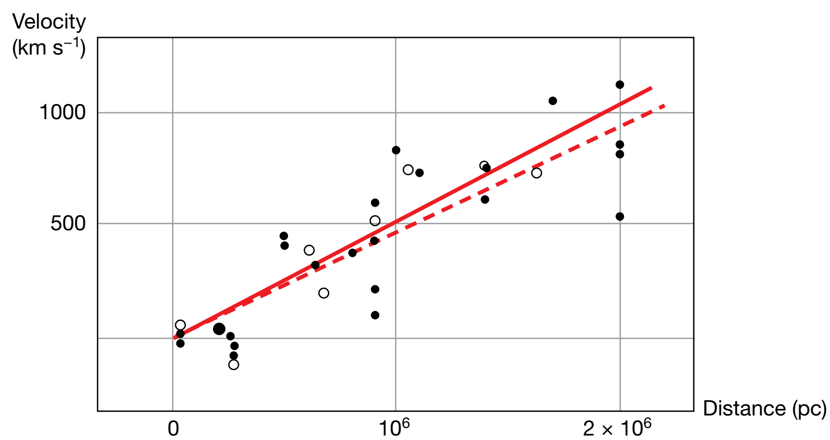
Hubble plotted the velocity (red-shift) of galaxies versus their distance from Earth and ambitiously fitted a straight line to the data, well aware that the distance calculations had large uncertainties. If the galaxies really did fit this straight-line rule, it would be easy to judge the distance to other galaxies; simply measure their red-shift and divide by the gradient of the line. This gradient became known as **Hubble's constant** (H) and the relationship between velocity and distance **Hubble's Law** ($\frac{1}{H}$ gives physicists a means of calculating the age of the universe). The value of Hubble's constant has been measured with increasing accuracy since Hubble's

FIGURE 12.18



time and discoveries in recent years have established its value to a small margin of uncertainty. Using the simplest scenario, that the universe has expanded at a constant rate, Hubble's early measurements put the age of the universe at 2 billion years, but the age of the Earth appeared to be greater than that. The age of the Earth has been dated using the proportion of radioactive isotopes in rocks, the rate of cooling from the Earth's original molten state, the time it would take to develop its geological features, and the time required for the evolution of life. These all pointed to an age greater than 2 billion years — the currently accepted figure for the age of the Earth is about 4.6 billion years.

FIGURE 12.19 Hubble's data. The solid dots are the results for galaxies treated individually and the solid line is the line of best fit for these data. The dashed line is fitted to the circles, which are the result of treating galaxies in clusters. One parsec (pc) is 3.09×10^{16} m or light-years. Notice the group of blue-shifted galaxies at about 2.5×10^5 parsecs. This corresponds to the Andromeda galaxy and its satellites.



Source: Adapted from Edwin Hubble, 'A relation between distance and radial velocity among extra-galactic nebula', *Proceedings of the National Academy of Science,* vol. 15, no. 3, 15 March 1929, Mount Wilson Observatory, Carnegie Institution of Washington. Communicated 17 January 1929.

It was not until the 1950s that Walter Baade identified two populations of stars that helped resolve the age problem. Baade noticed that Cepheid variables with significant amounts of heavier elements (Population I stars) had a different relationship between intensity and period than those made of little other than hydrogen and helium (Population II stars). This more than halved Hubble's constant, and therefore doubled the calculated value for the age of the universe. The adjustment resulted in a calculated universe age of 5 billion years; still young but at least it was older than the Earth. Improvements in measurement since have put the age of the universe at about 13.8 billion years.

12.3 SAMPLE PROBLEM 4

Use the solid line in the graph of Hubble's data to estimate the age of the universe. Compare this with his estimate of 2 billion years.

SOLUTION:

$$\begin{aligned} 2 \times 10^6 \text{ pc} &= 2 \times 10^6 \times 3.09 \times 10^{13} \\ &= 6.2 \times 10^{19} \text{ km} \end{aligned}$$

Hubble's constant is the gradient of the graph:

$$H = \frac{1100}{6.2 \times 10^{19}}$$
$$= 1.77 \times 10^{-17}\ s^{-1}$$

$$t = \frac{1}{H}$$
$$= \frac{1}{1.77 \times 10^{-17}}$$
$$= 5.64 \times 10^{16}\ s$$
$$= 1.8 \times 10^{9}\ \text{years}$$

This is 2 billion years to one significant figure, in agreement with Hubble's estimate.

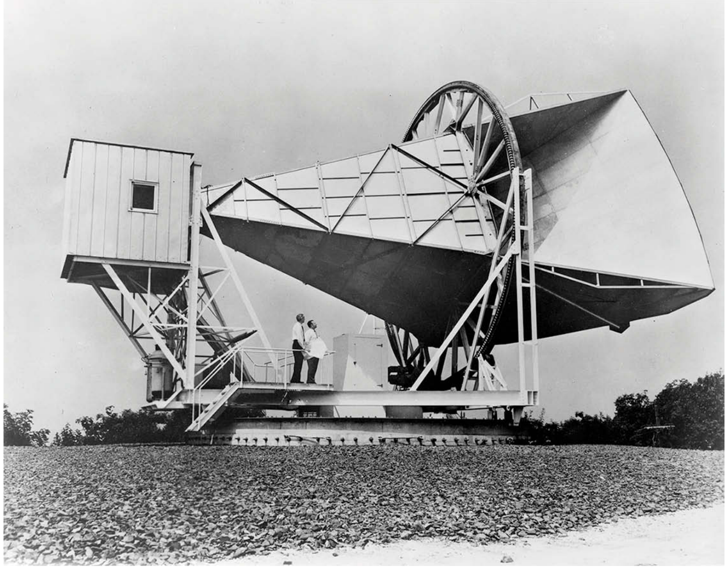
12.3 Exercise 2

1 Given the data in Hubble's graph, the currently accepted age of the universe, 13.8 billion years, is quite unexpected. What is Hubble's constant if the universe is 13.8 billion years old?

Other evidence for the expansion model of the universe included surveys showing that galaxy density in distant space was greater than the density of galaxies closer to Earth. This is expected with the expansion model because when we observe distant galaxies, we see them as they were billions of years ago, when the universe had undergone much less expansion.

When the universe cooled sufficiently for electrons to bond with nuclei forming hydrogen and helium atoms, an event called recombination, photons were able to travel freely through the universe without interacting with electrons. The wavelength of a photon depends on its energy. As the universe expanded, the wavelengths of the photons would have expanded, stretching out to form radio waves with much less energy than when the photons were initially released. These radio waves have become known as cosmic microwave background radiation (CMB) and correspond to a background temperature of about −270°C or 2.7 Kelvin. Arno Penzias and Robert Wilson discovered this radiation accidentally in 1965 when they were trying to eliminate some noise coming from their radio telescope. The identification of this noise was the turning point for the big bang theory.

FIGURE 12.20 Penzias and Wilson accidently discovered the cosmic microwave background radiation in 1965.



The name 'big bang' came from an early opponent of the theory, Fred Hoyle, in 1950. It is not a particularly accurate description as it implies that the creation of the universe was like an enormous explosion, which is not the case.

So in summary we have discovered that:

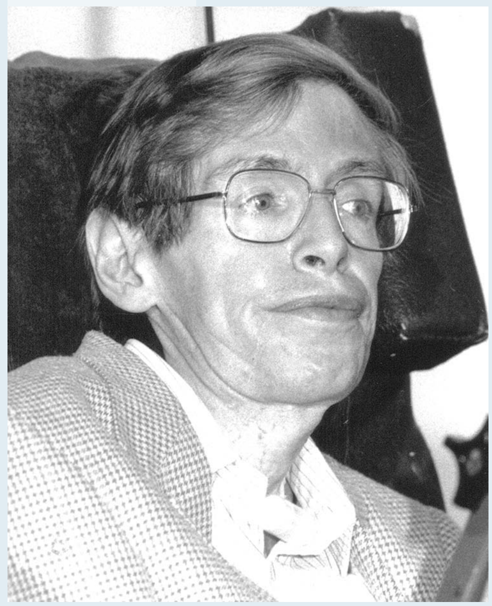
- the universe is vast, filled with billions of galaxies each containing billions of stars
- these galaxies are moving away from one another at a rate that increases the more distant the galaxies are
- the more distant galaxies (the older ones) are more densely packed than those nearer to us
- the predicted cosmic microwave background radiation has been detected.

Most scientists concluded that these discoveries meant that the universe had a beginning in a relatively small volume that has expanded ever since. This idea has been called the big bang theory.

PHYSICS IN FOCUS

Stephen Hawking and the big bang

FIGURE 12.21



The big bang theory raises many questions. What happened before the big bang? What is outside the universe? What is the universe expanding into? Famous physicist Stephen Hawking, in collaboration with others, has posed answers to these questions. He reminds us that, in the past, people contemplated what would happen if you sailed off the edge of the world. That question turns out to not need an answer because we have a completely different view of the shape of the world; it is a globe, so we never reach an edge. The question of what happened off the edge of the world only arose because of our misunderstanding of the shape of the world, thinking that it was flat.

In his PhD thesis in the mid 1960s, Stephen Hawking proved that Einstein's Theory of General Relativity required the universe to have a beginning in what was called a 'singularity'. A singularity is a point of infinite density that is achieved when we think of what happens if we run time backwards so that the expanding universe collapses back to its beginning. This was a stunning result for general relativity. If the universe was initially very dense, it must also have been incredibly hot and that energy should still be found throughout the universe. This energy was found soon after Hawking's initial work on singularities, in the form of the cosmic microwave background radiation. This radiation is central to the topic of this topic the origin of atoms.

Later, Hawking showed that time becomes like another dimension of space under extreme conditions. It makes no more sense to ask what happened before the big bang than to ask what is south of the South Pole. Hawking asks us to imagine the passing of time as being like decreasing degrees of latitude away from the South Pole. The latitude is 90° at the South Pole and as we move north it becomes 89°, 88° and so on. There is no 91° of latitude and there is no 'before the big bang'. Time began with the big bang and space has no edge to fall off, or to step outside of, to discover what the universe is expanding into.

eBook*plus* RESOURCES

Watch this eLesson: The expansion of the universe
Searchlight ID: eles-0038

Try out this interactivity: Shifting stars
Searchlight ID: int-6395

Explore more with this weblink: The expansion of the universe and Brian Schmidt

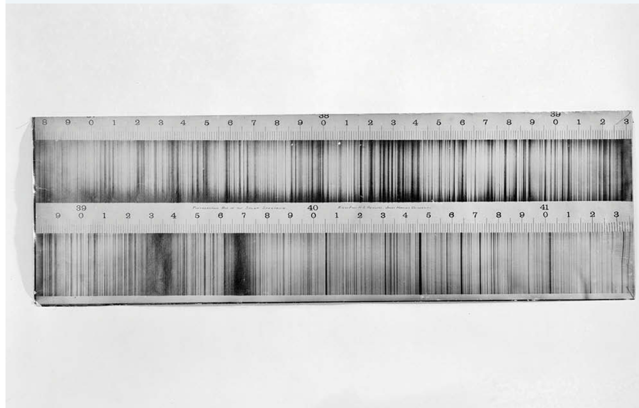
12.4 The power of stars

12.4.1 Stars and Einstein

The big bang theory provides an explanation for most of the atoms observed in the universe. It explains the abundance of hydrogen and helium, with about 75% of the mass of normal matter in the universe made up of hydrogen and 25% of helium, and a little lithium as well. This process of big bang nucleosynthesis described in Section 12.2 was all over in the first 20 minutes of the universe's existence. The big bang model roughly describes the abundance of atoms in the universe today, but we know there are many more atoms that exist that could not have been created by the processes of the big bang. The key elements that make us up, including carbon, nitrogen and oxygen would never have existed and neither would the heavier elements that make up our planet, such as silicon, iron and aluminium. Where did they come from?

The answer to this question answered another great scientific question: what powered the Sun and all the other stars? Early in the twentieth century, no one understood what could power the Sun. Chemical reactions do not produce sufficient energy. Scientists considered that maybe it was some form of nuclear reaction that could release some of the Sun's mass in the form of energy, as Einstein described in 1905 in the equation $E = mc^2$. However, scientists thought the Sun was made largely of iron, the most stable nucleus of all. The conclusion that the Sun was made of iron was based on measurements of tidal effects on the Earth that showed that the Earth was mostly made of iron, and on analysis of meteorites from space that showed that they too were composed largely of iron. This was prior to the big bang model. It was easy to assume that the Sun was made of iron. Even analysis of sunlight suggested it was made of iron. It was not until 1925 that Cecilia Payne (1900–1979) demonstrated that the Sun was mostly made of hydrogen, not iron at all. She did this using spectroscopy, a technique explained in Section 12.5.

FIGURE 12.22 The Rowland Solar Spectrum, 1886, depicting the Fraunhofer lines that were so difficult to interpret.



Stars form from clouds of atoms, mostly hydrogen and helium, collapsing due to the gravitational attraction between the atoms. This collapse results in a loss of gravitational potential energy of the atoms. This energy must be transformed into some other form of energy, and the result is the generation of a lot of thermal energy. The heat provides more than enough energy to the electrons in the Sun, the result being that the electrons are no longer bound to the nuclei. This creates a ball of plasma, ionised hydrogen and helium nuclei. A result of star formation in the early universe was that most of the atoms that formed in the era of recombination were reionised.

The heat has another consequence. The particles in the Sun move about with great energy until the star reaches a point of equilibrium where the collapse due to gravity is matched by the outward pressure due to the heating.

If the Sun was a ball of iron as previously thought, calculations have shown that the Sun's energy supply from gravitational collapse would only have lasted 15 million years. There was mounting evidence that the Earth was hundreds of times older than this and the Sun was showing no signs of cooling down. As the Sun was found to be a ball largely of hydrogen and helium, another energy source was available, hydrogen fusion. We will return to that in a later section.

Gravitational collapse by itself would only sustain the Sun radiating energy at its current rate for 15 million years. In Topic 11, we learned Einstein established that energy and mass are equivalent. $E = mc^2$ tells us that the energy radiated by the Sun, or any star, corresponds to a loss of mass. The Sun radiates 3.86×10^{26} Joule every second.

12.4 SAMPLE PROBLEM 1

Calculate the mass loss of the Sun every second due to the radiation of energy.

SOLUTION:

$$E = mc^2$$

$$\begin{aligned} m &= \frac{E}{c^2} \\ &= \frac{3.86 \times 10^{26}}{(3 \times 10^8)^2} \\ &= 4.29 \times 10^9 kg \end{aligned}$$

The Sun loses 4.29 billion kilograms every second due to the energy it radiates.

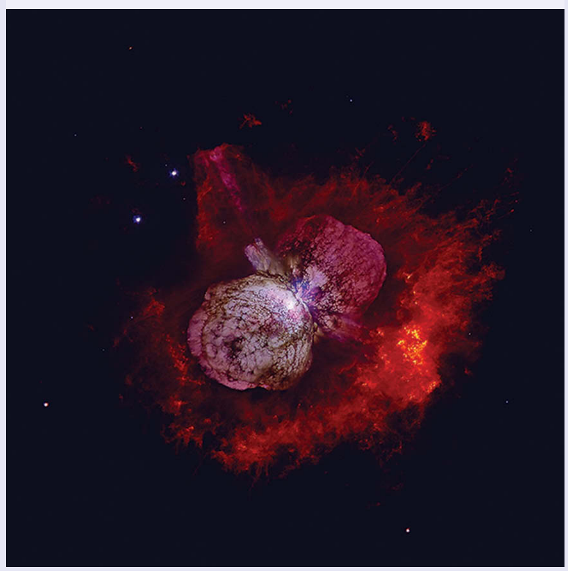
The amount calculated in sample problem 1 is the same whichever way the energy is produced. The mass of the Sun is 1.99×10^{30} kg, so if the Sun radiated at this rate and was able to convert all its mass into radiant energy, it would last 1.47×10^{13} years. However, the Sun is not converting all of its mass to radiant energy and is also constantly losing mass to the solar wind and gaining mass through collisions with other material. What process would sustain the Sun radiating at this rate for the 4.6 billion years that the Earth has existed and keep it going for whatever time the Sun still has left?

12.4 Exercise 1

1 One of the most luminous stars, Eta Carinae, is 80 times the mass of the Sun and 5 million times as luminous. If it and the Sun generate their energy in the same way and radiate energy at a constant rate throughout their lives, would you expect Eta Carinae or the Sun to last longer. Justify your argument using calculations.

But what is powering the Sun, and other stars like Eta Carinae?

FIGURE 12.23 Eta Carinae is a system of at least two stars enveloped in this nebula of material that it has shed.



12.4.2 Nuclear Fusion

Some sort of nuclear reaction was the most promising candidate for the powering of stars. Neither chemical reactions nor gravitational energy would sustain energy at the rate the Sun was producing it for long enough. In Section 12.2.3, we learned about how neutrons and protons underwent fusion in the early universe. We will now explore this fusion in greater detail.

When protons and neutrons come close enough, a force called the strong nuclear force binds them together to form a deuterium nucleus, which is hydrogen with a single neutron.

Nuclei are made of protons and neutrons, known collectively as nucleons. The number of protons is called the atomic number (Z) and this defines the element of the nucleus (Z = 1 is hydrogen, Z = 2 is helium and so on). The number of protons plus neutrons in a nucleus is called the mass number or nucleon number (A). Nuclei that share the same atomic number but have different mass numbers are called isotopes. These are represented symbolically by:

$${}^{A}_{Z}\mathrm{X}$$

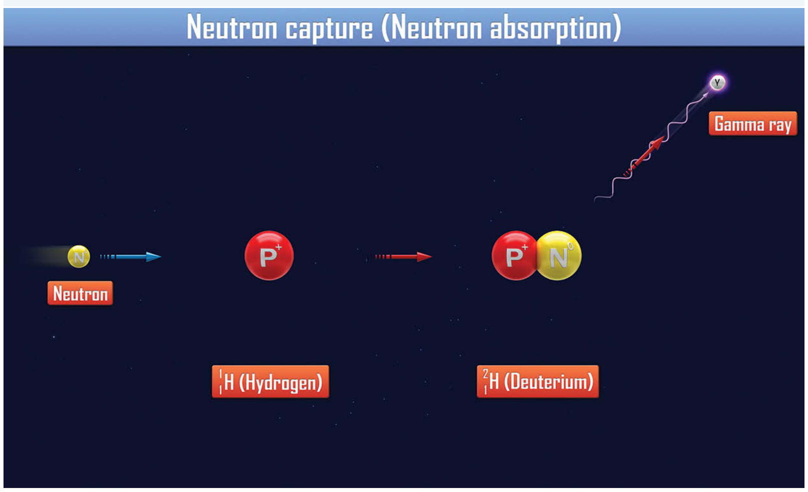
where X is the symbol for the element, A is the mass number and Z is the atomic number. In a nuclear reaction, the total of the mass numbers on each side of the equation must be equal. Similarly, the atomic numbers on each side of the equation must be equal.

In the early universe, when protons and neutrons existed in a ratio of 7 to 1 under conditions of extreme temperature and pressure, a reaction that occurred was:

$${}^{1}_{1}\mathrm{H} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{2}_{1}\mathrm{H} + \gamma$$

where n is a neutron.

FIGURE 12.24 Neutron capture.



The deuterons that formed (${}^{2}_{1}\mathrm{H}$) have less mass than the individual neutron and proton. The energy of the deuteron must be lower than the neutron and proton as separate particles because it requires energy to separate them. This energy is called the binding energy. In this fusion of nucleons, the binding energy is 2.2 MeV (1 electron volt = 1.6×10^{-19} J). That means that in the formation of each deuteron, 2.2 MeV was released. While 2.2 MeV is a small amount of energy, if large numbers of these fusions could occur in a star, that could be the source of energy.

However, this neutron and proton fusion requires a good supply of free neutrons, and free neutrons do not last long. There would need to be a source of neutrons. The protons in stars have existed from the early moments of the universe. The neutrons that survived from that period were those that were already captured by protons and formed deuterons and then mostly helium-4 nuclei. These are not the raw materials for this reaction to be the main power source of the Sun.

12.4 SAMPLE PROBLEM 2

Compare the mass of a free neutron with the total of a proton and electron. Use this to explain why free neutrons are unstable.

Data

Particle	Mass (kg)
proton	1.6726×10^{-27}
electron	9.11×10^{-31}
neutron	1.6749×10^{-27}

SOLUTION:

Mass of proton + mass of neutron $= 1.6726 \times 10^{-27} + 9.11 \times 10^{-31} = 1.6735 \times 10^{-27}$ kg, which is less than the mass of a neutron. This difference is called the mass defect. A free neutron requires more energy (equivalent to mass through $E = mc^2$) to keep it together than to simply fall apart (decay) into a proton and an electron.

A neutron bound in a deuteron cannot decay to form a proton, electron and neutrino because conservation of energy will not allow it. If we add the masses of the proton, proton and electron that would result, the total mass would be equivalent to 1.4 MeV more than the deuteron. It would be like a golf ball in a cup suddenly leaving the cup. The only way for this to happen is to remove the energy barrier (tip the cup on its side) or provide energy, perhaps by hitting the cup from underneath.

12.4.3 Fusion of protons

So nuclear fusion is a source of energy, but fusing neutrons and protons cannot be the major source of energy in stars as there are not enough free neutrons. What there is plenty of in the universe including in the centre of young stars is protons. Can protons fuse?

Protons exert an electrostatic force on one another that keeps them apart. To push them close enough together for the strong nuclear force to act would require extreme conditions. It turns out that these conditions are what are found in the Sun. There is another problem, though. The electrostatic force is very strong and when protons are close enough for the strong nuclear force to operate, this repulsion is stronger than the force of nuclear attraction. However, protons can emit a positron and neutrino in these high-energy conditions to become neutrons. It seems impossible for protons to decay into other particles, including one that has greater mass than the initial proton. However, this reasoning does not take into account $E = mc^2$ and the additional energy that is available from squeezing two protons so close together. What was impossible for the free proton is now likely to occur.

This gives the following fusion reaction:

$$^{1}_{1}\text{H} + {}^{1}_{1}\text{H} \rightarrow {}^{2}_{1}\text{H} + {}^{0}_{1}\text{e}^{+} + \nu$$

This reaction releases 0.42 MeV of energy. The positron will go on to annihilate an electron, and the neutrino will escape at near the speed of light. The conditions required for this reaction are present in the Sun.

12.4 Exercise 2

1. A deuteron has a mass of 3.3436×10^{-27} kg. Show that a deuteron has a smaller mass than the mass of a free neutron and free proton.
2. Use the difference in the mass of the deuteron and the free neutron and proton to find the binding energy of the deuteron (the energy required to separate the neutron and the proton).
3. Compare the combined mass of the two protons with the mass of the deuteron plus positron.

12.5 Analysing light from stars

12.5.1 Spectroscopy

Isaac Newton (1642–1727) published his book *Opticks* in 1704. In the first volume he demonstrated that light from the Sun can be dispersed into its constituent colours. Other theories about why rainbows formed, why prisms of glass produced a spectrum of colours and why soap bubbles appeared coloured involved the prism, raindrop or bubble altering the light. However, as Newton demonstrated, the prism, raindrop or bubble simply disperse the light according to its colour (wavelength), revealing information about the Sun. Newton's prisms showed the colour spectrum from the Sun to contain red, orange, yellow, green, blue, indigo and violet.

In 1802 William Wollaston (1766–1828) invented the spectroscope in an effort to explore the spectrum in more detail. He found the solar spectrum was not continuous but was crossed by a number of black lines. In 1814, Joseph von Fraunhofer (1787–1826) mapped the spectrum in much greater detail, finding 576 black lines. These have become known as Fraunhofer lines.

John Herschel (1792–1871), and W.H. Fox Talbot (1800–1877), a pioneer in photography, found that when chemical substances were heated in a flame and observed through a spectroscope, each chemical had a distinctive set of bright bands of colour forming its spectrum. This meant that scientists could identify chemicals simply by observing their spectra. Other scientists found that when sunlight is passed through gas before entering the spectroscope, it had extra black lines through its spectrum. This suggested that the black lines in the solar spectrum were due to light passing through gases in the Sun. These scientists had identified a method for determining the elements in stars.

FIGURE 12.25 A continuous spectrum produced by a spectroscope.

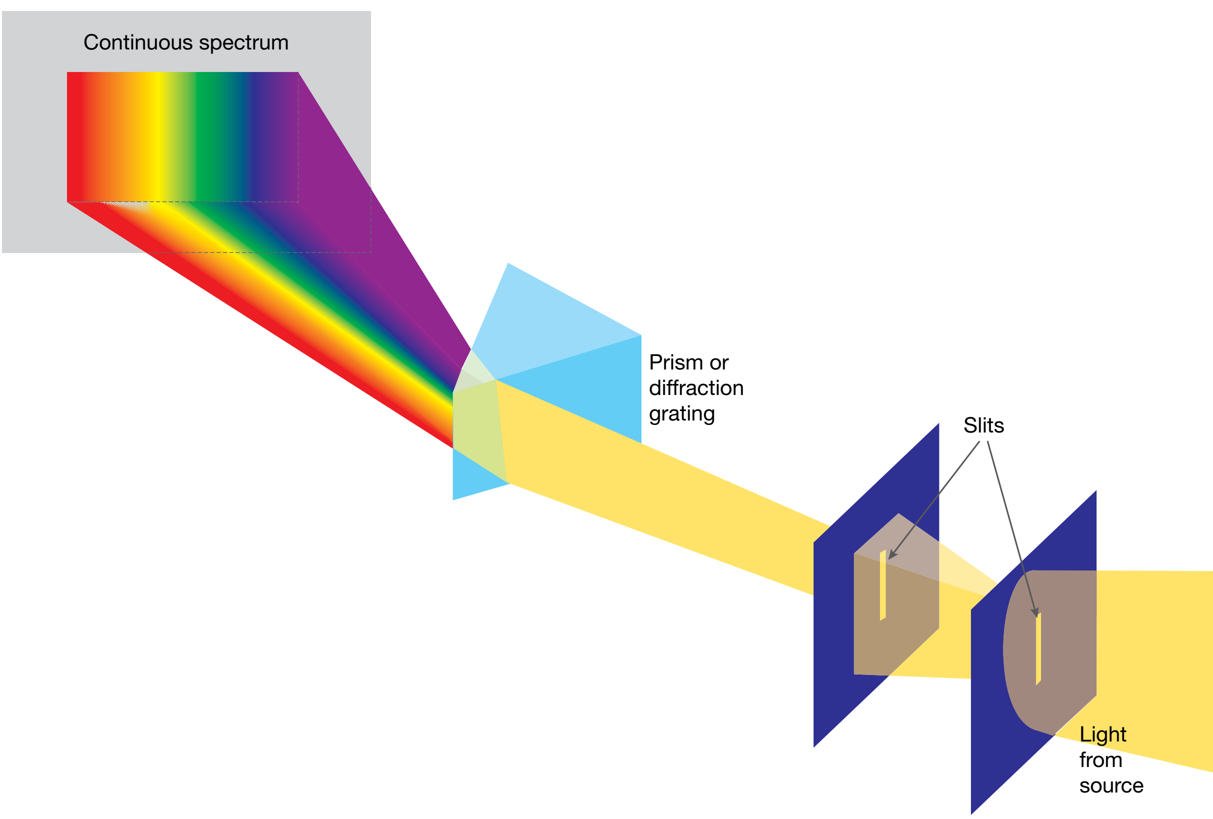


FIGURE 12.26 An absorption spectrum produced by shining light with a continuous spectrum through a cool gas.

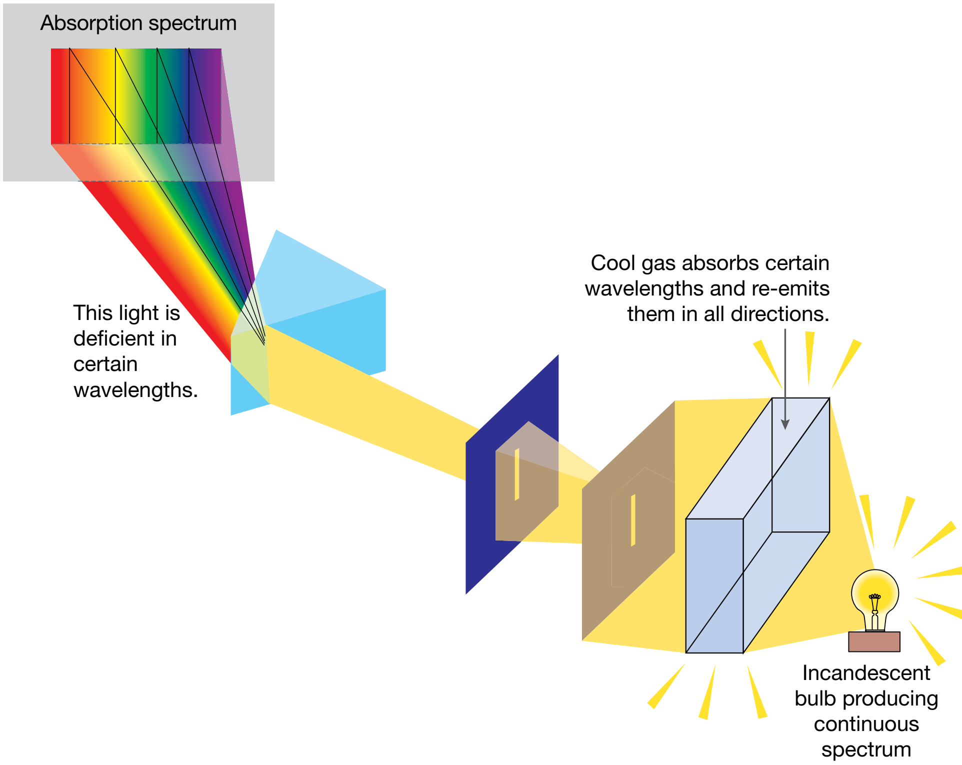


FIGURE 12.27 The absorption spectrum for hydrogen.

In 1859 Gustav Kirchhoff (1824–1887) with his friend Robert Bunsen (1811–1899) used Bunsen's burner to burn elements and clearly describe the cause of these spectral lines. They found that:

- a continuous colour spectrum is produced by glowing solids or dense gaseous bodies such as the Sun, which behaves very much like black body radiation
- if a gas is between the light source and the spectroscope, light is absorbed from the continuous spectrum at wavelengths or colours characteristic of the chemical components of the gas
- a glowing gas produces bright lines on a dark background at wavelengths or colours characteristic of the chemical components of the gas.

One of the first successes with this new tool of astrophysics was the spectroscopic analysis of planetary nebulae by William Huggins (1824–1910). Working in London in 1864, he found that these nebulae produced the bright line spectra of glowing gas, showing that they were clouds of gas rather than groups of very distant stars (see Table 12.2). Some of the nebulae documented by Hershel, however, showed that they were collections of stars, as they emitted continuous spectra interrupted by black lines. Huggins's investigations also provided very convincing evidence that the stars in the sky really are distant suns.

TABLE 12.2 Types of spectra and the celestial bodies that produce them.

Type of spectrum	Produced in the laboratory by	Celestially produced by:
Continuous blackbody	Hot solids, liquids, gases under pressure	Galaxies, inner layers of stars
Emission	Incandescent low-density gases	Emission nebulae, quasars
Absorption	Cool gases in front of continuous spectrum	The atmospheres of stars

12.5 SAMPLE PROBLEM 1

Describe how astronomers can determine the elements that are present in the Sun.

SOLUTION:

Astronomers observe the Sun's light through a spectroscope. The solar spectrum contains absorption lines that are characteristic of the elements found in the Sun's atmosphere.

12.5 Exercise 1

1 The Sun is a hot ball of gas. Explain why it does not produce an emission spectrum.

12.5.2 What produces the spectra?

We know that electrons move around the nucleus of an atom. When the atoms are part of a solid, or when the material is very dense, like the gases in a star, the electrons can exist with a range of energies. When the material is heated, the electrons gain energy. They then re-emit this energy in the form of packets of light energy called photons and fall back to a lower energy state. This process results in the emission of light with a range of energies or frequencies. We see this in the light emitted by hot wires in incandescent globes. The light produced has a continuous spectrum when passed through a spectroscope. The relative intensities of the different wavelengths of the spectrum determine the colour of the object. The hotter the object, the more the highest intensity wavelengths shift towards the blue end of the spectrum. Topic 14 addresses this process in more detail.

Absorption spectra

Atoms and molecules in gaseous form allow their electrons to take on only particular energies. When light with a continuous spectrum passes through a gas, most of the light will pass straight through. This is because most of the light is not at the specific energies that the electrons in the gas can absorb. The particular energies that are absorbed are unique to the element. When light with a continuous spectrum passes through a gas, these energies are absorbed and so are removed from the spectrum. When the light is passed through a spectrometer, it forms an absorption spectrum. The energies that have been absorbed indicate the elements that the light has passed though.

This is important for our understanding of stars, because the hot dense plasma at the centre of a star produces light with a continuous spectrum, behaving like an ideal black body radiator. As the light moves out from the centre, the gases in the star's cooler atmosphere absorb specific frequencies and produce light with an absorption spectrum. Analysis of this spectrum on Earth shows us what elements are present in the atmosphere of the star without having to go there to take samples.

Emission spectra

When a gas is heated, the electrons increase their energies to higher energy states. These higher energy states are temporary and energy can be released at any time by an electron dropping back down to one of the lower energy states allowed by that element. This process produces light of very specific frequencies — an emission spectrum.

Sometimes it is difficult to determine what fuzzy light sources in space are. Are they galaxies too distant to distinguish the individual stars, or are they clouds of gas or nebulae, glowing because they are being energised by nearby stars? Passing the light through a spectroscope answers this question. If the spectrum is an absorption spectrum, we know that it is produced by stars. If it is an emission spectrum, we know that it is produced by diffuse clouds of gas. This technique was important in the discovery of galaxies.

Consider the spectra in Figures 12.28 and 12.29.

FIGURE 12.28 The spectra of 13 different stars. The stars increase in temperature from bottom to top. The top star is bright at the blue end of the spectrum, but emits very little light at the red end of the spectrum. It is a blue star. The star at the bottom emits very little light in the blue wavelengths. It is a cool red star. The absorption lines are visible in the spectra.

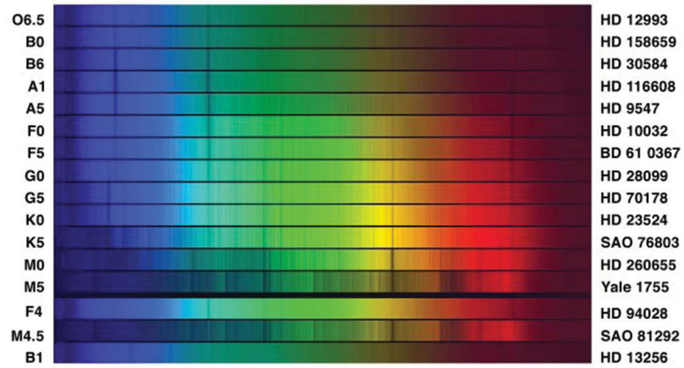
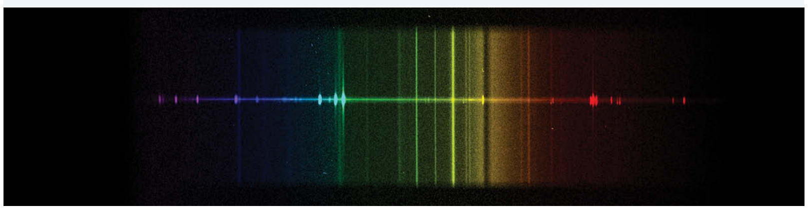


FIGURE 12.29 The spectrum of light from a nebula. Notice that it is an emission spectrum, distinguishing it from the absorption spectra of stars.



12.6 Classifying stars by their light

12.6.1 Spectral type

When the spectra of stars were first observed in detail in the nineteenth century, it seemed that the spectrum of every star was different. Gradually, some sense was made of the multitude of lines that crossed the spectra and stars were classified into spectral types. The system developed by Annie Jump Cannon (1863–1941) has been used since 1910. It classes stars as O, B, A, F, G, K or M according to the relative intensity of various absorption lines in their spectra. For example, for type A stars the lines of the hydrogen spectrum are very clear. The spectral classes are arranged in order of temperature from O, the hottest with a spectrum peaking in the ultraviolet, to M, the coolest with a spectrum peaking in the infra-red. The Sun is a type G star and these are yellow. A full description of the spectral classes is given in Table 12.3.

12.6 SAMPLE PROBLEM 1

Describe what happens to the wavelength and intensity of light from a star as its temperature increases.

SOLUTION:

The wavelength of the light emitted becomes shorter and the intensity increases.

TABLE 12.3 Spectral classifications and their corresponding features. Note that in astronomy the term 'metal' refers to any element other than hydrogen or helium.

Spectral class	Surface temperature (K)	Colour	Spectral features
O	28 000–50 000	Blue	Ionised helium lines Strong UV component
B	10 000–28 000	Blue	Neutral helium lines
A	7 500–10 000	Blue-white	Strong hydrogen lines Ionised metal lines
F	6000–7500	White	Strong metal lines Weak hydrogen lines
G	5000–6000	Yellow	Ionised calcium lines Metal lines present
K	3500–5000	Orange	Neutral metals dominate Strong molecular lines
M	2500–3500	Red	Molecular lines dominate Strong neutral metals

12.6 SAMPLE PROBLEM 2

You observe a red star. Estimate its surface temperature and spectral class.

SOLUTION:

The surface temperature is in the range from 2500 to 3500 Kelvin. It is an M class star.

The naming of the spectral classes seems strange and confusing today, but the naming came before the underlying physics was understood. The classes were arranged in this order once the relationship between spectral class and temperature was discovered using the technique for determining the surface temperature of stars, described earlier, in Section 8.7.1 of Topic 8. To memorise the sequence OBAFGKM, many have used the cheeky mnemonic Oh Be A Fine Girl/Guy, Kiss Me.

The spectral classes are each divided into 10 classifications by a number from 0 to 9 following the spectral class letter. For example, the Sun is classified as a G2 star, which makes it a little hotter than a G1 and a little cooler than a G3.

12.6 SAMPLE PROBLEM 3

One of the brightest stars in the sky is the closest to us. In fact it is a system of three stars called Alpha Centauri. These three stars in order of size have the following spectral classes:
Alpha Centauri A, G2
Alpha Centauri B, K1
Proxima Centauri, M5.

1. Which is most like our Sun?
2. Which is the dimmest?
3. What colour is Alpha Centauri B?

SOLUTION:

1. Alpha Centauri A has the same spectral class as the Sun so has very similar temperature. It is very similar in size and stage of life.
2. Proxima Centauri is the dimmest as it is the coolest and smallest. In fact, it is so dim, you need a powerful telescope to see it.
3. Orange.

12.6 Exercise 1

1 Describe how astronomers can determine the surface temperature of a star.
2 List the following stars from coolest to warmest.

Star name	Spectral Class
Sirius	A1
Canopus	A9
Arcturus	K1
Vega	A0
Rigel	B8
Betelgeuse	M5
Antares	M1

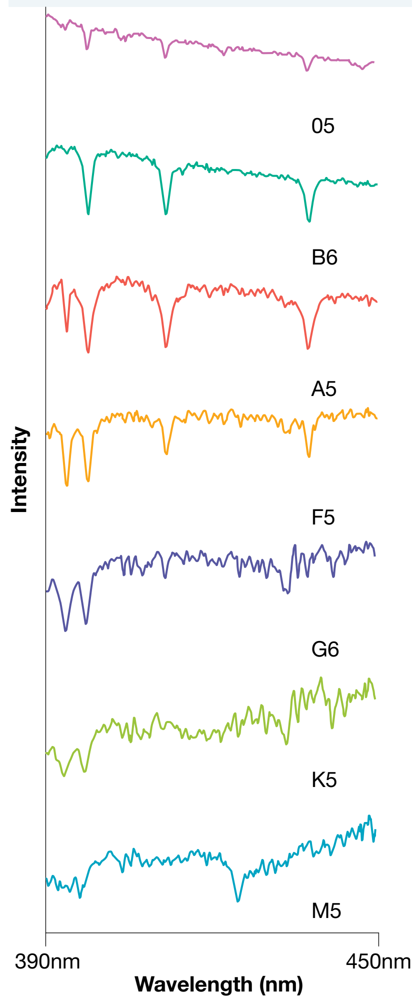
3 Identify the spectral class (OBAFGKM) of the star with a surface temperature of 6100 K.

Figure 12.30 shows some examples of graphs of the intensity of light emitted from different stars in the range of wavelengths 390 nm to 450 nm. Each graph is labelled with the spectral class that astronomers allocate to it based on this information. Notice that as the spectral class changes from M to O, the intensity of the light increases — hotter stars in this selection are brighter. Also notice that the graphs look different for each class. When first observed, the astronomers did not know the temperatures or how far away the stars were for them to know their absolute intensities, but they could see that different features were present in different stars and grouped them accordingly, forming the spectral classes.

The dips in intensity in each of the graphs depict wavelengths of light that have been partially absorbed as the light passes through the outer layers of the star. The interior of the star as a dense ball of plasma closely resembles a black body and emits a black body continuous spectrum. The outer layers, however, are cooler and less dense. They contain atoms at various stages of ionisation. The hottest stars, the O class, are hot enough to have helium that has lost one of its two electrons in the gaseous outer layers of the star. These stars have intensity spectra with dips in intensity at the wavelengths that are absorbed by ionised helium.

Cooler stars have different ionised states of atoms, and the coolest even contain some simple molecules. This provides a large number of dips in intensity in the spectrum, making them difficult to interpret. This was the case with the Sun when the many spectral lines confused astronomers into thinking that the Sun was mainly composed of iron, rather than being mostly hydrogen and helium.

FIGURE 12.30 Comparison of the 1D intensity spectra for Main Sequence stars.



Comparison of intensity spectra for stars of different spectral classes.

12.6 SAMPLE PROBLEM 4

On what basis were stars assigned to different spectral classes?

SOLUTION:

When the spectra of different stars were analysed, astronomers noticed that some stars shared features in their spectra, while other stars had quite different features. They grouped stars according to the ones that had similar spectral features.

12.7 Hertzsprung–Russell Diagrams

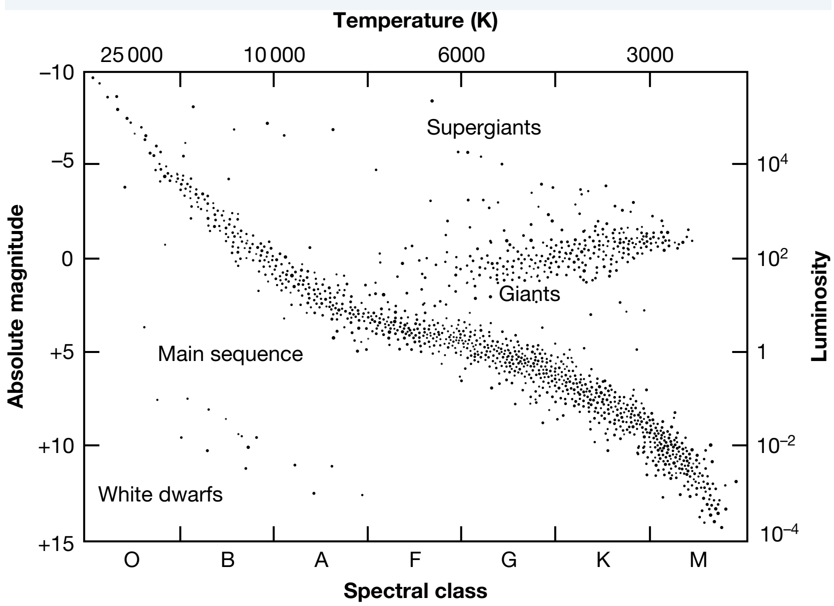
12.7.1 H–R diagrams

In 1911 in Denmark, Ejnar Hertzsprung (1873–1967) plotted star luminosities (or equivalently, absolute magnitudes) versus their spectral types (or equivalently, temperatures). In 1913, Henry Norris- Russell (1877–1957) did the same thing at Princeton. The result is known as a Hertzsprung–Russell (H–R) diagram. H–R diagrams provide a wonderful synthesis of the data that we have about stars.

Stars on the left side of the H–R diagram are hot and those on the right are cool. Those towards the top of the diagram are very luminous, while those towards the bottom are dim. Most stars cluster along a diagonal line from the bright and hot down to the dim and cold. These are known as **main sequence stars**. They are all doing much the same thing, fusing hydrogen in their cores, just like the Sun. The hotter, brighter stars are more massive than the cooler, dimmer stars. Stars lie on the main sequence for most of their existence.

Many stars do not lie on the main sequence. Above the main sequence are stars that are brighter than their spectral class would predict for a main sequence star. These stars are very large and are known as giants and supergiants. Many are red because of their cooler surface temperature, but are still very bright because they have such a large radiating surface. These stars can be quite unstable and their luminosity varies with time. They are stars that were once on the main sequence but are now swollen by heating that is now occurring in their outer layers. The Sun will eventually become a giant.

FIGURE 12.31 An H–R diagram showing various star types.



Below the main sequence is a cluster of hot, dim stars known as white dwarfs. White dwarfs are made mainly of carbon, and fusion has ceased in their cores. They are simply cooling down.

The stars on the right-hand end of an H–R diagram (spectral class O, B and A) are blue in colour due to their temperatures. Stars classed F are white, and G are white to yellow. Those classified K and M are red. In fact, main sequence M stars are also known as red dwarfs because of their size and colour.

The H–R diagram shows that while spectral class categorises stars by their temperatures, there are other groupings of stars. This could be seen in the intensity spectra too. Some stars had the same temperature but the dips or lines in the spectra had different widths. These widths corresponded to stars that radiated different amounts of light — that is, they had different luminosities. Another level of classification was added, as shown in Table 12.4. The Sun is actually a G2 V star: G2 tells us the temperature, with V telling us what type of G2 star it is.

TABLE 12.4 Luminosity classes.

Symbol	Class	Example
0	Extreme, luminous supergiants	
I	Luminous supergiants	Betelgeuse
II	Bright giants	Canopus
III	Normal giants	Aldebaran
IV	Subgiants	Procyon
V	Main Sequence	Sun
sd	Sub dwarfs	
wd	White dwarfs	

12.7 SAMPLE PROBLEM 1

1. Describe the luminosity and colour of a star located in the upper right-hand portion of the H–R diagram.
2. If you had two stars on the H–R diagram that shared spectral class O, what could you say about their relative surface temperatures?
3. Where would Sun-like stars lie on the H–R diagram at the end of their existence?
4. Where are the most massive stars on the H–R diagram?

SOLUTION:

1. It has high luminosity and is red.
2. They have very similar surface temperatures.
3. At the bottom right. They have cooled off and emit little light.
4. At the top on the far left. They generate the most heat and have large surface areas radiating the light, so are highly luminous.

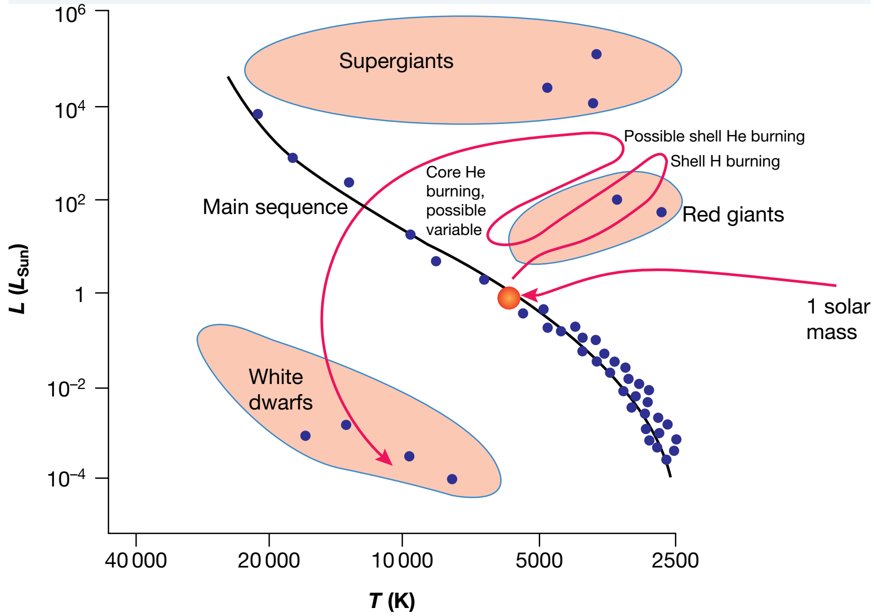
12.7.2 Population I and population II stars

During World War II, Walter Baade discovered that there appeared to be yet another classification of stars. He called these classes Population I and Population II stars. Population I stars contain a greater variety of elements, while Population II stars contain little more than hydrogen and helium. Population I stars tend to be hot and blue, and all fit on the main sequence, whereas Population II stars are much more varied. Many are red giants and the top left corner of the main sequence in the H–R diagram is nearly empty of Population II stars. The Sun is a Population I star.

The explanation for these two types of star is that Population I stars are young. They have more heavy elements in them because they are formed from the remnants of older stars, which produced heavy elements (helium and heavier) during the processes of fusion. Population I stars are all on the main sequence because they are still fusing hydrogen in their cores.

The most massive stars are found in the top left corner of the main sequence. This area of the H–R diagram is nearly empty of Population II stars. Massive stars have a much greater pull of gravity on the gases that they are made from, so equilibrium between gravitational collapse and the radiation pressure from the core is reached at much higher temperatures than in the smaller stars. This accelerates the rate at which fusion occurs and so these stars ‘live fast and die young’. The most massive Population II stars have had time to complete the fusion of hydrogen in the core and move on to fusing helium. Few Population I stars, however, are old enough to have completed this stage of their lives. By identifying the two star populations, Baade had discovered a way for astrophysicists to view star populations at two points in time. This opens a window of research as stellar evolution takes billions of years, so it is not possible to just sit and wait to see what happens.

FIGURE 12.32 This H–R diagram shows the 'path' of a star of one solar mass throughout its lifetime.



12.7.3 Stellar death

When the hydrogen in the core is exhausted, the star moves off the main sequence as it cools and expands into a red giant. Fusion of helium in the core begins. Once the helium in the core is exhausted, fusion of helium around the carbon core will cause the star to expand. In larger stars, fusion of nuclei can continue up to iron-56, but beyond iron it takes more energy to cause fusion than the fusion reaction produces, so the cycle stops. While fusion is occurring in shells around the core, the star pulses as a variable star. Due to its size and temperature it remains high above the main sequence on the H–R diagram. When a star about the size of the Sun throws off most of the remaining hydrogen and helium to form a planetary nebula, it can no longer sustain fusion and cools as a white dwarf, now below the main sequence as it is dim but still quite hot. White dwarfs are only about the size of Earth, but are much denser. They slowly cool, moving to the right in the H–R diagram as they become cold black dwarfs. These processes are summarised in Table 12.5.

TABLE 12.5 Fusion in different star types.

Star group	Energy-producing reactions
Main sequence	Nuclear fusion of H to He in core
Red giants	Nuclear fusion of He to C in core, with H fusion continuing in shell
Supergiants	Multiple nuclear fusions possible in shells, forming elements up to iron in core
White dwarfs	No nuclear fusion reactions occurring

12.7 SAMPLE PROBLEM 2

A star is observed to have a luminosity 1000 times that of the Sun and a surface temperature of about 10 000 K. Use an H–R diagram to predict what stage of evolution this star is in.

SOLUTION:

This star is on the main sequence, so it is fusing hydrogen in its core.

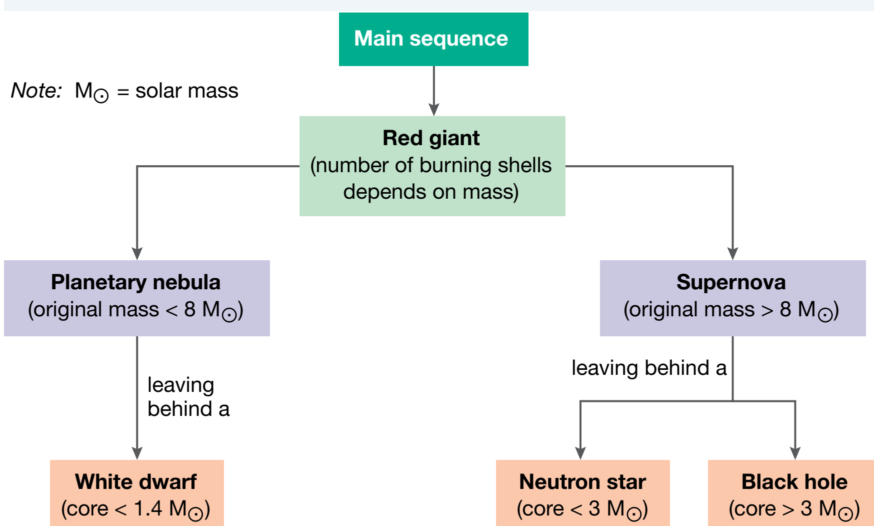
12.7 Exercise 1

1 Two stars have luminosities 10 000 times that of the Sun. One is a main sequence star and the other is a supergiant. Compare the temperatures, masses and ages of the stars.

More massive stars have a violent end. When fusion ends in these stars, they start to collapse very rapidly. This process leads to enormously high temperatures and more fusion, including the formation of elements heavier than iron-56. The outer layers come crashing inwards and bounce off the core in a **supernova**, blasting a rich soup of elements into space. A supernova can outshine a whole galaxy for a period of time. Stars between 8 and 50 solar masses end in this way and are called Type II supernovae. They are mostly found in the spiral arms of galaxies, where stars form. These massive stars have lifetimes in the millions of years, rather than the billions of years of solar mass stars. Their death provides the material for new stars in the galaxy that contain elements other than hydrogen and helium; the stars that could form planets like those in the solar system.

Type I supernovae are found in all parts of all types of galaxies. They form from old, low mass stars. Stars with less than 8 solar masses lose a lot of their mass as they form a planetary nebula following the red giant phase. They are too small to go supernova at this stage, so begin to cool as white dwarfs. Many stars, however, are in a binary system. If a white dwarf has a partner that is a red giant in a close orbit, the white dwarf will **accrete** hydrogen gas from the giant star. As the hydrogen falls onto the white dwarf, it undergoes fusion, heating the surface of the star rapidly and causing a nova, where the star glows a million times as brightly as the Sun. If the white dwarf is able to accrete sufficient hydrogen to have a total mass of more than 1.4 solar masses, it will collapse. Tremendous heat and pressure is generated, causing fusion reactions on a massive

FIGURE 12.33 Stellar evolution.



scale throughout the star and releasing incredible amounts of energy. These Type I supernovae are thousands of times brighter again than the nova.

With this energy source rapidly depleted, the star collapses. This time the mass is too great for protons and electrons to remain apart. The electrons and protons combine to form neutrons. The density of the resulting neutron star is the same as the density of atomic nuclei. The whole star has a diameter of little more than 10 km, but it has a mass greater than that of the Sun.

It is easy to distinguish Type I and Type II supernovae by their spectra. Type I supernovae have very little remaining hydrogen so the lines for hydrogen are missing from their spectra, while hydrogen lines are clear in Type II. Supernovae are rare, occurring at a rate of only about one per century per galaxy, with the last in the Milky Way recorded by Kepler in 1604. In 1987, a supernova was observed in the Large Magellanic Cloud. Photos of the cloud taken before the supernova showed the star before the massive explosion. It offered a wonderful opportunity to study supernovae and featured in the popular press.

If the core of the star that remains following a supernova has a mass of more than about 3 solar masses, gravity is unstoppable. Without the energy supply of fusion to hold the star up, no known force can support the matter against the intense gravity and the remnant collapses to form a **black hole**. These exotic objects have a gravitational pull so great that even light cannot escape.

FIGURE 12.34 The Crab Nebula, the remnant of a supernova observed in 1054; at the centre a neutron star sweeps a beam of radiation past us as it rotates. These types of neutron stars are called pulsars.

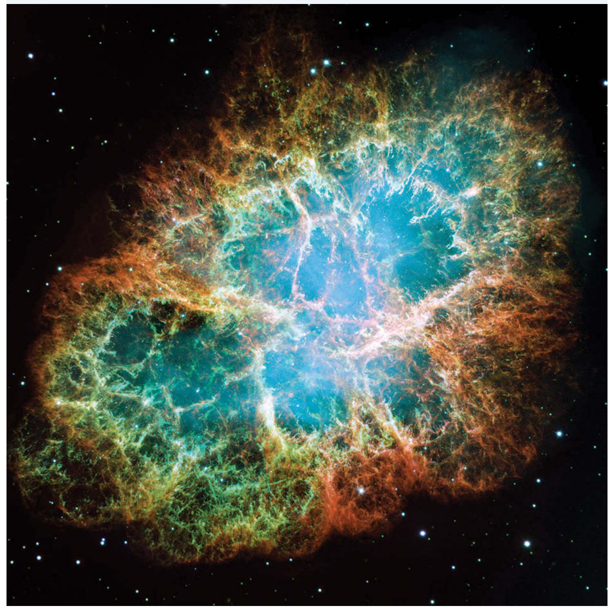
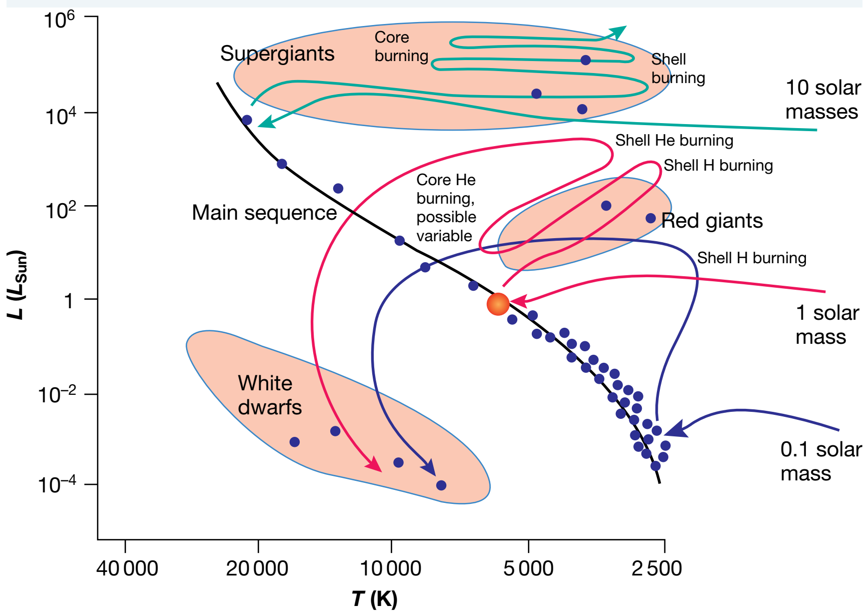


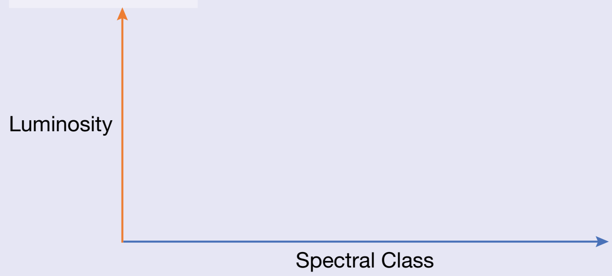
FIGURE 12.35 Stars of different masses evolve in different ways.



12.7 Exercise 2

1 On the following sketch of an H–R diagram, draw a horizontal line and a vertical line through it to divide it into four quadrants. Use the words 'hot' and 'cool' in each quadrant to describe the stars found there. Use the words 'bright' and 'dim' to describe the luminosity of the stars in each quadrant.

FIGURE 12.36



12.8 Where the atoms are made

12.8.1 Atom factories

Early in this topic, we learned how the first atoms in the universe formed, about 380 000 years after the universe began. The hydrogen, helium and trace of lithium that was produced then still accounts for most of the atoms in the universe, although a significant proportion of these have been stripped of their electrons again in the formation of stars. What about the other nearly 90 elements found in nature?

Cecilia Payne's discovery that the Sun was mostly hydrogen, and Einstein's $E = mc^2$ formula, led Arthur Eddington in the 1920s to propose that the fusion of hydrogen to form helium was the missing energy source of the Sun. By 1932, Australian physicist Mark Oliphant observed heavier nuclei being formed from protons in a particle accelerator. Finally, in 1939 Hans Bethe published his calculations of how energy could be produced in the Sun by the fusion of protons. This work won him the Nobel Prize in Physics.

Elements up to iron can be formed in the cores of stars through fusion, if the star is big enough.

12.8.2 Nuclear fusion: the enduring energy source of stars

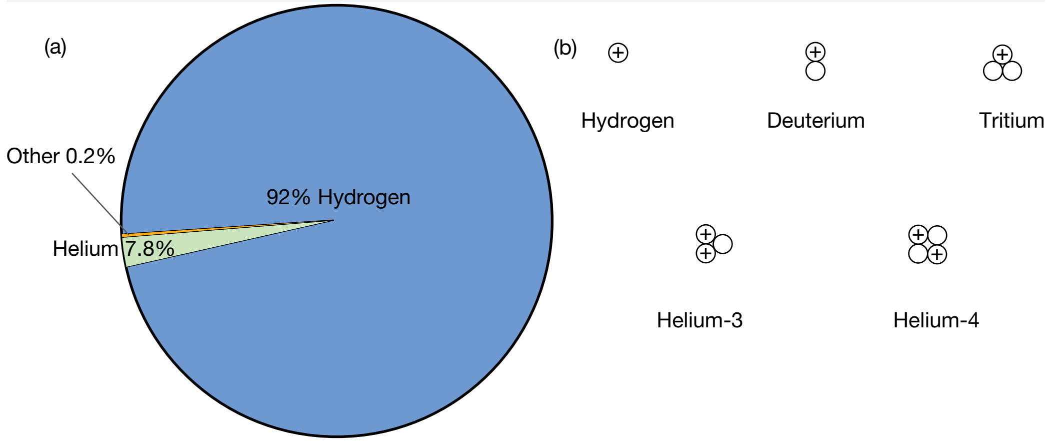
An understanding of how the Sun generates its energy did not come until well into the twentieth century. What could be the source of energy that seems to last for millions and even billions of years without any signs of running out?

All particles exert a force of gravity on other particles, and it is the force of gravity that dominates on the large scale of the universe. Clouds of hydrogen and helium gas that were present in the early universe eventually collapsed under the force of gravity. The larger and more massive the cloud, the stronger the gravitational force that pulled the gas particles in the cloud together.

Temperature is directly linked to the average kinetic energy of atoms. The cloud of gas shrinks because the force of gravity accelerates the atoms inward, the average speed of the atoms increases, therefore the temperature rises. The high temperatures in the centre of these collapsing gas clouds ensure that the cloud remains in a gaseous state and does not condense into liquid or solid. The more the cloud collapses, the more the gas particles will collide, which eventually provides enough resistance to prevent further collapse.

Consider the Sun's composition. Hydrogen makes up about 92% of the nuclei in the Sun. The rest is mostly helium, about 7.8%. Carbon, nitrogen and oxygen are the next most common elements, in total making up less than 0.1% of the atomic nuclei in the Sun. The Sun is so hot in its centre that the electrons are not attached to atoms; the nuclei and electrons swarm around separately, forming what is known as plasma. When talking of hydrogen in stars, we usually mean individual protons. Some of these protons will be attached to one or two neutrons, forming the different isotopes of hydrogen — deuterium and tritium. In the same way, helium-3 and helium-4 isotopes are formed.

FIGURE 12.37 (a) The proportion of atomic nuclei in the Sun (b) The most common isotopes present.



Arthur Eddington (1882–1944) was famous for his measurement of the bending of starlight around the Sun, which helped verify Einstein's General Theory of Relativity. Around 1926, Eddington proposed fusion to be the energy source of the Sun.

The nuclei of atoms are held together by a force known as the strong nuclear force, which acts over only very short distances within atoms. In the centre of the Sun, the pressure is so great and the temperature so high that protons are pushed together with enormous force. However, the electrostatic repulsive force that exists between all positive charges resists this coming together. Eddington calculated that the centre of the Sun would reach temperatures of about 15 million degrees Celsius simply by contraction due to gravity. Physicists at the time thought this was not high enough for the protons to get sufficiently close for the strong nuclear force to overcome the electrostatic force, allowing the protons to fuse. But quantum mechanics developed in the 1920s showed that what the protons would do was not clear cut; some would be able to fuse and others would not. It was a matter of chance.

This probability is high only when the protons have a lot of kinetic energy (which they do when the temperature is very high) and becomes significant at about 15 million degrees Celsius! Some of the protons come close enough to fuse into a single nucleus, releasing energy in the process. This **fusion** of protons does not happen easily, otherwise we would have a wonderful source of energy to replace fossil fuels and nuclear fission.

Maybe one day a practical fusion reactor will be produced, but so far attempts to achieve controlled fusion have produced energy for a very brief time. The thermonuclear (hydrogen) bomb achieves fusion in an uncontrolled way, but it requires a rather large 'match' to light the fuse. These bombs use a fission (atom) bomb to achieve the necessary temperatures. In the Sun, the gases are pulled closer together by gravity until the temperature is high enough for fusion to occur. The energy released by the fusion reactions prevents the Sun from contracting any further.

In 1938, Hans Bethe (1906–2005) at Cornell University was trying to understand what happens to protons when they are forced close together in the Sun. One of the reaction chains that Bethe found is known as the PP1, or proton–proton chain, in which two protons are forced together. A nucleus of two protons represents helium, but a nucleus of two protons alone is very unstable. One of the protons undergoes positive beta decay.

That is, it releases a positive beta particle (called a positron or positively charged electron), another particle called a neutrino and about 1.44 MeV of energy. This can be summarised in the following nuclear equation:

$$ {}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{2}_{1}\mathrm{H} + \beta^{+} + \upsilon + 1.44\,\mathrm{MeV} $$

where υ is the symbol for the neutrino.

The isotope of hydrogen formed is called deuterium. Deuterium has a chance of undergoing fusion with another proton in the following reaction:

$$ {}^{2}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + \gamma + 5.49\,\mathrm{MeV} $$

where γ is a gamma ray.

Two of these helium nuclei can fuse to form the much more stable helium-4 isotope:

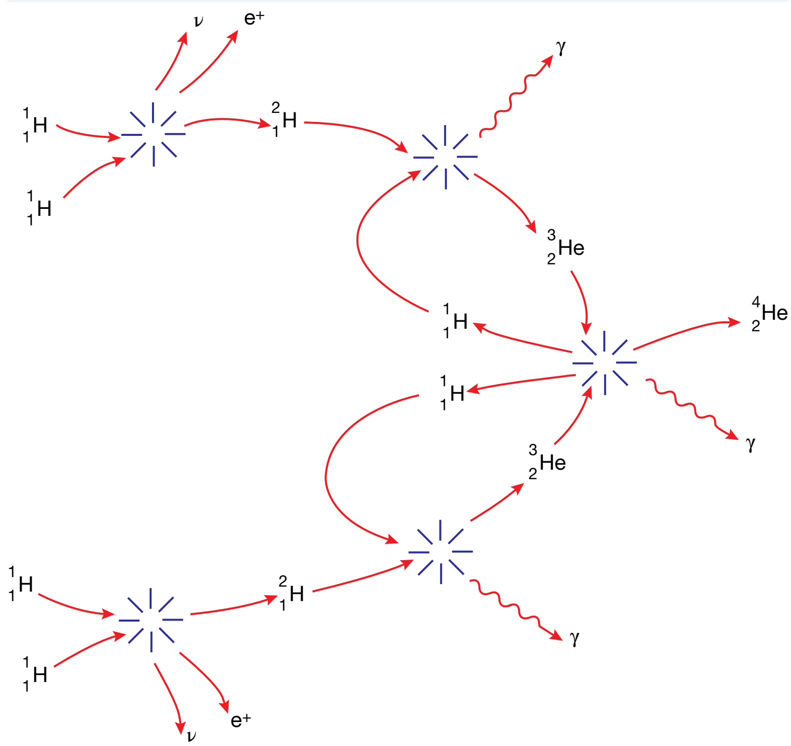
$$ {}^{3}_{2}\mathrm{He} + {}^{3}_{2}\mathrm{He} \rightarrow {}^{4}_{2}\mathrm{He} + 2\,{}^{1}_{1}\mathrm{H} + 12.9\,\mathrm{MeV} $$

This is the most common of several chains of nuclear reactions that occur in the Sun; all of them start with protons and end up with helium nuclei.

The mass of the products of each of the chains of fusion reactions is less than the total mass of the protons that undergo fusion in the chain. Einstein's statement about the equivalence of mass and energy, described by $E = mc^2$, explains what happens to the 'missing' mass. Each reaction results in a loss of mass as energy is released.

How can we be sure that this theoretical model of the Sun describes how it actually works? Experiments with particle accelerators and the hydrogen bomb have shown that fusion of hydrogen does occur and that a lot of energy is produced in the process. The neutrinos and gamma rays produced in the reactions interact very weakly with matter and therefore travel out of the Sun into space. Their presence can be detected in laboratories on Earth and in orbit, but because neutrinos interact so weakly with matter their detection is difficult. Photons of light take millions of years to make their way from the centre to the surface of the Sun because of their continual absorption and re-emission. Neutrinos, on the other hand, leave the Sun's surface in seconds. If we could see neutrinos during the day we would see them radiating from the Sun. At night we would see them coming through from the other side of the Earth, as most of them travel straight through it uninterrupted! Recently, the levels of neutrinos predicted by the fusion models were shown to agree with those detected in experiment, further validating the already well-established theory of fusion in the Sun.

FIGURE 12.38 The proton–proton chain PP1.



12.8 SAMPLE PROBLEM 1

When two hydrogen nuclei in the centre of a star fuse, 1.44 MeV of energy is released. How much mass is lost by the Sun in this reaction?

SOLUTION:

$$1.44\text{ MeV} = 1.44 \times 10^{6}\text{ eV} \times 1.6 \times 10^{-19}$$

$$= 2.3 \times 10^{-13}\text{ J}$$

$$m = \frac{E}{c^2}$$

$$= \frac{2.3 \times 10^{-13}\text{ J}}{(3.0 \times 10^{8}\text{ m s}^{-1})^2}$$

$$= 2.56 \times 10^{-30}\text{ kg}$$

So the Sun loses 2.56×10^{-30} kg in this reaction.

Both the mass and energy involved in each fusion reaction may seem tiny. However, the Sun contains in the order of 10^{57} protons that can fuse in its lifetime to form helium.

12.8 Exercise 1

1 Given that the Sun radiates 3.828×10^{26} joules per second, determine the number of proton-proton fusions occurring each second if this was the only source of its power.

12.8.3 What happens next?

After about 10 billion years most of the hydrogen in the core of the Sun will have fused to form helium. Once the hydrogen in the Sun's core is exhausted, it will cool and contract a little. The temperature at the core's edge will rise, due to the extra contraction, causing the hydrogen that remains at the edge to undergo fusion. This fusion closer to the surface of the star will heat its outer layers, causing them to expand and then cool. At this point the Sun will become a **red giant**.

The Sun as a red giant will be about 1000 times as bright as it is now and have a radius about 100 times its current size. Its core will be small and extremely dense and hot, while the outer layers will have very low density and will be quite cool. After a time, the collapsing core will reach a temperature where helium begins to fuse. This will produce enormous amounts of energy throughout the whole core. The rapid heating that occurs in this time is known as a *helium flash*. The core will reach an enormous 350 million degrees Celsius, causing the star to expand and cool. Then the Sun will continue burning helium in the core and hydrogen on the edge of the core. The helium fusion will involve the following reactions:

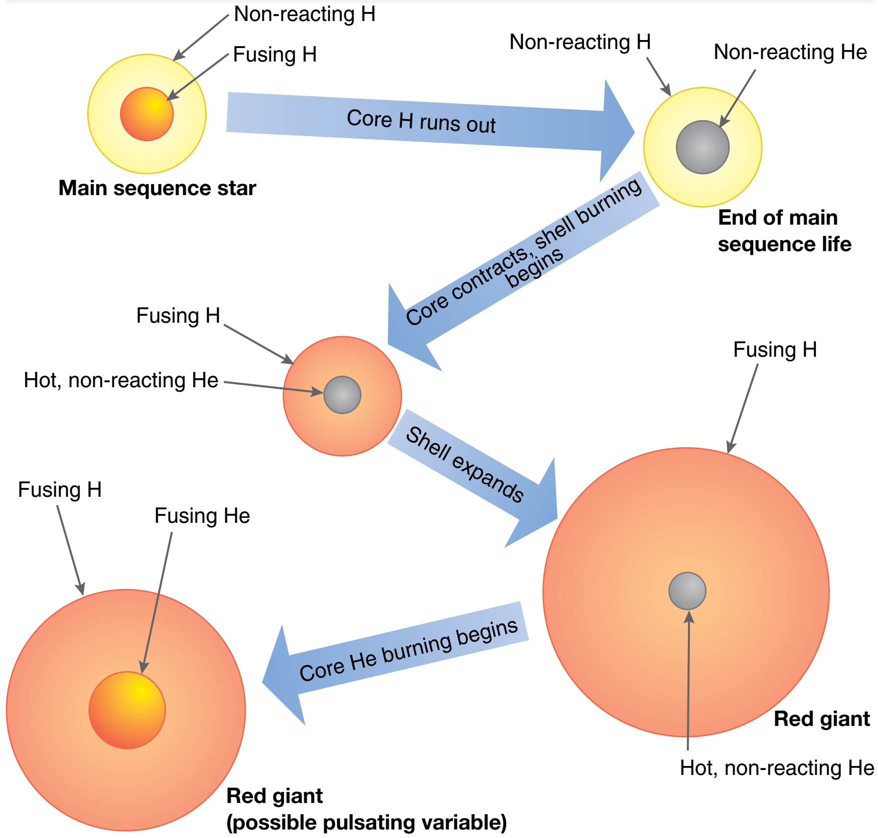
$$^{4}_{2}\text{He} + {}^{4}_{2}\text{He} \leftrightarrow {}^{8}_{4}\text{Be} + \gamma \quad \text{then} \quad {}^{8}_{4}\text{Be} + {}^{4}_{2}\text{He} \rightarrow {}^{12}_{6}\text{C} + \gamma$$

This is known as the **triple alpha process** because each helium nucleus is an alpha particle and three are required to make the carbon-12 nucleus. The double-headed arrow in the first equation, which indicates that the reaction occurs in both directions, is included because beryllium-8 is not very stable and tends to disintegrate into two helium-4 nuclei if it does not fuse quickly with another helium-4 nucleus.

In contrast, carbon-12 is a very stable nucleus and eventually it will replace the hydrogen and helium that currently form the Sun's core. Helium will continue to fuse on the edges of the carbon core, again causing the outer layers to expand. The Sun will become a bit unstable at this time, its size pulsing in and out every

10 000 years or so due to the sensitivity of the rate of helium fusion to temperature. As the outer layers expand, the temperature of the Sun drops, causing fusion to stop. The outer layers then cool and contract until the temperature again reaches the point where helium fusion can occur. Eventually this pulsing will throw off the outer layers of the Sun, which will form a ring called a **planetary nebula**.

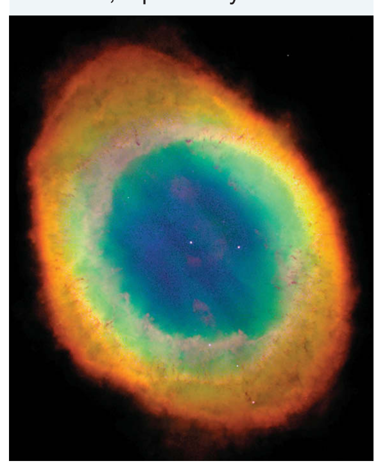
FIGURE 12.39 The development of a star from the hydrogen fusion stage through to the helium fusion stage.



The Sun is not large enough to cause carbon to undergo fusion. The mainly carbon remnant is known as a **white dwarf** and this gradually cools over a few billion years to form a **black dwarf**.

All of this happens over billions of years and so has not been observed. Astrophysicists have given us reason to be confident this is the process our Sun will go through, but red giants, planetary nebulae and white dwarfs are all present in our galaxy.

FIGURE 12.40 The Ring Nebula, a planetary nebula.



12.8.4 The CNO (carbon, nitrogen, oxygen) cycle

While the proton-proton chain is the dominant fusion process at work in the Sun, there is an alternative, known as the CNO (Carbon, Nitrogen, Oxygen) cycle. The CNO cycle features especially in stars larger than the Sun where the temperature is hotter in the core. It occurs in the Sun but it accounts for less than 2% of the Sun's fusion reactions. Like the proton-proton chain, the CNO cycle fuses protons to form helium nuclei. The difference is that it makes use of the small amount of carbon, nitrogen and oxygen in the star to enable a different chain of reactions. Stars such as the Sun contain many elements because they were formed relatively late in the universe's existence when there had been generations of earlier stars.

Many of the earliest stars were very large so were very productive in creating new elements through fusion and other processes. Large stars move through their life cycle much faster than stars such as the Sun, so there was time in the eight or nine billion years prior to the Sun's formation for several cycles of star birth and death. Importantly as well, large stars end violently with supernova that are very effective both at generating different nuclei and dispersing them through a large region of space.

The CNO cycle:

1. A proton is captured by a carbon-12 nucleus to produce nitrogen-13 and a gamma ray.

   $^{12}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{13}_{7}\mathrm{N} + \gamma$

2. The unstable nitrogen-13 decays to produce carbon-13, a positron (antimatter version of electron) and a neutrino.

   $^{13}_{7}\mathrm{N} \rightarrow {}^{13}_{6}\mathrm{C} + \mathrm{e}^{+} + \nu$

3. The carbon-13 captures a proton to produce nitrogen-14 and a gamma ray.

   $^{13}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{14}_{7}\mathrm{N} + \gamma$

4. The nitrogen-14 captures a proton to produce oxygen-15 and a gamma ray.

   $^{14}_{6}\mathrm{N} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{15}_{7}\mathrm{O} + \gamma$

5. The unstable oxygen-15 decays to produce nitrogen-15, a positron and a neutrino.

   $^{15}_{8}\mathrm{O} \rightarrow {}^{15}_{7}\mathrm{N} + \mathrm{e}^{+} + \nu$

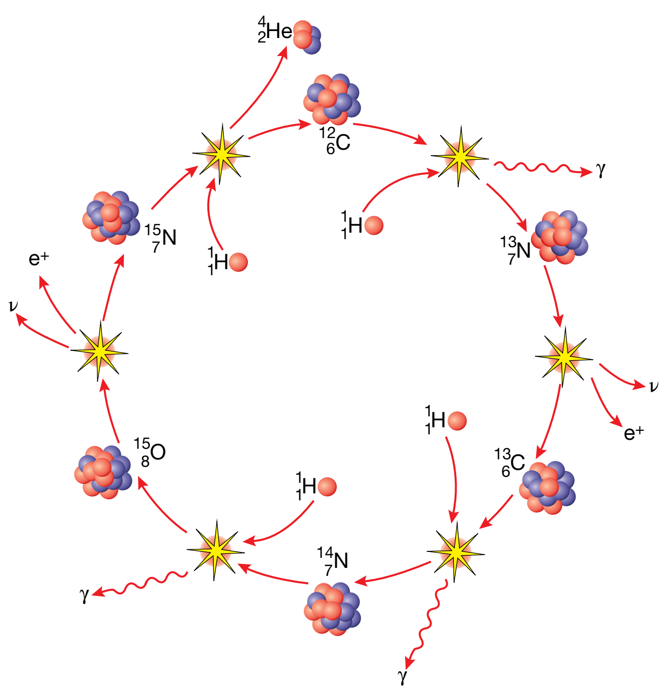
6. The nitrogen-15 captures a proton to produce carbon-12 and a helium nucleus.

   $^{15}_{7}\mathrm{N} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{12}_{6}\mathrm{C} + {}^{4}_{2}\mathrm{He}$

The carbon-12 was present at the start of the process so, overall, we have four protons combining to form a helium-4 nucleus.

$$4\,{}^{1}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He} + 2\mathrm{e}^{+} + 2\nu + 3\gamma + 24.7\mathrm{MeV}$$

FIGURE 12.41 The CNO cycle.



The two positrons almost immediately annihilate in interaction with electrons, producing a further 2 MeV.

The carbon that was present at the start of the cycle is present at the end so there is no net change in its amount. The nitrogen and oxygen isotopes are formed temporarily as part of the cycle. As the carbon, nitrogen and oxygen are not products or reactants in the cycle, they are referred to as catalysts.

12.8 SAMPLE PROBLEM 2

1. In which types of stars is the CNO cycle more likely than the proton-proton chain?
2. What is the difference in products of the proton-proton chain and the CNO cycle?

SOLUTION:

1. Stars on the main sequence with significantly more mass than the Sun have higher core temperatures. These are the conditions where the CNO cycle dominates. Sun-sized stars and smaller are powered predominately by the proton-proton chain of reactions. They also need to be stars that contain carbon-12.
2. There is no difference. They both produce a helium-4 nucleus, 2 positrons, 2 neutrinos and 3 gamma photons.

12.8 Exercise 2

1 Would the proton-proton chain or the CNO cycle have been more likely in the earliest stars?
2 What is the role of carbon, nitrogen and oxygen in the CNO cycle?

12.8.5 And all the other elements?

So far, we have only accounted for a few elements. Hydrogen, helium and a trace of lithium nuclei formed in the first few minutes of the universe. Some more helium is produced in stars such as the Sun, and carbon is formed through fusion of helium later on. There are still a lot of gaps in our periodic table!

More massive stars than the Sun can undergo fusion to produce many nuclei all the way to iron-56. Some lighter elements, including the stable isotopes of lithium, beryllium and boron, are formed by high-energy cosmic rays interacting with matter between stars, not inside stars. Some reactions in large stars produce free neutrons that can be captured by other nuclei. These captures either lead to a new stable isotope of the nucleus that captures the neutron, or an unstable nucleus that undergoes beta decay to form a new element. This is known as the s-process, or slow neutron capture process. It is one of the most important processes for forming heavier nuclei. Another process for forming nuclei heavier than iron is known as the r-process, or rapid neutron capture process. It occurs in the presence of heavy nuclei and an abundance of free neutrons, conditions found during a supernova explosion at the end of the life of a massive star.

12.9 Review

12.9.1 Summary

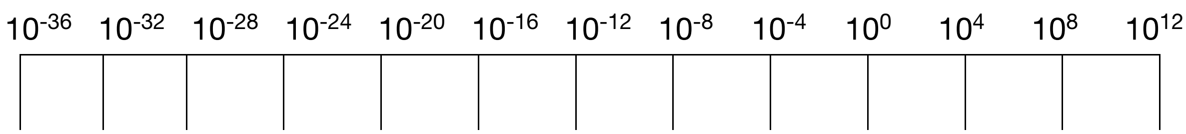
- The earliest moments of the universe were too hot for matter to form. As it expanded and cooled, different particles were able to form, first the Higgs boson, then quarks, electrons and neutrinos, then protons and neutrons, then some light nuclei including deuterons, helium-4 and a tiny amount of lithium-7.
- About 20 minutes after it began, the universe consisted of hydrogen and helium nuclei with 25% of the mass in helium and 75% in hydrogen; this proportion has only slightly changed in the nearly 14 billion years since.

- Atoms first appeared 380 000 years into the universe's existence when the universe was cool enough for electrons to stay bound to nuclei. Light was then free to travel through the universe, which is now detected as the cosmic microwave background radiation.
- Techniques for measuring the distance to distant stars and distinguishing nebulae from distant clusters of stars led to the discovery that our galaxy, the Milky Way, is only one of billions of galaxies.
- Cepheid variables, whose period of varying luminosity is related to their average luminosity, enable us to measure distances to distant stars.
- Spectra can be continuous for dense materials, including the interiors of stars. These closely resemble the black body spectrum of an ideal radiator.
- Light passing through a gas causes the atoms in the gas to change energy states, removing particular frequencies from the light. This creates an absorption spectrum.
- A glowing gas emits light at characteristic frequencies, called an emission spectrum.
- Stars emit absorption spectra and nebulae (gas clouds) emit emission spectra.
- Analysis of the spectra from galaxies revealed to Edwin Hubble that all distant galaxies are moving away from us, and the further away they are, the greater their red-shift.
- The spectra of more distant galaxies are all red-shifted by an amount that increases with distance, so red-shift is a very useful technique for measuring how far away distant galaxies are.
- The expansion of space provided the first evidence that the universe had a beginning, now determined to have occurred 13.8 billion years ago.
- Production of heavier elements did not occur until the first stars began fusing hydrogen and helium in their cores, 800 000 000 years after the beginning of the universe.
- The big bang is the name given to the theory that describes the universe beginning from a point of infinite density and expanding to create space and time as we see it today.
- The key evidence for the big bang includes: the expansion of the universe, the higher density of galaxies in the past, the proportion of elements in the universe, and the cosmic microwave background radiation.
- The source of energy for the Sun and other stars needs to supply an enormous energy production for very long periods of time. Chemical reactions and gravitational energy would have been exhausted long ago.
- Einstein's equivalence of mass and energy through the equation $E = mc^2$ can be used to determine how much mass the Sun is losing per second as a result of the energy it is radiating.
- $E = mc^2$ can be used to determine which nuclear reactions can occur in a star. To produce energy, the mass of the products of the reaction must be less than the mass of the total mass of the individual particles prior to the reaction.
- Proton-proton fusion to form deuterium via emission of a positron is energetically feasible in the Sun, as are other fusions leading to the production of helium-4.
- From electromagnetic radiation, astrophysicists are able to determine the temperatures of stars and the elements they contain.
- The spectra of stars appear different depending on the temperature of the star. This is due to the ionisation states of different atoms in the atmosphere of the stars at the different temperatures. These differences have been used to classify stars by spectral type in order from hottest to coldest: OBAFGKM.
- More massive stars shine more brightly and are hotter than low-mass stars and also pass through their life cycle more quickly.
- The Hertzsprung–Russell (H–R) diagram is a graph of the luminosity of stars against their temperature or colour.
- In an H–R diagram, most stars are found on a diagonal line called the main sequence, from hot and luminous down to cool and dim.
- H–R diagrams reveal the characteristics of the star, such as its colour, surface temperature, size and stage of its life cycle.
- Main sequence stars fuse hydrogen into helium in their cores.

- Stars above the main sequence in an H–R diagram have consumed all the hydrogen in their cores and have expanded in size to form red giants due to fusion heating the outer layers of the star.
- Stars below the main sequence in an H–R diagram are ending their life cycle; fusion has finished and most of their material has been shed, forming a planetary nebula. They are remnants of stars cooling down as white dwarfs.
- Stars are powered largely by fusion. Main sequence stars, which include all stars for most of their lives, fuse protons through the proton-proton chain or the CNO cycle to produce helium-4.
- The proton-proton chain does not rely on the presence of any other nuclei, but the CNO cycle relies on carbon-12 being present. Carbon, nitrogen and oxygen act as catalysts in this process but any that is produced is also consumed in the process, so the only gain in nuclei at the end of the cycle is helium-4.
- Once the proton supply in the star's core is exhausted, the core contracts, causing the temperature to rise, allowing fusion on the edge of the core. This causes the outer layers of the star to expand into a giant or a supergiant, and the star leaves the main sequence.
- When the temperature of the core is sufficiently hot, helium-4 can begin to fuse through the triple alpha process to form carbon. Larger stars can continue fusion to produce elements up to iron.
- Heavier elements are produced by neutron capture, mostly in supernovae — explosions that occur at the end of the lives of very large stars. Collisions between cosmic rays and nuclei in interstellar space are also important in producing some of the elements in the universe.

12.9.2 Questions

1. What atoms were present in the first 100 000 years of the universe's existence? Explain.
2. Why were heavier nuclei, such as carbon and oxygen, not created during the intense temperatures and pressures in the early universe?
3. Order the following events in time from first to last: inflation, nuclear fusion, particle–antiparticle annihilation, the formation of atoms, and ignition of the first stars.
4. Describe the hypothesis of inflation and one problem that it solves.
5. Telescopes are not able to see anything prior to recombination, even in theory. What tools do physicists use to improve their understanding of what happens to matter in the conditions in the early universe?
6. What prevented nuclear fusion prior to its commencement some seconds after the beginning of the universe?
7. Draw a time line of the first 20 minutes after the big bang. Use a scale where each centimetre represents an increase by $\times 10^2$ seconds as in the following example:



8. The ratio of hydrogen and helium in the universe was determined in the first 20 minutes. Explain why the ratio ended up approximately 25% helium and 75% hydrogen.
9. What did Hubble discover in 1929 that led to the formulation of the big bang theory?
10. How did the red-shift of the most distant galaxies compare with the red-shift of galaxies closer to us?
11. What is the major cause of the red-shift of the light that has travelled from the most distant galaxies?
12. The light from the Andromeda galaxy is blue-shifted. Explain why it is not red-shifted like the light from most galaxies.
13. State Hubble's Law.
14. Sketch a graph of red-shift versus distance that summarises Hubble's observations of galaxy red-shifts.
15. How is the big bang different from normal explosions?
16. List the key evidence in favour of the big bang theory.
17. How does the big bang theory explain the predominance of hydrogen and helium in the universe?

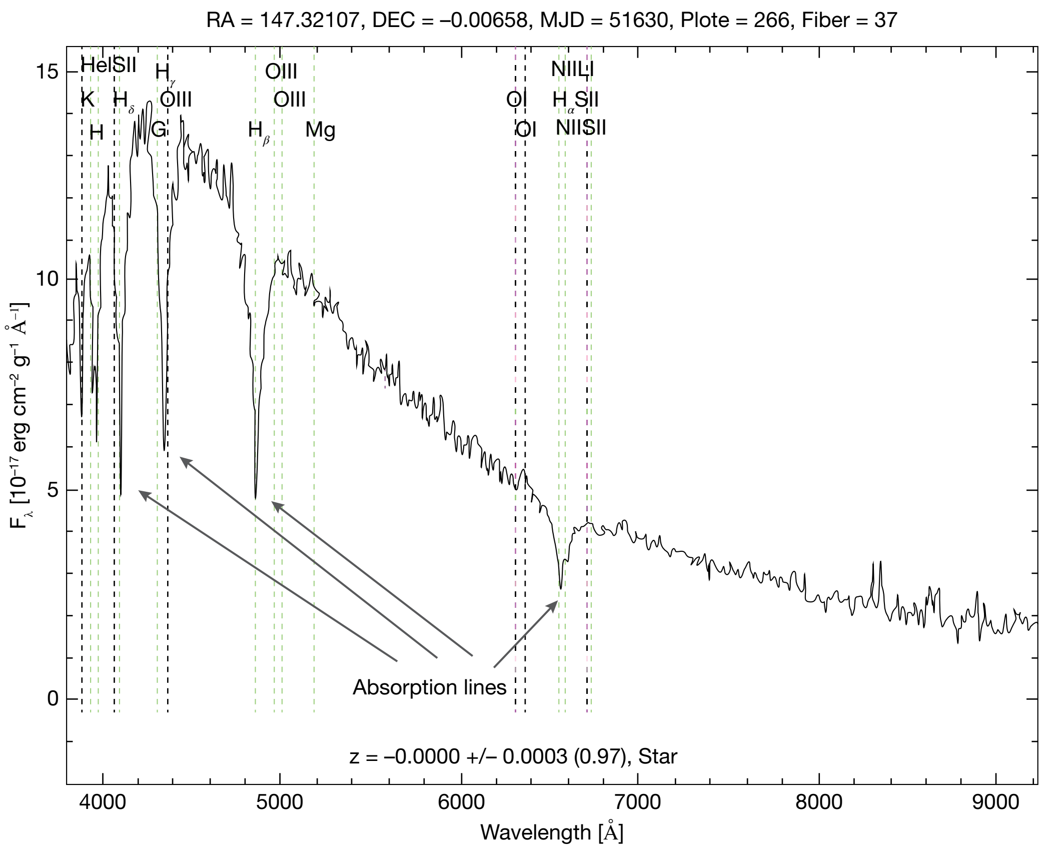
18. Why was the cosmic microwave background discovery so important in establishing the big bang theory?
19. How is the expansion of the universe related to the early formation of matter?
20. The Sun radiates 3.86×10^{26} W.
    (a) How much energy does it produce per hour?
    (b) What is the mass equivalent of this energy?
21. How many electron volts of energy does the Sun produce per second?
22. Astronomers often talk about the luminosity of stars using the unit of solar luminosity ($L_{\odot}$). $1\,L_{\odot} = 3.86 \times 10^{26}$ W. The luminosity of Proxima Centauri, the closest star to Earth other than the Sun, is $0.00005\,L_{\odot}$.
    (a) What is its luminosity in watts?
    (b) How much mass is Proxima Centauri losing per year to produce this luminosity?
23. The fusion of two protons to form a deuteron releases 0.42 MeV. The mass of free protons is 1.6726×10^{-27} kg. When 1 kg of coal is burned in a chemical reaction, it releases about 30 MJ of energy. Compare the proportion of mass lost in the process of the 2 protons with the proportion of mass lost from the burning of 1 kg of coal.
24. How did scientists know that chemical energy and gravitational energy were not enough to fuel the Sun?
25. Describe the difference in appearance of an emission spectrum and an absorption spectrum.
26. Describe the situations that would produce an absorption spectrum.
27. Clouds of gas in space are called nebulae. Explain why a nebula usually produces an emission spectrum whereas gas around a star produces an absorption spectrum.
28. Describe the role of spectroscopy in the discovery of galaxies.
29. Describe the role of spectroscopy in the development of the big bang theory.
30. What can spectroscopy tell us about stars?
31. Which of the following most closely approximates a black body?
    (a) the interior of a star
    (b) the atmosphere of a star
    (c) a dark nebula
    (d) a planet
    (e) a bright nebula.
32. Place these colour stars in order of surface temperature from hottest to coldest: white, blue, red, orange, yellow.
33. In what three ways is the surface temperature of stars represented in the stellar spectrum, assuming that the distance to the stars is taken into account (controlled for)?
34. The different spectral classes are represented by the letters O, B, A, F, G, K and M. What is the reason for the spectra of different classes being different?
35. A star called Koo She has a spectral class of A0V. Describe the star in terms of temperature, colour, the fusion reactions occurring in its core.
36. Look up a table of the 30 brightest stars in the night sky and complete the following table to show how common each class is among these bright stars:

Spectral Class	Frequency
O	
B	
A	
F	
G	
K	
M	

Do you think that this sample represents the distribution of spectral classes of all stars? Why?

37. By comparing the absorption lines in Figure 12.42 with those of the different spectral types in Figure 12.31, determine the likely spectral class of this star. What colour would it be?

FIGURE 12.42



38. What is the cause of the absorption lines? Why would they be different for stars in different classes?
39. What are some key differences you would expect to see in the spectrum of an M class star with the one in the graph in question 37?
40. (a) Sketch an H–R diagram. Include the spectral classes on the horizontal axis.
    (b) Circle and label the main sequence.
    (c) Is it normal for stars to move along the main sequence during their life spans? Explain.
    (d) What event will cause the Sun to leave the main sequence?
    (e) Circle and label the stars that have mainly fusion of helium and heavier elements in their cores.
    (f) In what circumstances is it possible to have a very massive star positioned at the right-hand side of the diagram?
    (g) What section of the diagram contains remnants of stars that no longer have fusion reactions as a source of energy?
    (h) Andre measures the composition and temperatures of two stars to be the same. He expects to place them in the same region of the H–R diagram. One of the stars is brighter than the other. What does this tell him about the two stars?
    (i) A star lies on the main sequence. What does this tell you about the star?
    (j) What colour are stars on the left-hand side of the diagram?
41. What are the key sources of energy in a star?
42. What two elements make up the vast majority of the nuclei in the universe?

43. The Sun radiates energy at $3.86 \times 10^{26}\,\mathrm{J\,s^{-1}}$. Assume that this is all the result of the fusion of hydrogen to helium following the proton-proton chain. We can summarise the fusion of these reactions by the equation: $4\,{}^{1}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He} + 2\mathrm{e}^{+} + 2\nu + 3\gamma + 26.76\,\mathrm{MeV}$
   (a) How many of these fusion reactions would occur in the Sun per second?
   (b) Given that the Sun has about 10^{57} hydrogen nuclei but that only about 10% of those will fuse in the core, how long do you predict the Sun will continue to fuse hydrogen? (Give your answer to one significant figure.)
44. One of the reactions that occurs in the Sun is the fusion of helium-3 and helium-4.
   (a) Complete the equation: ${}^{3}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} \rightarrow \underline{\qquad} + \gamma + 1.59\,\mathrm{MeV}$
   (b) Use the energy released by the reaction to determine the mass difference between the nuclei on the left- and right-hand sides of the equation.
   (c) Where could the helium isotopes for this reaction have come from?
   (d) This is an intermediate reaction in a chain of reactions that will occur in the Sun. What is the final product of this chain of reactions?
45. Explain why, if hydrogen and helium make up the vast majority of the nuclei in the universe, they are relatively rare on Earth.
46. The CNO cycle is another process that occurs in the core of the Sun resulting in the fusion of protons to form helium-4. Identify two features of a star that are essential for substantial fusion via the CNO cycle that do not need to be present for the proton-proton chain.
47. (a) When the Sun's hydrogen in the core is exhausted, fusion in the core will continue but the Sun will move off the main sequence and begin fusing helium-4 through a triple alpha process (a helium-4 nucleus is also known as an alpha particle). Why doesn't this process also result in beryllium-8 accumulating in the core? (Hint, look up the half-life of beryllium-8.)
   (b) Why does the star expand greatly during this phase?

eBook*plus* RESOURCES

Complete this digital doc: Investigation: Expansion of the universe
Searchlight ID: doc-16173

Complete this digital doc: Investigation: Hertzsprung–Russell diagrams
Searchlight ID: doc-17064

PRACTICAL INVESTIGATIONS

Investigation 12.1 Expansion of the universe

Aim

To model the expansion of the universe in two dimensions, using the surface of a balloon.

Apparatus

- balloon
- fine permanent marker
- string (about 30 cm long)
- ruler.

Method

1. Blow a small amount of air into the balloon so that its diameter is about 10 cm. Hold the opening of the balloon so that none of the air escapes.

2. Using the marker, place a dot on the end of the balloon, opposite its opening, to represent the Earth. Put five other dots around that end of the balloon and number them. Make sure they are at a variety of distances from 'Earth'. These dots represent galaxies.
3. Measure the distance from the Earth to each of the galaxies — the string makes measuring distances on the curved surface of the balloon easier. Record your measurements in a table like Table 12.6.
4. Measure the distance from one of the other galaxies to Earth and each of the other galaxies. Record your measurements in a table like Table 12.7.
5. Puff a few more breaths into the balloon and close the opening. Measure the distance from Earth to each of the galaxies again. Also, re-measure the distance from your chosen galaxy to each of the other galaxies. Record your measurements in the Distance 2 column of the appropriate table.

TABLE 12.6 Distance from Earth

Galaxy	Distance 1 (cm)	Distance 2 (cm)	D2 – D1(cm)
1			
2			
3			
4			
5			

TABLE 12.7 Distance from galaxy 1

Galaxy	Distance 1 (cm)	Distance 2 (cm)	D2 – D1(cm)
2			
3			
4			
5			
Milky Way (Earth's galaxy)			

Analysis

For both sets of measurements:

1. Compare Distance 2 with Distance 1. Relate this to Hubble's observation of the universe.
2. What is the relevance of the results from the galaxy 1 measurements?
3. Explain why the difference between Distance 2 and Distance 1 (D2−D1) gives an indication of apparent velocities of galaxies.
4. Graph D2−D1 versus Distance 1. What does this show you about the recession of galaxies?

TOPIC 13
The structure of the atom

13.1 Overview

13.1.1 Module 8: From the universe to the atom

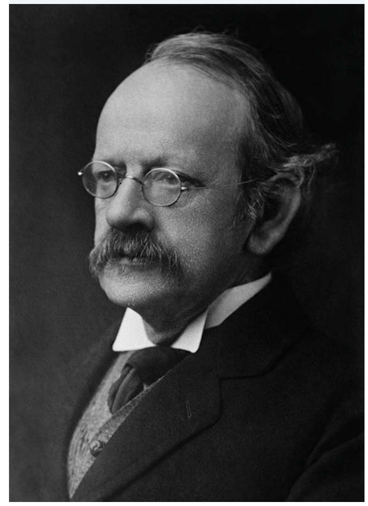
Structure of the atom

Inquiry question: How is it known that atoms are made up of protons, neutrons and electrons?

Students:

- investigate, assess and model the experimental evidence supporting the existence and properties of the electron, including:
  - early experiments examining the nature of cathode rays
  - Thomson's charge-to-mass experiment
  - Millikan's oil drop experiment (ACSPH026)
- investigate, assess and model the experimental evidence supporting the nuclear model of the atom, including:
  - the Geiger–Marsden experiment
  - Rutherford's atomic model
  - Chadwick's discovery of the neutron (ACSPH026)

FIGURE 13.1 The famous physicist J.J. Thomson (1856–1940) was celebrated for his experiments with the electron.



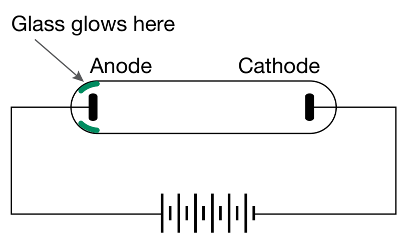
13.2 Cathode rays and the electron

13.2.1 The discovery of cathode rays

In the early part of the nineteenth century, the discovery of electricity had a profound effect on the study of science. By the 1850s, much was known about which solids and liquids were electric conductors or insulators, and it was thought that gases were electric insulators.

The development of a vacuum, using pumps to remove the air from glass tubes, was also being actively researched at this time. As improved vacuum pumps were developed, scientists were able to experiment with gases at very low pressures. In 1855, a German physicist, Heinrich Geissler (1814–1879), refined a vacuum pump so that it could be made to evacuate a glass tube to within 0.01 per cent of normal air pressure. Geissler's friend Julius Plucker (1801–1868) took these tubes and sealed a metal plate, called an electrode, to each end of the tube. The electrodes made electrical connections through the glass and were sealed to maintain the partial vacuum in the tube. These were then connected to a high-voltage source, as illustrated in Figure 13.2. To their surprise, the evacuated tube actually conducted an electric current. What puzzled them more was the fact that the glass at the positive end, or **anode**, of the vacuum tube glowed with a pale green light. What type of invisible 'ray' caused this glow or **fluorescence**?

FIGURE 13.2 Production of cathode rays in a discharge tube, as used by Plucker.

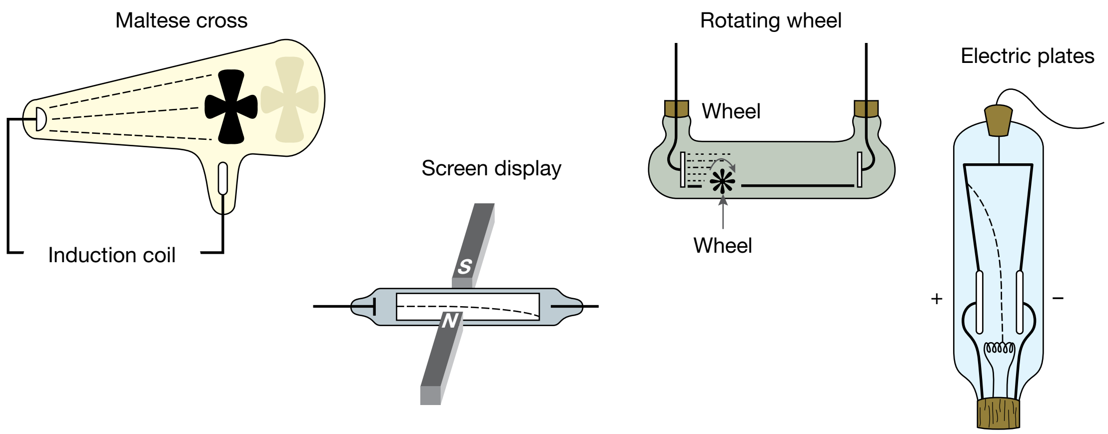


Whatever it was must have originated at the negative electrode, or **cathode**, of the vacuum tube. Another physicist, Eugene Goldstein (1850–1930), who was studying these same effects, named the rays that caused the glow 'cathode rays', and the tubes became known as **cathode ray tubes** or **discharge tubes** (see Figure 13.3). Early experimenters used these tubes to investigate all the properties of cathode rays and

X-rays. Some modified the cathode ray tube to include a rectangular metal plate covered in zinc sulphide inside the tube. This plate had a horizontal slit cut into the end nearest the cathode and the plate was slightly bent so that the cathode rays formed a horizontal beam. When the cathode rays struck this material, it appeared fluorescent and showed the path of the rays through the tube.

Cathode ray tubes were then refined and developed and were used in television sets, computers and many other applications.

FIGURE 13.3 A variety of early discharge tubes used in experiments.



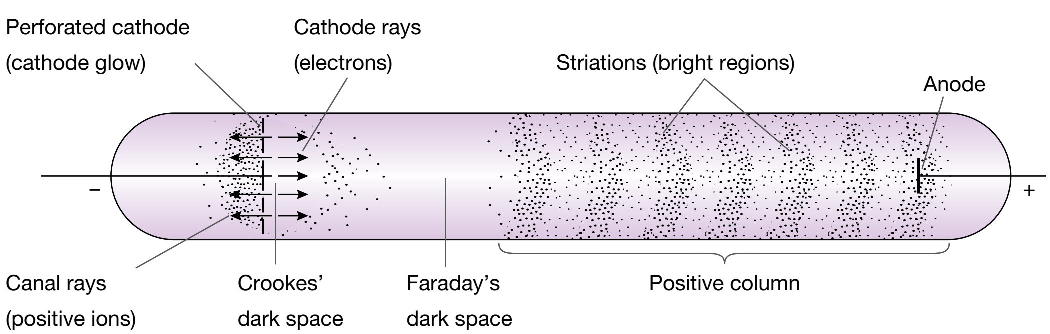
13.2.2 Discharge tubes

Discharge tubes evacuated to different air pressures were found to produce different effects.

For example, in the first practical investigation for this topic, an induction coil acts as a step-up transformer, delivering a high voltage across the set of discharge tubes. At low pressures, electrons can accelerate to faster speeds before colliding with gas particles. Initially, a current will flow even though nothing can be seen. The first effect that can be observed is a steady luminous discharge known as a 'glow discharge'. As the pressure is lowered further, a number of colourful effects can be seen.

At first, most of the tube is occupied by a bright luminous region called a 'positive column', which appears to start from the anode and is broken up into a series of bands or striations (see Figure 13.4). Near the anode, a weaker glow can be seen. The striations are separated by 'dark spaces'. These discharges and spaces are named after some of the scientists who examined them, for example, 'Ashton's dark space', 'Crookes' dark space' and 'Faraday's dark space'. The colours of the discharge depend on the gas used. In low pressure air, the positive column is a brilliant pink and the negative glow is deep blue.

FIGURE 13.4 Some of the effects observed in discharge tubes.



PHYSICS IN FOCUS

Everyday uses of discharge tubes

Neon signs colour the night in every city street. They are long tubes with most of the air removed. A small amount of gas is introduced which, when excited by a high potential, glows with a characteristic colour. For example, when the added gas is neon, the kinetic energy of the electrons is sufficient to ionise the gas around the cathode, causing the emission of a reddish light.

Fluorescent tubes in the home contain mercury vapour at low pressure. The light produced is in the ultraviolet region of the electromagnetic spectrum. To produce visible light, a thin coating of a powder is spread on the inside surface of the tube. The ultraviolet radiation causes this coating to fluoresce with the familiar bright white light.

FIGURE 13.5 Discharge tubes are used in the neon lights that are often a feature of streetscapes and venue advertising at night.

13.2.3 The effect of electric fields on cathode rays

You are familiar with three types of fields: gravitational, electric and magnetic. An electric field exists in any region where an electric charge experiences a force. There are two types of charge — positive and negative. We define the direction of the electric field as the direction in which a positive charge will experience a force when placed in an electric field.

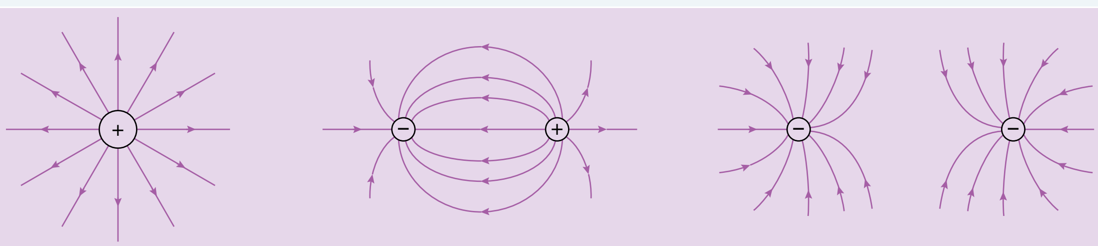
This definition of an electric field allows us to describe the fields around a charge (see Figure 13.6). Using Faraday's 'lines of force', we see that these lines radiate from a point at the centre of the charge. For a positive charge, lines of force leave the centre of the charge and radiate in all directions from it. For a negative charge, the lines are directed radially into the centre of the charge.

If a positive charge is placed near another positive charge, it will experience a force of repulsion, that is, a force that acts in the direction of the arrow.

A number of rules apply to the interpretation of these lines of force diagrams (see Figure 13.6).

- Field lines begin on positive charges and end on negative charges.
- Field lines never cross.
- Field lines that are close together represent strong fields.
- Field lines that are well separated represent weak fields.
- A positive charge placed in the field will experience a force in the direction of the arrow.
- A negative charge placed in the field will experience a force in the direction opposite to the arrow.

FIGURE 13.6 Electric fields around charges.



(a) Electric field around a single positive charge

(b) Opposite charges

(c) Like charges

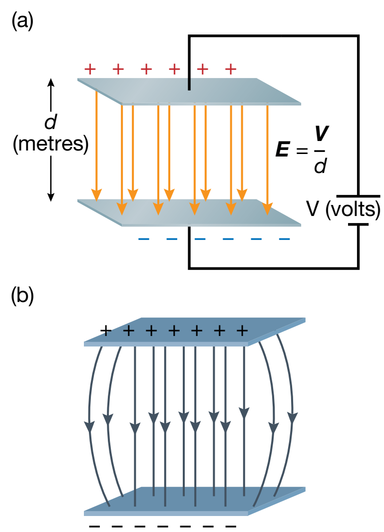
Uniform electric fields

A uniform electric field can be made by placing charges on two parallel plates that are separated by a small distance compared with their length. These electric fields are very useful in and were used by prominent scientists such as Robert Millikan and J.J. Thomson when investigating the properties of small charged particles.

Consider the electric field between two plates that are separated by d metres, as shown in Figure 13.7a.

The magnitude, or intensity, of an electric field is determined by finding the force acting on a unit charge placed at that point. The symbol for electric field is $\boldsymbol{E}$.

FIGURE 13.7 (a) Electric field (E) between two parallel plates. (b) The electric field is uniform except at the edges of the plates where it bulges slightly.



$$\boldsymbol{E} = \frac{\boldsymbol{F}}{q}$$

where
$\boldsymbol{E}$ = electric field intensity (in newtons per coulomb)
$\boldsymbol{F}$ = electric force (in newtons, N)
q = electric charge (in coulombs, C).

When the potential difference, or voltage, is applied to the plates, a uniform electric field is produced. The strength of this field is the same at all points between the plates, except near the edges where it 'bulges' slightly (see Figure 13.7b).

The magnitude of the electric field, $\boldsymbol{E}$, is given by:

$$E = \frac{V}{d}$$

where
V = potential difference, in volts.

This can be derived by recalling that potential difference is the change in potential energy per unit charge moving from one point to the other. The amount of energy or work is given by:

$$W = qV.$$

Also, the work done by a force is the product of the force and the distance moved, d. In this case, $F = qE$. Hence, the amount of work is given by:

$$W = Fd = qEd$$

It follows that: $V = Ed$ or $E = \frac{V}{d}$.

Remember that *work done is equal to the gain in energy*.

A small positive charge released next to the positive plate will experience a force that will accelerate the charge. The charge will increase its kinetic energy.

$$W = qV = \frac{1}{2}mv^2$$

This shows that the amount of work done depends only on the potential difference and the charge, and is the same for both uniform and non-uniform electric fields.

13.2 SAMPLE PROBLEM 1

What is the electric field strength between two parallel plates separated by 5.0 mm if a potential difference of 48 volts is applied across them?

SOLUTION:

$$\begin{aligned} V &= 48 \text{ volts} \\ d &= 5.0 \text{ mm} \\ &= 5.0 \times 10^{-3} \text{ m} \\ E &= \frac{V}{d} \\ &= \frac{48}{5.0 \times 10^{-3}} \\ &= 9600 \text{ V m}^{-1} \text{ or N C}^{-1} \end{aligned}$$

13.2 SAMPLE PROBLEM 2

How much work is done moving a charge of 3.6 μC through a potential difference of 15 volts?

SOLUTION:

$$\begin{aligned} q &= 3.6 \times 10^{-6} \text{ C} \\ V &= 15 \text{ volts} \\ W &= qv \\ &= 3.6 \times 10^{-6} \times 15 \\ &= 5.4 \times 10^{-5} \text{ J} \end{aligned}$$

13.2 SAMPLE PROBLEM 3

Two parallel plates are separated by a distance of 5.0 mm. A potential difference of 200 volts is connected across them. A small object with a mass of 1.8×10^{-12} kg is given a positive charge of 12 μC. It is released from rest near the positive plate. Calculate the velocity gained as it moves from the positive plate to the negative plate.

SOLUTION:

$$\begin{aligned} d &= 5.0 \times 10^{-3} \text{ m} \\ V &= 200 \text{ V} \\ q &= 1.2 \times 10^{-5} \text{ C} \\ m &= 1.8 \times 10^{-12} \text{ kg} \end{aligned}$$

$$\begin{aligned} qV &= \frac{1}{2}mv^2 \\ v^2 &= \frac{2\left(1.2 \times 10^{-5} \times 2.00 \times 10^2\right)}{1.8 \times 10^{-12}} \\ v &= 5.2 \times 10^4 \text{m s}^{-1} \end{aligned}$$

13.2 Exercise 1

1 A research scientist wishes to make a uniform electric field of strength 200 N C^{-1} between two parallel plates separated by 12.5 cm.

(a) State the electric field strength in V m^{-1}.

(b) Calculate the potential difference applied to the plates to make an electric field strength as specified.

(c) A positive test charge $q = 2.0 \times 10^{-4}$ C is placed in between the plates. Determine the size and direction (towards the positive plate or towards the negative plate) of the electric force acting on the test charge.

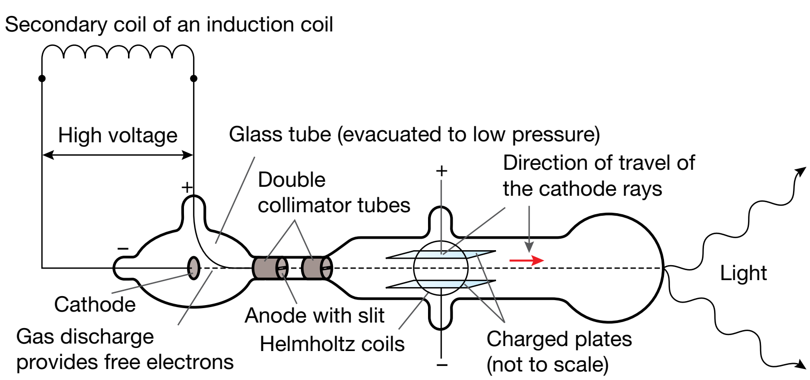
(d) If the plates were moved further apart, what effect, if any, would this have on the size of the electric force acting on the test charge?

13.2.4 The work of J.J. Thomson — determining the charge/mass ratio for cathode rays

J.J. Thomson

The work of English physicist Joseph John Thomson (1856–1940) centred around cathode rays. By incorporating charged plates inside the cathode ray tube, Thomson was able to verify an earlier hypothesis by Crookes that cathode rays would be deflected by electric fields (see Figure 13.8). In Thomson's experiment, the cathode rays passed between parallel plates connected to a battery. He observed that the direction of the rays moved towards the positively charged plate, showing that the rays behaved as negative charges.

FIGURE 13.8 The apparatus for J.J. Thomson's experiments with cathode rays.



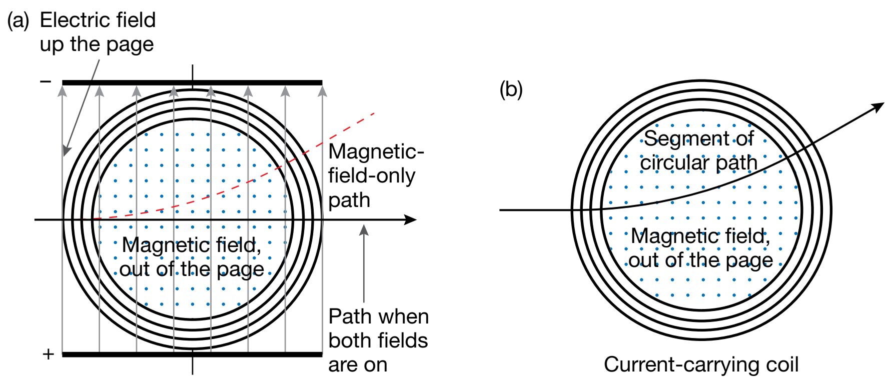
J.J. Thomson was an intuitive and brilliant experimentalist. Following on from his experiment showing that cathode rays were deflected by electric fields, he succeeded in measuring the charge-to-mass ratio of the cathode ray particles, called electrons. Thomson built a cathode ray tube with charged parallel plates (called capacitor plates) to provide a uniform electric field and a source of uniform magnetic field. Using this apparatus, he investigated the effect of cathode rays passing through both fields (see Figure 13.9). The fields were oriented at right angles to each other and this had the effect of producing forces on the cathode rays that directly opposed each other (see the upcoming 'Physics fact').

Thomson's experiment involved two stages:

1. varying the magnetic field and electric fields until their opposing forces cancelled, leaving the cathode rays undeflected. This effect is shown in Figure 13.9a. By equating the magnetic and electric force equations, Thomson was able to determine the velocity of the cathode-ray particles.
2. applying the same strength magnetic field (alone) and determining the radius of the circle path travelled by the charged particles in the magnetic field (see Figure 13.9b).

Thomson combined the results and obtained the magnitude of the charge-to-mass ratio for the charged particles that constituted cathode rays.

FIGURE 13.9 (a) A beam of negatively charged particles left undeflected by the combination of a magnetic field out of the page, and an electric field up the page (b) A negatively charged particle deflected by a magnetic field out of the page. The mechanics of circular motion describes the path, with the centripetal force provided by the magnetic force acting on the particle.



PHYSICS FACT

When charged particles enter an electric field they follow a trajectory under the influence of an electric force (see Figure 13.10).

Similarly, when a charged particle enters a magnetic field, it experiences a magnetic force. The direction of this force is given by the right-hand palm rule (see Figure 13.11).

We can combine these two effects by arranging the electric field, magnetic field and the velocity of the particle at right angles to each other.

For example, by adjusting the strengths of the electric and magnetic fields, their effects on the motion of a charged particle can cancel each other out. The particles can then travel along a straight path.

In Figure 13.8, there are two sets of electric fields. The first accelerates the electrons through a set of collimators to produce a narrow beam. This beam then passes through a combination of electric and magnetic fields that can be adjusted.

FIGURE 13.10

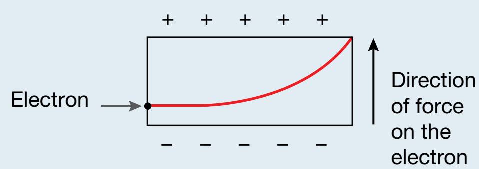


FIGURE 13.11

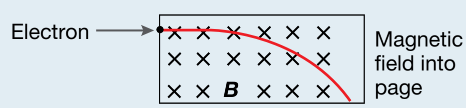
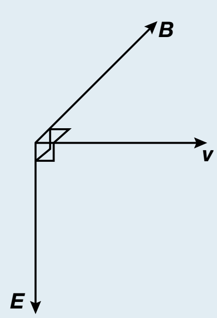


FIGURE 13.12

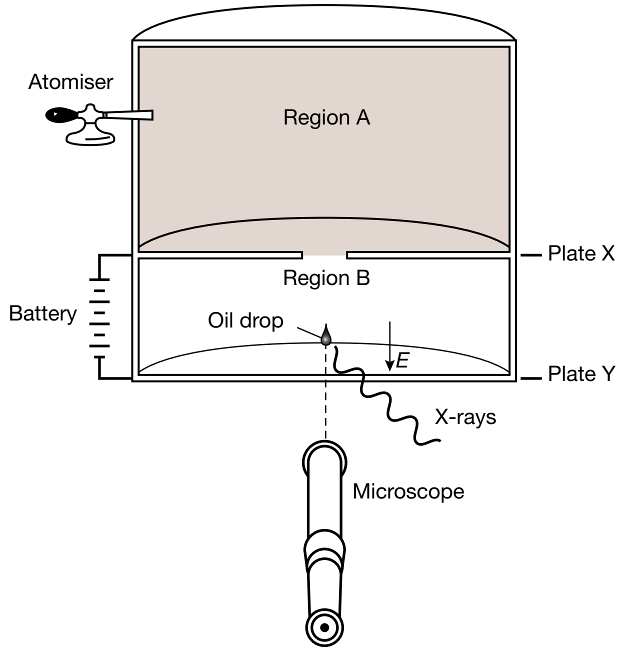


13.2.5 The work of Millikan — determining the quantum of charge — the charge on an electron

In 1909, American physicist Robert A. Millikan was able to use the uniform electric field created between two parallel plates to investigate the properties of a charge. His set-up (see Figure 13.13) involved an atomiser that sprayed a fine mist of oil drops into his apparatus (region A).

Some drops drifted into region B and came under the influence of the electric field, E. As the oil drops entered this region they were momentarily exposed to a beam of X-rays, resulting in some of the oil drops becoming charged. The gravitational field of the Earth exerts a force directed vertically down (weight) that can be counteracted by an electric field produced between parallel plates by a source of variable voltage. The spaces between the plates could be viewed through a microscope. By careful adjustments of the voltage it was possible for one drop to be held stationary, or made to travel with uniform velocity. That is, the forces acting on the drop were balanced.

FIGURE 13.13 Apparatus for Millikan's oil drop experiment.



$$\text{Weight force (down)} = \boldsymbol{F}_g = m\boldsymbol{g}$$
$$\text{Electric force (up)} = \boldsymbol{F}_E = \boldsymbol{E}q$$

For this drop to be suspended between plates, $\boldsymbol{F}_g = \boldsymbol{F}_E$.

Having suspended an oil drop, Millikan could then determine the charge on that particular oil drop by solving for q; that is, $q = \dfrac{m\boldsymbol{g}}{\boldsymbol{E}}$.

Millikan needed to determine the mass of the oil drop. His approach was to measure the terminal velocity of the oil drop when the electric field was turned off and it fell under the force of gravity alone. By using equations from fluid mechanics, he could calculate the radius of the oil drop. By using an oil with a known density, he was able to determine the mass of the oil drop.

Millikan's remarkable findings, for which he won a Nobel prize in 1923, showed that the charge on an oil drop was not of just any arbitrary value. Instead, the charge always occurred in 'packets' or multiples of some smallest value. This value was calculated to be 1.6×10^{-19} C and was called the 'elementary charge', the charge found on an electron.

13.2 SAMPLE PROBLEM 4

An oil drop of mass 6.8×10^{-6} g is suspended between two parallel plates that are separated by a distance of 3.5 mm, as shown in Figure 13.13.

(a) What is the electric field strength between the plates?
(b) What is the charge that must exist on the oil drop?
(c) How many excess electrons must be present on the oil drop?

SOLUTION:

(a) Using the equation $\boldsymbol{E} = \frac{V}{d}$:

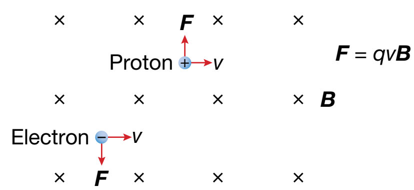
$$\boldsymbol{E} = \frac{110}{3.5 \times 10^{-3}}$$
$$= 3.1 \times 10^4 \text{ Vm}^{-1} \text{down.}$$

13.2.6 The effect of magnetic fields on cathode rays

Magnetic fields exert forces on electric currents, that is, on moving charged particles. If a particle with charge q is moving with velocity v perpendicularly to a magnetic field of strength B, the particle will experience a magnetic force F, given by $F = qvB$.

The direction of the force is given by the right-hand rule. (If the particle has a positive charge, the direction of the conventional current is that of the velocity; if the particle has a negative charge, the direction of the conventional current is opposite to that of the velocity.) This is illustrated in Figure 13.14 for an electron (negative charge) and a proton (positive charge).

FIGURE 13.14 Forces on an electron and a proton moving perpendicularly to a magnetic field.



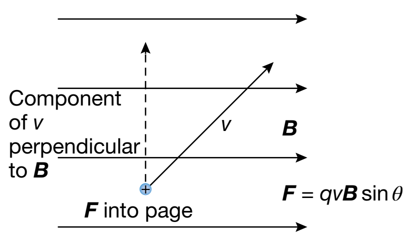
If the velocity is at an angle θ to the magnetic field, the force is given by

$$\boldsymbol{F} = qv\boldsymbol{B} \sin \theta.$$

To find the direction of the force, use the component of the velocity perpendicular to the magnetic field and the right-hand rule. This is illustrated in Figure 13.15 where the direction of the force is into the page.

If the charged particle is moving parallel to the magnetic field, $\theta = 0$, and therefore $F = 0$.

FIGURE 13.15 Force on a charged particle moving at an angle, θ, to a magnetic field



13.2 SAMPLE PROBLEM 5

An electron of charge -1.6×10^{-19} C is projected into a region where a magnetic field exists, as shown in Figure 13.16. If the velocity of the electron is 2.5×10^4 m s^{-1}, determine:

(a) the force on the electron at the instant it enters the magnetic field

(b) the shape of the path that the electron follows.

SOLUTION:

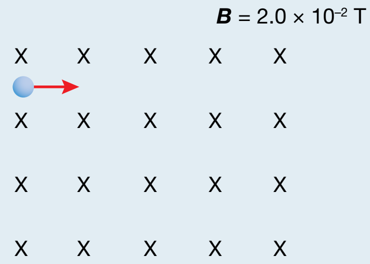
(a) $F = qvB \sin\theta$

$= 2.0 \times 10^{-2} \times 1.6 \times 10^{-19} \times 2.5 \times 10^{4}$

$= 8.0 \times 10^{-17}$ N downwards.

(b) The path that the electron follows will be circular. This is because the magnetic force is always acting perpendicular to the velocity of the electron.

FIGURE 13.16



13.2 Exercise 2

1 A pair of horizontal parallel plates (45 mm separation) is arranged such that the potential difference applied across the plates is 56 V. An oil drop is suspended between the plates in equilibrium.

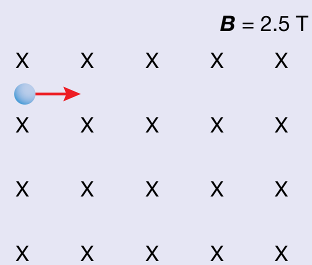
(a) Which of the plates, top or bottom, would be positive?

The oil drop is known to have excess electrons equal to 5.0×10^{8}.

(b) Determine the mass of the oil drop.

2 A horizontal beam of electrons is fired into a region containing a uniform magnetic field of 2.5 T directed vertically downwards.

FIGURE 13.17



(a) The electrons in the beam followed a circular path. Explain why.

(b) An analysis of the path taken by the electrons determined the force acting on each of them to be 2.0×10^{-15} N. Use this information to calculate the velocity of the electrons.

13.2.7 Cathode rays — waves or particles

In 1875, twenty years after their discovery, William Crookes (1832–1919) designed a number of tubes to study cathode rays (some of these are shown in Figure 13.3). He found that the cathode rays did not penetrate metals and travelled in straight lines. It was initially thought that the rays may be an electromagnetic wave because of the similarity in their behaviour to light. This was discounted when Crookes discovered that the cathode rays were deflected by magnetic fields, an effect that did not occur with light.

In a paper read to the Paris Academy of Science in 1885, Jean Perrin (1870–1942) described the two main hypotheses concerning the nature of cathode rays:

'Some physicists think, with Goldstein, Hertz and Lenard, that this phenomenon is like light of very short wavelength. Others think, with Crookes and J.J. Thomson, that these rays are formed by matter which is negatively charged and moving with great velocity, and on this hypothesis their mechanical properties, as well as the manner in which they curve in a magnetic field, are readily explicable.'

The way that physicists set out to understand the nature of cathode rays shows how the scientific method is used to solve problems. That is, observations from experiments are interpreted and a hypothesis developed to explain what is thought to be happening. Opposing models may arise, with supporters of each side arguing strongly for their belief. The argument may eventually be resolved either by improved experiments or with greater understanding of the phenomenon.

In this case, the debate about whether cathode rays were electromagnetic waves or streams of charged particles remained unsolved until 1897, when J.J. Thomson showed beyond doubt that the rays were streams of negatively charged particles, which we now call electrons. Why was the debate so prolonged? The problem was the apparently inconsistent behaviour of rays. For example, the following observations from cathode ray experiments fitted the wave model:

- they travelled in straight lines
- if an opaque object was placed in their path, a shadow of that object appeared
- they could pass through thin metal foils without damaging them.

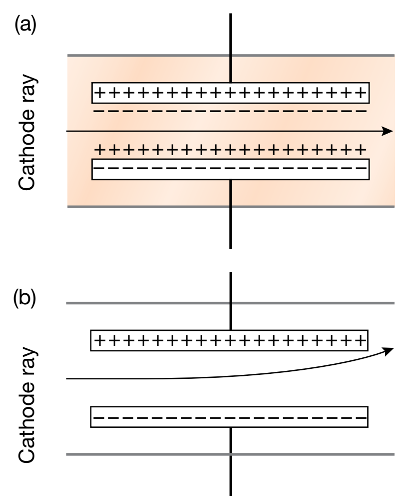
The following observations fitted the particle model:

- the rays left the cathode at right angles to the surface
- they were obviously deflected by magnetic fields
- they did not appear to be deflected by electric fields
- small paddlewheels turned when placed in the path of the rays
- they travelled considerably more slowly than light.

The main restriction for the charged particle theory was the absence of deflection in electric fields. However, Thomson showed that this was due to the rays themselves. He stated:

'... on repeating the experiment I first got the same result, but subsequent experiments showed that the absence of deflection is due to the conductivity conferred on the rarefied gas by the cathode rays. On measuring this conductivity ... it was found to decrease very rapidly with the exhaustion of the gas ... at very high exhaustions there might be a chance of detecting the deflection of cathode rays by an electrostatic force.'

FIGURE 13.18 The path of cathode rays (a) at high gas pressure and (b) at low gas pressure.



Within the tube, the cathode rays ionised the gas. The ions were attracted to the plate with the opposite charge and the line-up of ions effectively neutralised the charge on the plate, allowing the cathode rays to pass by unaffected.

After evacuating the chamber, Thomson observed deflection and that the particles were always deflected towards the positive plate, which confirmed that they were negatively charged particles. The deflection of cathode rays in tubes of different gas pressure is shown in Figure 13.18.

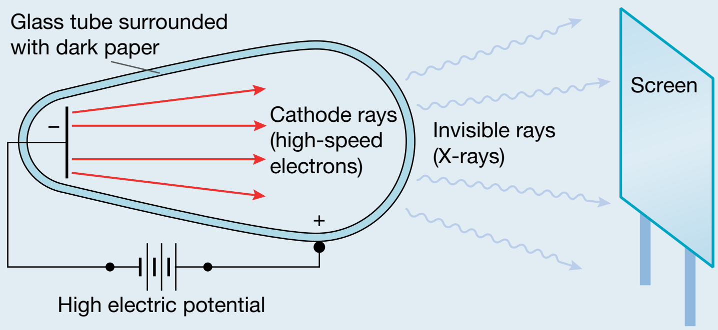
The ability of cathode rays to penetrate thin metal foils was still unexplained. The answer lay, not simply with the properties of cathode rays, but with the model of the atom. If the atom was not a solid object, but much more open, it might be possible for very small particles to pass through thin foil. Although not considered at this time, Ernest Rutherford (1871–1937) would use a similar approach to change the model of the atom.

PHYSICS FACT

X-rays: discovery and application

In 1895, a type of radiation was discovered by Wilhelm Röntgen (1845–1923) while he was experimenting with cathode rays. He found that, in a dark room, a screen covered with a sensitive fluorescent material (barium platino-cyanide) glowed when it was placed near the end of a cathode ray tube (see Figure 13.19).

FIGURE 13.19 Röntgen's apparatus



Since cathode rays could not pass through the glass at the end of the tube, he deduced that this fluorescence must be due to a new form of radiation. He called this radiation 'X-rays' as their properties were not known. Later research showed that X-rays were produced when high-speed electrons interacted with matter, such as the glass in the cathode ray tube.

X-rays were later found to be electromagnetic waves, similar to light but with a much smaller wavelength.

Among the many characteristics that make X-rays so useful are the fact that they can:

- penetrate many substances
- expose photographic film
- cause certain substances to fluoresce
- be reflected and refracted.

The most common use of X-rays is in the field of medicine, for diagnosing illness or injury as well as treating illnesses such as cancer. X-ray machines are used widely — to check luggage at airports, analyse the welding of metal parts in an aircraft wing, and look at things that we otherwise could not see.

FIGURE 13.20 X-ray of a normal knee joint (human).

13.3 The development of the classical model of the atom

13.3.1 Development of the Thomson 'plum pudding model'

TABLE 13.1 A timeline of modern physics.

Year	Event
1864	James Clerk Maxwell develops the mathematical theory of electromagnetism.
1875	Sir William Crookes observes cathode rays.
1880s	Michelson–Morley experiments
1885	Heinrich Hertz produces electromagnetic waves artificially.
1888	Heinrich Hertz discovers the photoelectric effect.
1895	Wilhelm Roentgen discovers X-rays.
1896	Henri Becquerel discovers radioactivity.
1897	J.J. Thomson discovers the electron.
1899	Max Planck describes energy 'quanta'.
1905	Albert Einstein explains the photoelectric effect and publishes the Special Theory of Relativity.
1911	Ernest Rutherford proposes the 'solar system' model of the atom.
1912	Robert Millikan determines the electric charge on the electron.
1913	Niels Bohr proposes the quantum model of the atom.
1914	James Franck and Gustav Hertz experimentally demonstrate the existence of discrete energy states of atoms.
1919	The proton is discovered.
1920	Rutherford proposes the existence of the neutron.
1923	Louis de Broglie proposes that electrons have a wave nature.
1926	Max Born, Werner Heisenberg and Pascual Jordan develop the theory of mechanics.
1927	Werner Heisenberg develops the uncertainty principle.
1932	James Chadwick experimentally confirms the existence of the neutron.

On Christmas Eve 1899, just a few days before the beginning of the new century, Max Planck first described the energy of atoms in terms of packets called 'quanta'. Looking back now, many physicists regard this moment as the birth of what we now refer to as modern physics. This new form of physics arose from the physics discoveries of the previous four hundred years or so — what is generally known now as classical physics.

The nineteenth century hummed with discovery, and the sweeping changes that came with the Industrial Revolution in Europe led many to believe that Science had at last mastered the natural world. Building upon the prodigious work of Isaac Newton and his contemporaries in the seventeenth century who had finally drawn together the threads of motion, mass and force to explain gravity and the movement of heavenly bodies, the new scientists such as Michael Faraday and H.F.E. Lenz went on to explain electricity and magnetism. In the early 1800s, Thomas Young in England and Augustin Fresnel in France confirmed the wave nature of light, settling a long-standing dispute as to the nature of light and overturning the favoured theory that light was made up of particles. Around 1870, James Clark Maxwell theorised that accelerating electric charges would produce electromagnetic radiation, the existence of which was first demonstrated by Heinrich Hertz in 1887; Hertz later went on to confirm that light itself was a form of electromagnetic radiation.

Another aspect of light studied by early physicists such as J. Fraunhofer was that of the characteristic spectra that gases produced when heated or when an electric discharge was passed through them. We will look at this in a little more detail as spectra of gases play an important part in the development of quantum theory.

The invention of an efficient vacuum pump by Heinrich Geissler in 1855 led to the discovery of a different type of electric discharge that was able to travel through a tube that had been evacuated to a very low pressure. The nature of these cathode rays, as they were called, was not explained until 1897 when J.J. Thomson identified them as being made up of negative particles called electrons.

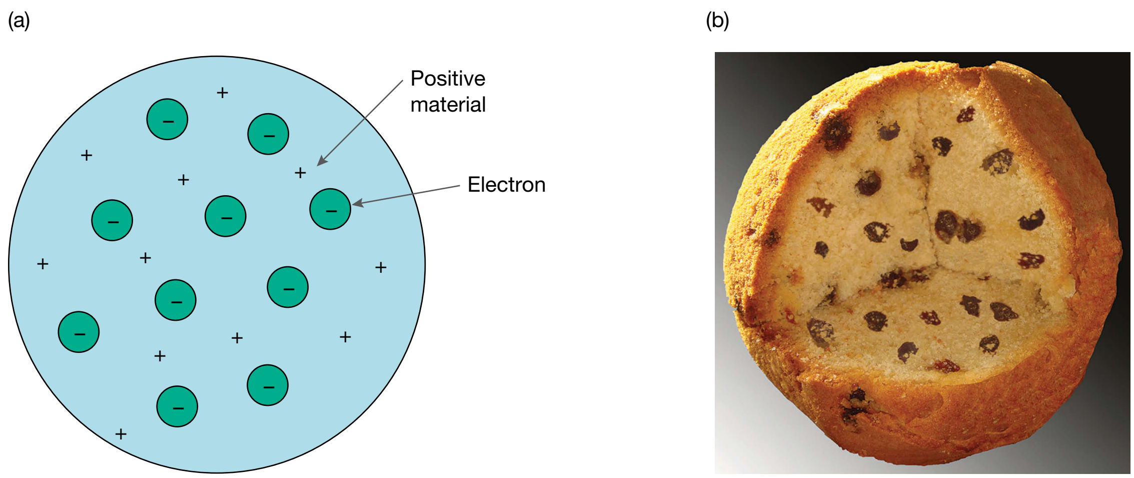
However, for all the discoveries that were made, the end of the nineteenth century left many questions still unanswered and had even produced new questions. Some optimistically declared that these were merely minor details that would be eventually cleared up. In fact, what had been deemed ‘minor details’ led to the development of modern physics and an entirely new way of looking at the universe.

13.3.2 Rutherford’s model of the atom

A precursor to the Rutherford model of the atom was the J.J. Thomson model of the atom, often referred to as the ‘plum-pudding model’. Thomson was able to demonstrate that a mysterious radiation known at the time as ‘cathode rays’ consisted of negatively charged particles with a charge-to-mass ratio of approximately $1.8 \times 10^{11}\ \mathrm{C\,kg^{-1}}$. This was achieved by passing a beam of cathode rays through a magnetic field and observing the beam to curve, indicating the rays were, in fact, negatively charged particles. He named these particles ‘electrons’.

Thomson’s discovery led him to a new model of the atom. The atom, according to Thomson, must consist of a positively charged fluid with his negatively charged electrons scattered throughout the atom. The model was known as the plum pudding model of the atom because Thomson envisioned that the electrons were strewn through the positive body of the atom just as plums are in a plum pudding.

FIGURE 13.21 (a) Thomson’s plum pudding model of the atom. (b) Distribution of plums in a plum pudding.



One of the younger physicists who assisted J.J. Thomson in his electron experiments was the New Zealander Ernest Rutherford. After leaving Thomson's laboratory, Rutherford went to McGill University in Montreal, Canada, where he extensively studied how alpha particles (a form of radiation) were deflected by a variety of materials and how the paths of alpha particles were affected by electric fields.

Rutherford did nothing more with alpha particle scattering until he moved to Manchester, England, in can be made by placing1907, where he inherited Hans Geiger, a German physicist who specialised in the detection of radioactivity, as his assistant. Rutherford was inspired to return to his investigations of the scattering of alpha particles, this time by very thin metal foils. As part of these experiments, Rutherford suggested to an undergraduate student named Ernest Marsden (who was being trained in radioactive detection techniques by Geiger) that he should investigate if alpha particles fired at thin metal foils were deflected in the same way as they were by other materials. Rutherford expected that all the alpha particles would pass through the thin foils with only a small deflection from their original path — a matter of a few degrees of angle change. However, Marsden observed that, while most of the alpha particles fired at a thin gold foil were deflected from their path by a small amount (if at all), a very small fraction of the alpha particles (about 1 in 8000 particles) were deflected by angles greater than 90°; some of these were found to be reflected back in the direction from which they had come! In one of the last lectures that Ernest Rutherford ever gave, he described his reaction to Marsden's discovery of deflection of alpha particles through large angles as 'the most incredible event that has ever happened to me in my life. It was almost as incredible as if you had fired a fifteen-inch shell at a piece of tissue paper and it came back and hit you'.

FIGURE 13.22 Ernest Rutherford (1871–1937).

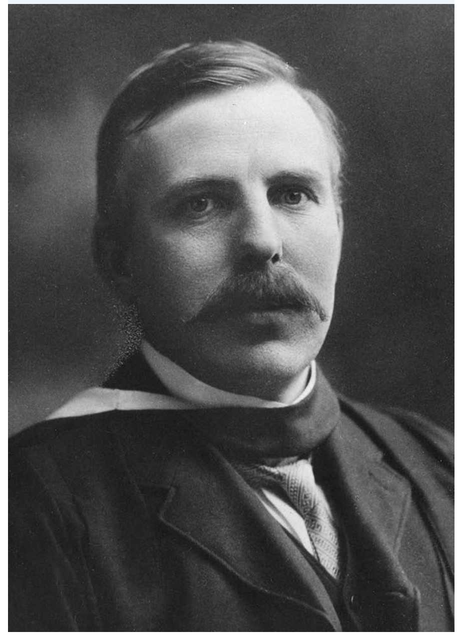
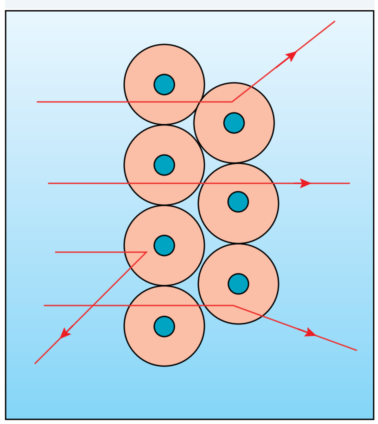


FIGURE 13.23 The deflection of alpha particles by atoms in a thin gold foil in Marsden's experiment.



Ernest Rutherford was awarded the Nobel Prize in Chemistry in 1908. The award that year was a matter of intrigue. In an attempt to award Nobel prizes to two atomists (Planck and Rutherford) in the same year, Dr Arrhenius, Director of the Nobel Institute for Physical Chemistry, arranged for Rutherford to be nominated for the chemistry prize and for that prize to be determined before the physics prize. In the end, Planck was opposed for the physics prize, which he was eventually awarded ten years later.

Looking at the results of the experiments of Geiger and Marsden, Rutherford realised that they were similar to those that you would expect when two charged bodies interacted — an electrostatic force arising between the foil atoms and the alpha particles or if one mass had collided elastically with another. However, according to Thomson's model of the atom, electrons were small, negatively charged and distributed evenly through a positive material. Rutherford knew that such small charges as the electrons could have little electrostatic effect on the much larger alpha particles. It came to him that the distribution of the positive and negative charge in an atom was not uniform at all, but that the majority of the atom was made up of empty space with all of the positive charge located in a mass in the centre with the much smaller negative charges around the outside.

Two years after Geiger and Marsden published a paper on the deflection of alpha particles by thin metal foils, Rutherford explained their results by proposing a nuclear atom. In this model, electrons orbited around the outside of a much more massive positive centre or 'nucleus' much as the planets orbit the sun. (For this reason, the Rutherford model is sometimes referred to as the 'planetary model of the atom'.) This large nucleus contained 99.9% of the atom's overall mass. When the alpha particles passed through the empty regions of the atoms far from the nucleus, they were deflected from their paths by only a small amount; the closer to the nucleus their path was, the greater the electrostatic repulsion that they experienced and, so, the greater the deflections.

On the basis of his nuclear model, Rutherford calculated the relative numbers of alpha particles that would be scattered through different angles. When Geiger began a series of careful experiments in which he made observations of the number of alpha particles scattered through different angles, his experimental results matched Rutherford's predictions. Rutherford's model of the atom replaced that of Thomson.

FIGURE 13.24 The apparatus used in 1911 to study alpha particle scattering. In this version, the microscope and scintillation screen can be rotated to observe the alpha particles at different angles. Polonium was used as the alpha particle source, the metal foil used was gold and the apparatus was evacuated.

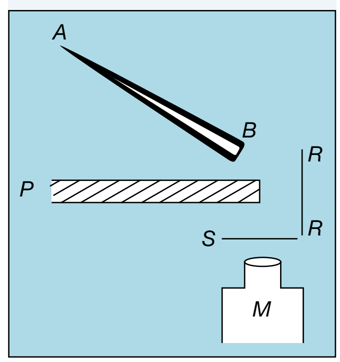
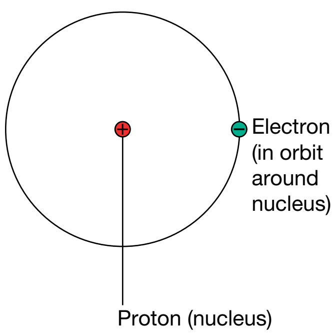


FIGURE 13.25 Diagram of the Rutherford model of the atom of hydrogen.



The radius of a hydrogen atom is about 2.1×10^{-11} m. The radius of a proton is about 0.85 femtometres (0.85×10^{-15} m). Physicists sometimes call this unit a fermi, named after Enrico Fermi. The ratio of the radius of this atom to the radius of its nucleus is about 2.5×10^4. This would make it very difficult to construct an accurate scale model of an atom in your laboratory. If your laboratory was 10 m across and this represented the diameter of the atom, the diameter of the nucleus would have to be 4×10^{-4} m or 4 tenths of a millimetre in diameter!

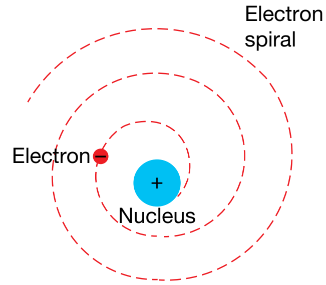
13.3.3 Limitations of the Rutherford model of the atom

While Rutherford's model provided a better explanation of observed experimental results, it had a number of shortcomings. First, if the electrons were in orbit around the nucleus, they would be accelerating and so, according to Maxwell's theories of electromagnetism, they would be emitting electromagnetic radiation. However, in giving off this radiation, the electrons would then have to be losing energy and their orbits would deteriorate until they finally spiralled into the positive nucleus. If this were the case, then:

- The frequency of the radiation given off by the electrons as they spiralled closer and orbited faster would become increasingly larger. In reality, this was never observed.
- The atoms would be unstable. Once again, this was not true as the general stability of matter indicates that atoms themselves are also very stable.

The second difficulty that arose from the Rutherford model was that it was unable to explain why atoms emitted characteristic spectral lines. This was a puzzle that had yet to be solved.

FIGURE 13.26 Rutherford's model implied that electrons would lose energy and spiral into the nucleus.



13.4 The neutron

13.4.1 The discovery of the neutron

After the discovery of the nucleus, it seemed logical to assume that the nucleus contained protons and electrons. It was possible to explain radioactive transmutations in terms of emission of alpha and beta particles from the nucleus of protons and electrons. However, there were major problems with the idea of a nucleus containing these constituents.

In 1920, Rutherford proposed that a neutral particle, with mass comparable to that of a proton, must be another constituent of the nucleus. He named this particle a neutron. In future research he and James Chadwick (1891–1974) continued to look out for any result that would suggest the existence of such a particle.

Experiments involving artificially induced radioactivity

The radioactivity that we have encountered so far has been associated with natural alpha, beta and gamma emitters. Rutherford was the first to use alpha particles to produce nuclear reactions.

The first artificially induced transmutation

In 1919, Rutherford bombarded nitrogen gas with alpha particles from bismuth-214. A positively charged particle that was more penetrative than an alpha particle was produced. This particle was identified as a proton.

What had occurred, as shown in Figure 13.27, was that the alpha particle had combined with the nitrogen nucleus and a proton had been emitted. The alpha particles from the bismuth-214 source were able to approach the nucleus very closely and occasionally make contact with it. The equation for this reaction is:

$$^{4}_{2}\text{He} + ^{14}_{7}\text{N} \rightarrow ^{17}_{8}\text{O} + ^{1}_{1}\text{H}$$

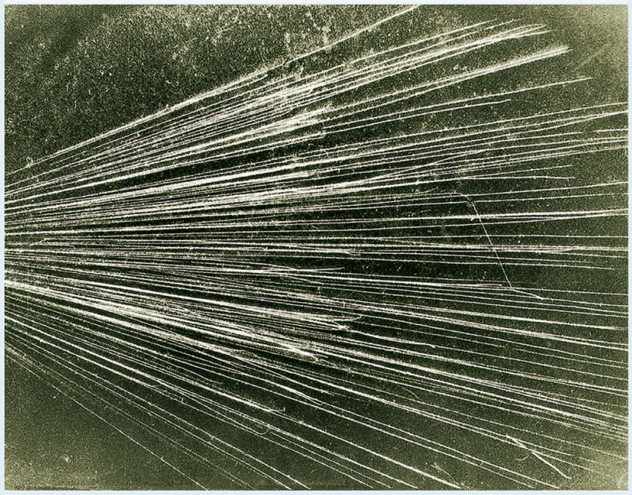
PHYSICS FACT

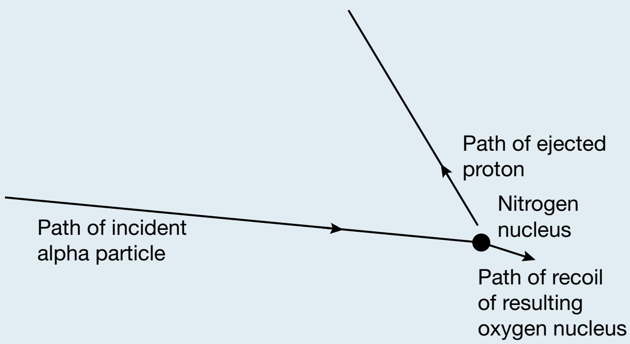
Alpha-particle-induced nuclear reactions

When the first alpha particle scattering experiments were performed, low-energy alpha particles were used and those that approached a gold nucleus (containing 79 protons) were strongly repelled. In the alpha-particle-induced reaction with nitrogen, the alpha particles had a much higher energy than those used in the early experiments and there was only a weak repelling force from a nitrogen nucleus that contained only 14 protons. An energetic alpha particle was able to make contact with the nitrogen nucleus.

Various writers have commented that Rutherford was fortunate that he did not use a source of very powerful alpha particles when he performed his first alpha particle scattering experiments!

FIGURE 13.27 The photograph shows alpha particle tracks through a cloud chamber filled with nitrogen gas. The diagram helps to identify an event where a nitrogen nucleus has been struck by an alpha particle. A proton is ejected upwards and the resulting oxygen nucleus recoils downwards.





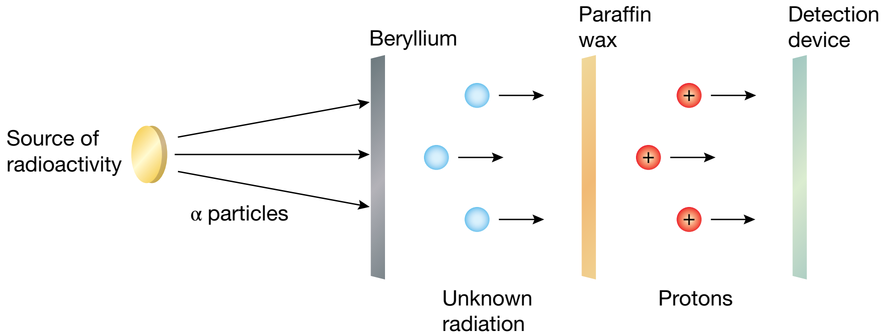
An artificially induced radioactivity

In 1930, Bothe and Becker (in Germany) fired alpha particles at beryllium and found that a highly penetrating radiation was produced. The radiation seemed to be similar to gamma rays (high-energy photons) but it was much more highly penetrating than the gamma rays previously observed. It was found to have an energy of about 10 MeV, again much higher than that previously observed for gamma rays.

In France, Frédéric Joliot (1900–1958) and his wife Irène Curie (1897–1956) (daughter of Marie Curie), studied this mysterious radiation and let it fall on a block of paraffin. Paraffin is a hydrocarbon very rich in hydrogen atoms. They found that the radiation knocked protons (hydrogen nuclei) from the paraffin (see Figure 13.28). The energy of the protons was about 5 MeV. Of course, now that charged particles (protons) were involved, it was much easier to determine their properties. They also found that many more protons than expected were emitted from the paraffin. If gamma rays had been responsible, their very high penetrating power would have resulted in fewer interactions with protons.

The high energy of the protons (5 MeV) was a problem because applying the conservation of energy and conservation of momentum to the collision between a gamma ray and a proton yielded a value for the incident gamma ray of at least 50 MeV. This was a major dilemma because the energy of the incident alpha particles was only about 5 MeV. In other words, if this was the correct interpretation, there had to have been a tenfold increase in energy in the interaction!

FIGURE 13.28 The reaction of alpha particles with beryllium produced a mysterious radiation that knocked protons out of paraffin.



PHYSICS FACT

Rutherford's prediction of the neutron

In his Bakerian lecture of 1920, Rutherford had suggested that 'it may be possible for an electron to combine much more closely with the hydrogen nucleus than is the case in the ordinary hydrogen atom'. He later used the term neutron. It is worth noting that Rutherford's conjecture about the existence of the neutron had not received wide publication and it had not been read by either Joliot or his wife. Some years later Joliot commented on the fact that he had not read Rutherford's Bakerian lecture and that, had he done so, it was possible or probable that he and his wife would have identified the neutron before Chadwick.

13.4.2 Chadwick identifies the neutron

Just over two weeks after reading the paper of the Joliot–Curies, James Chadwick (1891–1974) had completed his work and submitted a paper on 'Possible existence of a neutron' (1932). In that time, Chadwick applied conservation of energy and conservation of momentum to the interaction of a neutral particle (of mass similar to that of a proton) with a proton. Chadwick made measurements of the recoil of nuclei of hydrogen and nitrogen after interactions with his proposed neutron. The measurements were difficult but led to the mass of a neutron being calculated to be 1.15 times that of a proton.

At this time (1932), there was doubt expressed about whether or not the conservation laws of classical physics would apply to nuclear processes. Some leading physicists were adamant that they would but others, including Bohr, thought otherwise. In fact it was 1936 before Bohr dropped his ideas of non-conservation of energy.

As Chadwick's neutron identification depended on the conservation laws and there was doubt expressed about them at the time, he concluded his paper 'Up to the present, all the evidence is in favour of the neutron ... [unless] the conservation of energy and momentum be relinquished at some point'.

The nuclear equation for the reaction of alpha particles with beryllium is:

$$^{9}_{4}\text{Be} + ^{4}_{2}\alpha \rightarrow ^{12}_{6}\text{C} + ^{1}_{0}\text{n}.$$

There were no naturally occurring neutron emitters but now, with a high-energy alpha particle source (such as polonium) and some beryllium, it was possible to produce neutrons and conduct neutron scattering experiments.

PHYSICS FACT

Problems with electrons and protons in close association

It is worth noting that there are major difficulties with the concept of electrons and protons in close association either as a single particle (the neutron) or generally in the nuclei of atoms. The masses of atoms could not be explained in terms of numbers of protons and electrons. Another problem involved the de Broglie wavelength of an electron. How could an electron with an energy of a few MeV be confined to a region with a radius of 5×10^{-15} m when its de Broglie wavelength was large compared to this radius?

These difficulties were overlooked at the time because, after all, an alpha particle seemed to be 4 protons and 2 electrons combined very tightly together.

13.5 Review

13.5.1 Summary

- Cathode ray tubes were used to investigate the properties of cathode rays.
- Cathode rays were found to be negatively charged particles.
- Charged parallel plates produce a uniform electric field.
- The strength of a uniform electric field, E, in volts per metre, produced by parallel plates, separated by a distance, d, and charged by an applied voltage, V, is given by:

  $$E = \frac{V}{d}$$

- Millikan showed that charge came in discrete packets, multiples of the smallest amount: 1.6×10^{-19} C. This was achieved by having oil drops in vertical equilibrium so that an upward electric force balanced a downward gravitational force such that $q\boldsymbol{E} = m\boldsymbol{g}$, allowing $q = \frac{m\boldsymbol{g}}{\boldsymbol{E}}$ to be used to determine the charge, q.
- A charged particle moving with a velocity, $\boldsymbol{v}$, at an angle, θ, through a magnetic field of strength, $\boldsymbol{B}$, experiences a force, $\boldsymbol{F}$. The magnitude, in Newtons, is given by:

  $$\boldsymbol{F} = q\boldsymbol{v}\boldsymbol{B}\sin\theta.$$

  $\boldsymbol{F}$, $\boldsymbol{v}$ and $\boldsymbol{B}$ are all vector quantities, and each has a direction associated with it.
- Thomson's experiment, using perpendicular electric and magnetic fields, allowed him to determine the charge-to-mass ratio of an electron. The value he determined was 1.759×10^{11} C kg^{-1}.
- The discovery of cathode rays, which were identified by Thomson to be electrons, led Thomson to a new model of the atom, one that contained electrons immersed in a positively charged fluid.
- Rutherford's model of the atom had a central positively charged mass called the nucleus that was orbited by electrons.
- The Rutherford model of the atom explains alpha particle scattering very well. The planetary model of electron orbits necessary for the Rutherford model means that atoms self-destruct due to accelerating charged particles emitting light in accordance with Maxwell's equations.

- The properties of the nucleus could not be explained by assuming it contained protons and electrons. The existence of a new particle was predicted; this new particle was called the neutron and it was discovered in 1932 by Chadwick.

13.5.2 Questions

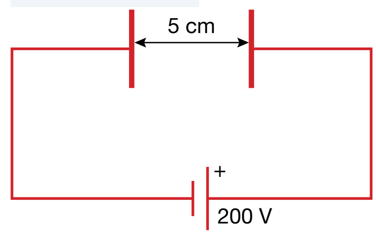
1. Why can cathode rays be observed and manipulated within a vacuum tube and not in air?
2. Look up the meaning of 'conservation of charge'. In a discharge tube, cathode rays were formed and moved from the cathode (the negative electrode). If these rays carried an electric charge, where was the corresponding amount of positive charge?
3. Draw the electric field lines between a positive and negative charge of equal magnitude. In which area is the electric field strongest?
4. Calculate the electric force on a charge of 1.0×10^{-6} C placed in a uniform electric field of $20\,\text{N}\,\text{C}^{-1}$.
5. Draw the electric field lines between two parallel plates placed 5.0 cm apart.
6. The electric field between parallel plates can be considered 'uniform' only in the region between the plates that is well away from the edges of the plates. What is meant by this statement?
7. A pair of parallel plates is arranged as shown in Figure 13.29. The plates are 5.0 cm apart and a potential difference of 200 V is applied across them.
   Data:

   Charge on electron $= -1.6 \times 10^{-19}$ C

   Mass of electron $= 9.1 \times 10^{-31}$ kg

   Mass of proton $= 1.67 \times 10^{-27}$ kg

   FIGURE 13.29

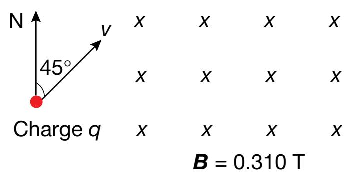
   

   (a) Calculate the magnitude and direction of the electric field between the plates.
   (b) Calculate the force acting on an electron placed between the plates.
   (c) Calculate the force acting on a proton placed between the plates.
   (d) Explain why these two forces are different.
   (e) Calculate the work done in moving both the electron and the proton from one plate to the other.

8. Negatively charged latex spheres are introduced between two charged plates and are held stationary by the electric field. Each sphere has a mass of 2.4×10^{-12} kg and the strength of the field required to counter their weight is $4.9 \times 10^{7}\,\text{N}\,\text{C}^{-1}$. Sketch this arrangement, identifying the positive and the negative plate, and determine the charge on the spheres.
9. Two parallel plates are separated by a distance of 10.0 cm. The potential difference between the plates is 20.0 V.
   (a) Calculate the electric field between the plates, assuming the field to be uniform.
   (b) A charge of $+2.0 \times 10^{-3}$ C is placed in the field. Calculate the force acting on this charge.
10. A beam of electrons moves at right angles to a magnetic field of flux density 6.0×10^{-2} T. The electrons have a velocity of $2.5 \times 10^{7}\,\text{m}\,\text{s}^{-1}$. What is the magnitude of the force acting on each electron?
11. A stream of doubly ionised particles (missing two electrons and therefore carrying a positive charge of twice the electronic charge) move at a velocity of $3.0 \times 10^{4}\,\text{m}\,\text{s}^{-1}$ perpendicular to a magnetic field of 9.0×10^{-2} T. What is the magnitude of the force acting on each ion?
12. An electron is travelling at right angles to a magnetic field of flux density 0.60 T with a velocity of $1.8 \times 10^{6}\,\text{m}\,\text{s}^{-1}$. What is the force experienced by the particle?
13. Given that the mass of the electron in question 12 is 9.1×10^{-31} kg, what is the acceleration of the particle in the direction of the force acting on it?

14. A charge of 5.25 mC, moving with a velocity of 300 m s^{-1} due north east, enters a uniform magnetic field of 0.310 T directed vertically downwards, into the page. Calculate the magnetic force on the charge.

FIGURE 13.30



15. If charged particles enter a magnetic field at angles other than at right angles, describe their path.
16. List the properties of cathode rays that can be described:
    (a) as wave motion
    (b) by a particle model.
17. Explain how the properties of cathode rays were demonstrated using the evacuated tubes in which a metal cross was mounted in the path of the rays, and in which a small paddle wheel was able to roll along glass rails.
18. Describe the path of an electron when it enters the region between parallel plates across which a potential difference of 1500 V is applied. Sketch the arrangement for an electron entering with a horizontal velocity of 2.4×10^4 m s^{-1} at right angles to the electric field.
19. Describe the conditions needed for an electron entering a magnetic field to undergo uniform circular motion.
20. In a tube similar to that used in Thomson's electromagnetic experiment, a magnetic field of 1.00×10^{-2} T is sufficient to allow the electrons to pass through the electric deflection plates. The plates are 10 mm apart and have a potential difference of 300 V across them.
    (a) What is the strength of the electric field between the plates?
    (b) What was the speed of the electrons as they entered the region between the plates?
    (c) What was the strength of the magnetic force acting on the electrons?
21. What is the evidence to support the claim that an atom consists of a small positively charged nucleus whose size is approximately $\frac{1}{10\,000} \times$ the size of an atom?
22. What is the problem that arises with an atomic model where electrons are in a planetary-type circular orbit around the nucleus?
23. Write out the equation for the nuclear reaction that Chadwick made use of to produce free neutrons.
24. There are two fundamental principles in physics that were used to assert that a neutral particle of mass similar to that of a proton must exist. What are the names of these two fundamental principles?
25. Once the neutron was isolated in the early 1930s, why did the atomic masses of elements make more sense?
26. Why are neutrons difficult to detect?
27. Why is the existence of a neutron inside a nucleus necessary to better explain beta decay?

eBook*plus* RESOURCES

Complete this digital doc: Investigation: Discharge tubes
Searchlight ID: doc-26321

Complete this digital doc: Investigation: Properties of cathode rays
Searchlight ID: doc-26322

PRACTICAL INVESTIGATIONS

Investigation 13.1: Discharge tubes

Aim

To observe the effect that different gas pressures have on an electric discharge passed through a discharge tube.

Apparatus

power pack
two plug–plug leads
one set of discharge tubes (with varying pressures)
induction coil
two plug–clip leads

Theory

The high voltage produced by the induction coil is applied across the terminals inside the discharge tubes. One plate (the cathode) becomes highly negative and releases a ray (cathode ray or electron). The electron passes through the gas in the tube and excites electrons in the atoms of the gas contained in the tube. The pressure of the gas determines the density of the atoms and therefore the nature of the collisions that take place between the electrons and atoms. Therefore, different discharge effects under different pressures can be observed (refer back to Figure 13.4).

Method

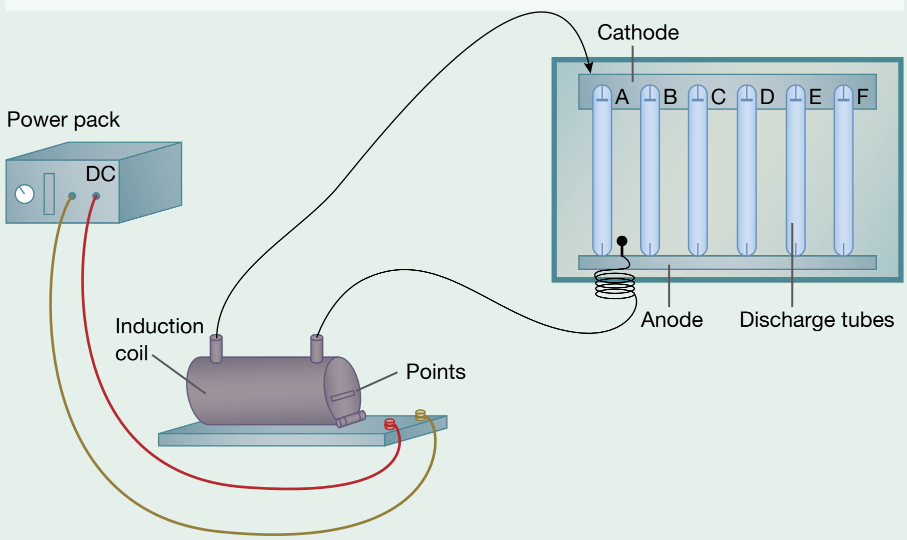
Safety note

When the induction coil is connected to the discharge tube, X-rays are produced. However, it is the cathode rays hitting the glass or metal within the discharge tube that creates the X-rays, not the induction coil. If the experiment uses a minimum operating voltage, these X-rays will be of a low energy and are significantly reduced after passing through the glass.

We need to deal with induction coils with extreme care because of the high voltages associated with them. Your teacher will set up the equipment related to the induction coil.

1. Attach the induction coil to the power pack using the two plug–plug leads. Adjust the points on the induction coil to obtain a continuous spark from the coil. Switch off the power pack.

FIGURE 13.31 Set-up for Practical Investigation 10.1.



2. Set the power pack at the correct setting for the induction coil (usually 6 volts) and turn it on.
3. Attach the negative terminal of the induction coil to the cathode of the discharge tube marked with the highest pressure (40 mm Hg) and attach the positive terminal to the other end as shown in Figure 13.31. Switch on the power pack.
4. Sketch a diagram of the pattern observed in this tube and describe it carefully.

5. Repeat the above procedure using each of the discharge tubes and see if you can observe streamers, Faraday's dark space, cathode glow, Crookes' dark space, striations and the positive column. Carefully describe each pattern, identifying each of the effects mentioned. (Tubes to be used should be 40 mm Hg, 10 mm Hg, 6 mm Hg, 3 mm Hg, 0.14 mm Hg and 0.03 mm Hg . These are represented as A, B, C, D, E and F in Figure 13.31).

Questions

1. What effects were common throughout all tubes?
2. If the striations are produced by electrons (cathode rays) striking atoms and causing light to be released, give an explanation for the occurrence of variation in the patterns for different pressures.

Investigation 13.2: Properties of cathode rays

Aim

To determine some of the properties of the rays that come from the cathode of a discharge tube.

Apparatus

two power packs
two plug–plug leads
one pair of magnets
induction coil
four plug–clip leads
discharge tubes (maltese cross, electric plates, rotating wheel, screen display)

Theory

This experiment will most likely be performed as a class demonstration by your teacher. The discharge tubes used are illustrated in Figure 13.3 and are similar to those Sir William Crookes would have used.

Method

Before starting, it would be advisable to read the 'Analysis' section of this experiment so as to plan what you should record during the experiment.

1. Connect the power pack to the induction coil and set it at 6 volts. Adjust the points on the induction coil so that a strong steady spark is being produced, as in practical investigation 13.1.
2. Connect the terminals of the induction coil to the discharge tube containing the maltese cross (Crookes' tube). Observe the end of the tube containing the cross when the cross is down and when it is up.
3. Replace the Crookes' tube with the tube containing the electric plates and connect the terminals of the plate to its high DC voltage supply. Observe the effects of the electric field on the cathode rays.
4. Connect the tube with the fluorescent screen display to the induction coil and record the effect of placing a set of bar magnets around the cathode rays as shown in Figure 13.3.
5. Finally, attach the tube containing the glass wheel on tracks to the induction coil and observe the effects that the cathode rays have on the wheel when the tube is horizontal.

Analysis

1. For each of the tubes placed in the circuit, sketch a diagram of the tube and the effect caused by the cathode rays.
2. Using the laws of electromagnetism, determine the charge that is evident on the cathode rays.

Questions

1. What are five properties of cathode rays that can be deduced from this experiment?
2. From these results, can we conclusively say that the cathode rays are electrons? Why or why not?

TOPIC 14
The atom and quantum mechanics

14.1 Overview

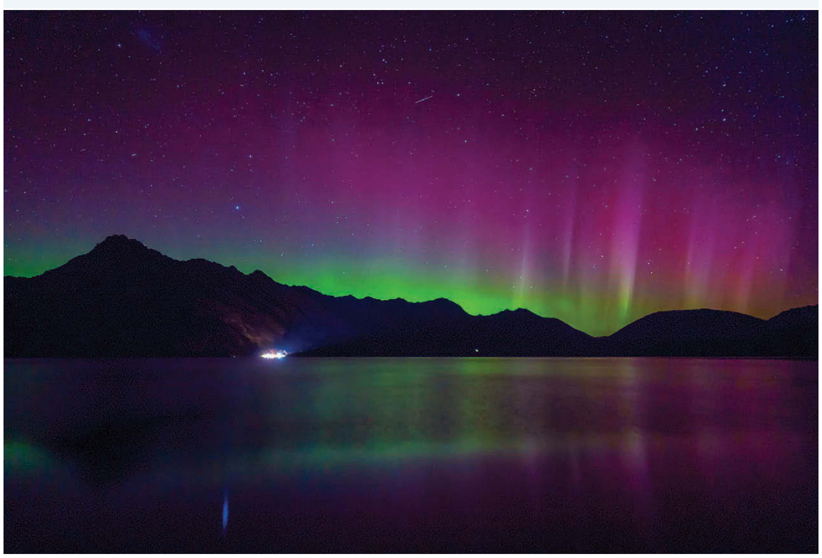
14.1.1 Module 8: From the universe to the atom

Quantum mechanical nature of the atom

Inquiry question: How is it known that classical physics cannot explain the properties of the atom?
Students:

- assess the limitations of the Rutherford and Bohr atomic models
- investigate the line emission spectra to examine the Balmer series in hydrogen (ACSPH138)
- relate qualitatively and quantitatively the quantised energy levels of the hydrogen atom and the law of conservation of energy to the line emission spectrum of hydrogen using:
  - $E = \mathrm{h}f$
  - $E = \frac{\mathrm{hc}}{\lambda}$
  - $\frac{1}{\lambda} = R\left[\frac{1}{n_f^2} - \frac{1}{n_i^2}\right]$ (ACSPH136)
- investigate de Broglie's matter waves, and the experimental evidence that developed the following formula:
  - $\lambda = \frac{\mathrm{h}}{mv}$ (ACSPH140)
- analyse the contribution of Schrödinger to the current model of the atom.

FIGURE 14.1 Photograph of 'aurora australis', the southern lights. In an aurora, atoms of the gases in the upper atmosphere emit radiation after being excited by interactions with charged particles from the Sun.



14.2 Limitations of the Rutherford atomic model

14.2.1 Rutherford unhinged

With all the success of Thomson and Rutherford in developing a picture and working model of the atom, several problems remained. In a planetary model of the atom, orbiting electrons that are accelerating should emit electromagnetic radiation continuously; however, this was never observed. Electrons should spiral into the nucleus as they emit this radiation. Following this model to its conclusion implied that all atoms are inherently unstable; again, this was not observed. Further, there was a second difficulty: excited atoms do emit electromagnetic radiation, but the spectrum is discrete, consisting of specific colours and unique for each element (much like the unique harmonic structure and timbre of a musical instrument), which are properties associated with waves. This puzzle concerning the stability of atoms was yet to be solved, but the emergence of quantum mechanics would unify all these confusing and often contradictory observations.

14.2.2 Bohr's model of the atom

Niels Bohr (1885–1962), a Danish physicist, was one of eleven Nobel prize winners who were trained by Rutherford. Before moving to Manchester to work with Rutherford, Bohr had worked for a short time at the Cavendish Laboratory under J.J. Thomson. Bohr and Thomson did not get along. At their first meeting, Bohr informed Thomson that one of Thomson's equations was wrong. There were several similar incidents and Bohr later recalled his disappointment that Thomson was not interested to learn that his work was incorrect. Bohr did acknowledge that his own lack of knowledge of the English language contributed to the failure of the two men to hit it off. One of Bohr's first contributions was to predict that a hydrogen atom would contain only one electron outside the positively charged nucleus. (At the same time, others predicted that one-electron atoms could not exist.)

Bohr attempted to apply the new quantum ideas of Planck and Einstein to the model of the hydrogen atom. Planck had managed to find an equation that solved the problem of the 'ultraviolet catastrophe' that troubled the theory of black body radiation. Planck held a traditionalist's view of physics and was opposed to the statistical processes of Boltzmann. After attempting to explain his black body equation, Planck reluctantly tried to derive it using the methods of Boltzmann. This involved dividing the energies up into small amounts and eventually should have finished with an integration in which all the energies would have been added together and would have experienced the problem of an infinite energy. However, before that final step, Planck realised that he had reached his equation for black-body radiation and therefore did not 'complete' the process. Einstein later showed that the problem of infinities will occur in any process where 'classical' theories and quantum theories are linked.

FIGURE 14.2 Niels Bohr.

Planck interpreted his result as meaning that the 'atomic oscillators' that produced the radiation could vibrate only with certain discrete amounts of energy. These discrete amounts of energy were called **quanta**.

Einstein later extended this idea to the radiation itself being quantised. Einstein's 'quanta of light' were later named 'photons' by Gilbert Lewis (1875–1946).

Bohr uses quantum theory to explain the spectrum of hydrogen

Bohr knew that, somehow, atoms must produce radiation that formed a characteristic spectrum for each element (see Section 14.2.4). Bohr realised that the 'atomic oscillators' of Planck were probably electrons in the atom. The Rutherford model failed to provide any information about the radius of the atom or the orbital frequencies of the electrons. Bohr attempted to introduce the quantum ideas of Planck to the atom, but at first failed.

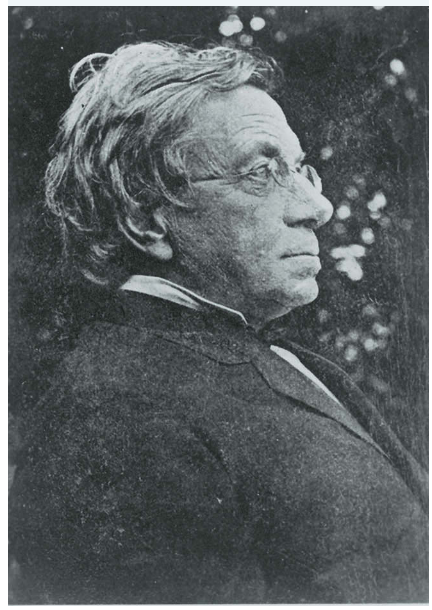
Early in 1913, Bohr was introduced to Balmer's equation (see Section 14.2.3) for the wavelengths of the spectral lines of hydrogen and it 'made everything clear to him'. After seeing this equation, Bohr realised how electrons were arranged in the hydrogen atom and also how quantum ideas could be introduced to the atom.

14.2.3 Balmer's equation for the hydrogen atom

Johann Jakob Balmer (1825–1898) completed a PhD in mathematics in 1849. He became a teacher at a girls' school in Basel, Switzerland, and had a desire to 'grasp the harmonic relationships of nature and art numerically'. Anders Angström (1814–1874) had measured the wavelengths of four of the spectral lines of hydrogen (now known as the Balmer series). Balmer found an equation that enabled him to calculate the wavelengths of these and, he believed, the infinite number of spectral lines emitted by hydrogen.

Balmer's equation was $\lambda = b\left(\frac{n^2}{n^2 - 2^2}\right)$ and the constant b was found empirically to be 364.56 nm.

FIGURE 14.3 Johann Jakob Balmer.



Janne Rydberg (1854–1919) modified Balmer's equation for wavelength to produce the familiar equation:

$$\frac{1}{\lambda} = \mathrm{R_H}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

where

λ = wavelength of the emitted radiation

$\mathrm{R_H}$ = Rydberg's constant $\left(\mathrm{R_H} = 1.097 \times 10^7\ \mathrm{m^{-1}}\right)$

n_f and n_i are integers.

The wavelengths of the visible lines of hydrogen correspond to $n_f = 2$ and $n_i = 3, 4, 5$ or 6. Of course, this is an **empirical equation** (Balmer played around with numbers until he arrived at something that worked).

Sometimes the equation $\frac{1}{\lambda} = \mathrm{R_H}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$ is known as the Rydberg equation. Sometimes it is called Balmer's equation. Rydberg had attempted to find his own equation for the spectral lines of hydrogen. He was unsuccessful and, as his contribution was to modify Balmer's equation, we will continue to refer to it as the Balmer equation.

14.2 SAMPLE PROBLEM 1

CALCULATING THE WAVELENGTHS OF HYDROGEN SPECTRAL LINES

Calculate the wavelength of the visible spectral line of hydrogen with the longest wavelength.

SOLUTION:

From Balmer's equation $\frac{1}{\lambda} = \mathrm{R_H}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$, we can see that the longest wavelength will occur when the term $\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$ is smallest.

As the visible spectral lines correspond to $n_f = 2$, the smallest value will be when $n_i = 3$.

$$\frac{1}{\lambda} = R_H\left(\frac{1}{2^2} - \frac{1}{3^2}\right)$$
$$= 1.097 \times 10^7\left(\frac{1}{2^2} - \frac{1}{3^2}\right)$$
$$= 1.524 \times 10^6$$
$$\lambda = 6.562 \times 10^{-7}\ \text{m}$$

The wavelength is 6.562×10^{-7} m. This is the wavelength of the red line in the hydrogen spectrum in Figure 14.4.

14.2 Exercise 1

1 A hydrogen atom makes a transition from the third excited state ($n = 4$) to the ground state. Use Balmer's equation to determine the wavelength and, hence, the energy of the photon emitted.

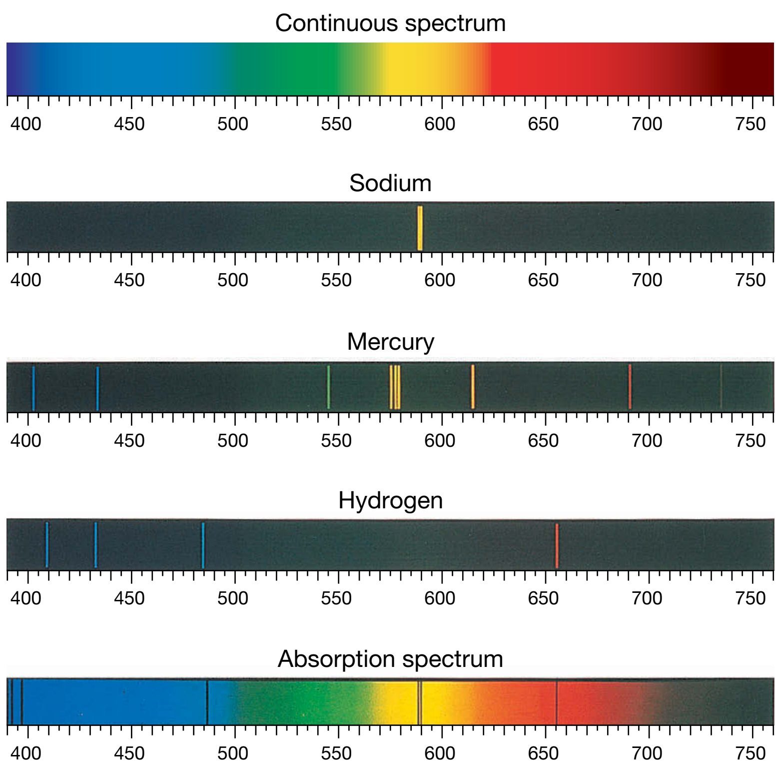
2 Is this photon in the visible part of the electromagnetic spectrum? If not, classify the photon.

14.2.4 The spectra of gases

There are three types of emission spectra: continuous spectra, bright-line spectra and band spectra. Continuous spectra are produced by incandescent objects, bright-line spectra are produced by excited gases, and band spectra are produced by excited molecules. We will consider the bright-line emission spectra of excited gases and also the absorption spectra of cool gases (as shown in Figure 14.4).

Spectral lines are produced as images of the slit that is an essential component of any spectroscope. After passing through the slit, the different wavelengths of light are diffracted by different amounts by a grating or dispersed by a prism by different amounts. Hence, the images of the slit corresponding to the different wavelengths are separated. When the slit is very narrow, closely spaced lines can be resolved (distinguished from one another). If the slit is wider, more light is admitted, but at the expense of the resolution.

FIGURE 14.4 A continuous white light spectrum, the emission spectra of excited atoms of the elements sodium, mercury and hydrogen, and an absorption spectrum. The red line in the hydrogen spectrum is known as the H_α line, and the other lines as H_β, H_γ and H_δ respectively.



An **emission spectrum** (see Figure 14.5) is produced when a gas is excited. A gas can be excited by heating it or by passing an electrical discharge through it. The emission spectrum is a series of narrow coloured lines on a dark background. Each element has its own characteristic spectrum and this can be used to identify the gas.

An **absorption spectrum** (see Figure 14.6) is produced when white light is passed through a cool gas. The atoms in the gas absorb energy from the white light. The atoms will then re-emit the energy that was absorbed. The energy will be emitted as light and it will be emitted in random directions. Therefore, the transmitted beam of light will be deficient in light at those energies or wavelengths. When this light is analysed, it will show a continuous spectrum of the white light with a series of narrow dark lines across it.

FIGURE 14.5 When the light emitted from an excited gas is passed through a prism, a series of coloured lines is observed. The lines correspond to the colours (wavelengths) of light emitted by the gas.

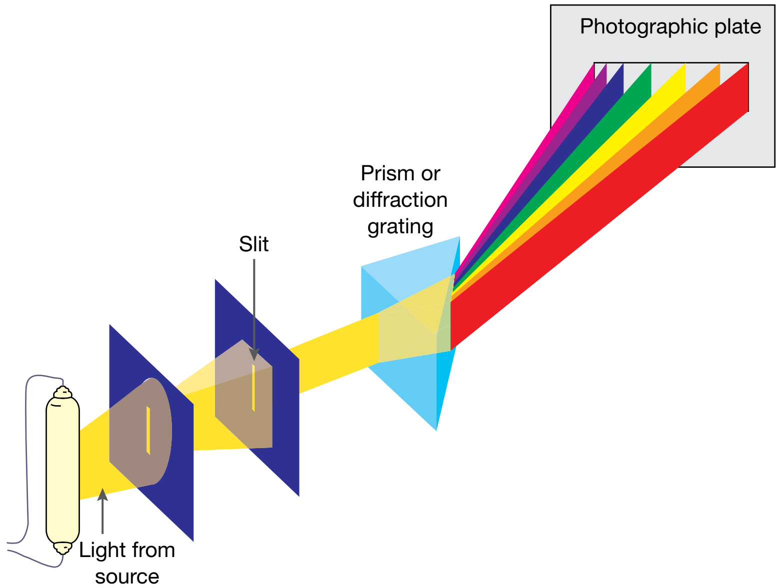
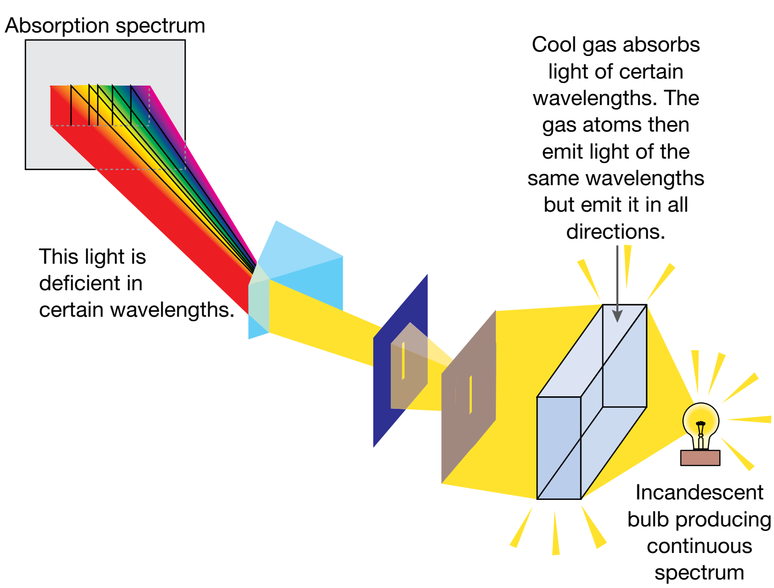


FIGURE 14.6 When white light is passed through a cool gas, the gas absorbs radiation from the light. After the light has been passed through a prism, the colours corresponding to the wavelengths of light absorbed by the gas are absent from the spectrum. An absorption spectrum of dark lines on a bright coloured background is observed.



14.2.5 Bohr's postulates

While Bohr believed that he knew the arrangement of electrons, he could not explain why the electrons were arranged in this way. Bohr published three papers between April and August 1913. In these papers, known as the great trilogy, he started with the problem of electrons in the Rutherford model and pointed out that the accelerating electrons must lose energy by radiation and collapse into the nucleus. He then applied quantum theory to the atom. He generally assumed that the orbits of the electron were circular.

Bohr was awarded the Nobel Prize for Physics in 1922 and in his Nobel lecture stated in reference to Rutherford's discovery of the nucleus:

'This discovery made it quite clear that by classical conceptions alone it was quite impossible to understand the most essential properties of atoms. One was therefore led to seek for a formulation of the principles of the quantum theory that could immediately account for the stability in atomic structure and the radiation sent out from atoms, of which the observed properties bear witness. Such a formulation I proposed [1913] in the form of two postulates.'

Bohr continued with rather lengthy statements of his postulates. Simpler statements are:

1. Electrons in an atom exist in '**stationary states**' in which they possess an unexplainable stability. Any permanent change in their motion must consist of a complete transition from one stationary state to another. When an electron is in a stationary state, it will orbit the nucleus without emitting any electromagnetic radiation.
2. In contradiction to the classical electromagnetic theory, no radiation is emitted from an atom in a stationary state. A transition between two stationary states will be accompanied by emission or absorption of electromagnetic radiation (a photon). The frequency, f, of this photon is given by the relation:

$$\mathrm{h}f = E_1 - E_2$$

where
h = Planck's constant
E_1 and E_2 = values of the energy of two stationary states that form the initial and final states of the atom.

Bohr then introduced what is generally known as his quantisation condition and is sometimes called his third postulate.

An electron in a stationary state has an **angular momentum** that is an integral multiple of $\frac{\mathrm{h}}{2\pi}$ (Planck's constant divided by 2π).

Bohr actually proposed that the kinetic energy of an electron was $\frac{n}{2\mathrm{h}f}$ but this reduces to the quantisation condition given if the orbits are circular.

In his first postulate, Bohr put forward one of the most audacious hypotheses ever proposed in physics by predicting that electrons exist in states in which they do not radiate energy. The second postulate involves the quantum of energy being emitted or absorbed when an electron jumps from one stationary state to another and hence explains the origin of spectral lines. The quantisation condition is really an intuitive guess.

Using these postulates together with the energy of electrons calculated from 'classical' physics applied to the Rutherford model, it is possible to derive a theoretical equation for the wavelengths of the spectral lines of hydrogen. It is a great success of the Bohr model that this theoretical equation is the same as the empirical equation of Balmer. Many famous physicists had addressed the problem of electrons being in non-uniform motion without radiating energy. This had become important after Thomson had discovered the electron and was not just associated with the Rutherford model.

14.2.6 Mathematics of the Rutherford and Bohr models

In the following sections we will derive an expression for the classical energy of the Rutherford hydrogen atom and then impose Bohr's postulates on that atom. This will enable us to calculate the energies of the stationary states of the hydrogen atom and then calculate the change in energy of an electron involved in a transition between two stationary states. Finally, this change in energy will enable us to calculate the frequency (or wavelength) of the spectral lines of hydrogen.

The 'classical' energy of the Rutherford hydrogen atom

When you studied the escape velocity of an object fired from the Earth, you found the total energy of the object was the sum of its kinetic energy and its gravitational potential energy. When this total energy was zero, the object had just enough energy to escape from the Earth. If the total energy was negative, the object was unable to escape the Earth.

In a similar way we can calculate the total energy of a proton and electron. This time it is the sum of the kinetic energy and the electrical potential energy. The zero point will be when the electron has just enough energy to escape from the proton.

Kinetic energy of electron:

$$E_{\mathrm{k}} = \frac{1}{2}m_{\mathrm{e}}v^2.$$

The electron is held in orbit around the proton by the electrical force of magnitude:

$$F = \frac{\mathrm{k}q_{\mathrm{e}}^2}{r^2}$$

where

q_e = magnitude of the charge on the proton and electron (1.602×10^{-19} C).

We know that this electrical force provides the centripetal force of magnitude:

$$\begin{aligned}
F_{\mathrm{c}} &= \frac{m_{\mathrm{e}}v^2}{r} \\
F_{\mathrm{c}} &= F_{\mathrm{E}} \\
\frac{m_{\mathrm{e}}v^2}{r} &= \frac{\mathrm{k}q_{\mathrm{e}}^2}{r^2} \\
\frac{1}{2}\frac{m_{\mathrm{e}}v^2}{r} &= \frac{1}{2}\frac{\mathrm{k}q_{\mathrm{e}}^2}{r^2} \\
\frac{1}{2}m_{\mathrm{e}}v^2 &= \frac{1}{2}\frac{\mathrm{k}q_{\mathrm{e}}^2}{r} \\
E_{\mathrm{k}} &= \frac{1}{2}\frac{\mathrm{k}q_{\mathrm{e}}^2}{r}.
\end{aligned}$$

The potential energy of the electron is given by:

$$E_{\mathrm{p}} = -\frac{\mathrm{k}q_{\mathrm{e}}^2}{r}.$$

The total energy is the sum of the kinetic and potential energies.

$$\begin{aligned}\text{Total energy} &= \mathrm{E_k} + \mathrm{E_p} \\ &= \frac{1}{2}\frac{\mathrm{k}q_\mathrm{e}^2}{r} - \frac{\mathrm{k}q_\mathrm{e}^2}{r} \\ &= -\frac{1}{2}\frac{\mathrm{k}q_\mathrm{e}^2}{r}\end{aligned}$$

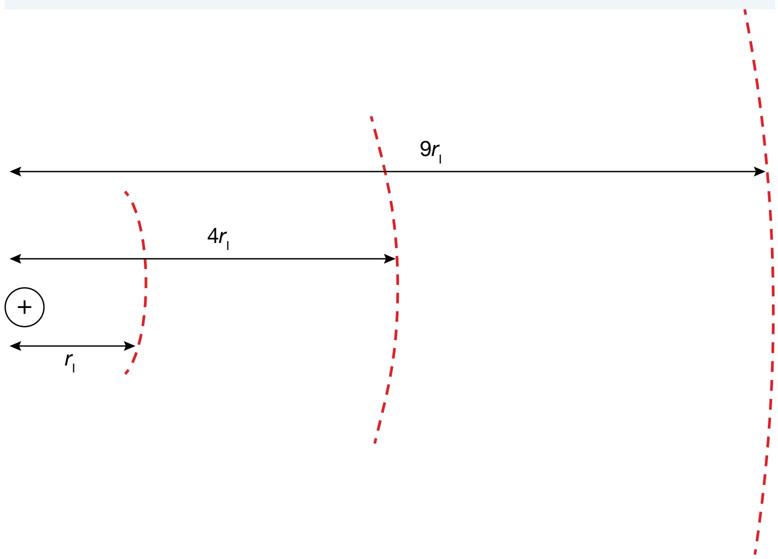
This is the total 'classical' energy of Rutherford's hydrogen atom.

Radii of the 'stationary states' of the Bohr hydrogen atom

When Bohr's quantisation condition is applied to the 'classical' hydrogen atom, the electron is restricted to stationary states in which the angular momentum of the electron is an integer multiple of Planck's constant, divided by 2π.

$$\begin{aligned}\text{Angular momentum} &= \frac{n\mathrm{h}}{2\pi} \\ m_\mathrm{e}vr &= \frac{n\mathrm{h}}{2\pi}\end{aligned}$$

FIGURE 14.7 The relative radii of the orbits of an electron in different stationary states in a hydrogen atom.



In this equation, n is an integer, known as the **principal quantum number**. We can obtain an expression for the radius of the stationary states corresponding to each value of the integer, n:

$$\begin{aligned}m_\mathrm{e}vr &= \frac{n\mathrm{h}}{2\pi} \\ r &= \frac{n\mathrm{h}}{2\pi m_\mathrm{e}v} \\ r^2 &= \frac{n^2\mathrm{h}^2}{4\pi^2 m_\mathrm{e}^2 v^2}.\end{aligned}$$

From the earlier equation, $\frac{m_\mathrm{e}v^2}{r} = \frac{\mathrm{k}q_\mathrm{e}^2}{r^2}$, we can obtain an expression for v^2:

$$v^2 = \frac{\mathrm{k}q_\mathrm{e}^2}{m_\mathrm{e}r}.$$

Substituting this gives:

$$\begin{aligned}r^2 &= \frac{n^2\mathrm{h}^2}{4\pi^2 m_\mathrm{e}^2 \frac{\mathrm{k}q_\mathrm{e}^2}{m_\mathrm{e}r}} \\ r_n &= \frac{n^2\mathrm{h}^2}{4\pi^2 m_\mathrm{e}\mathrm{k}q_\mathrm{e}^2}\end{aligned}$$

where

r_n = the radius of the stationary state corresponding to the integer n.

The radius of the stationary state corresponding to $n = 1$ will be:

$$r_1 = \frac{1^2 h^2}{4\pi^2 m_e k q_e^2} = \frac{h^2}{4\pi^2 m_e k q_e^2}.$$

We can combine the expressions for r_n and r_1 to give $r_n = n^2 r_1$.

Energies of the 'stationary states' of the Bohr atom

If we now return to the classical energy of the Rutherford hydrogen atom $\left(\text{energy} = -\frac{1}{2}\frac{kq_e^2}{r}\right)$ and impose the restriction that the only possible correspond to values of radius given by $r_n = \frac{n^2 h^2}{4\pi^2 m_e k q_e^2}$, we can calculate the value of these energy states:

$$E_n = -\frac{1}{2}\frac{kq_e^2}{\frac{n^2 h^2}{4\pi^2 m_e k q_e^2}} = -\frac{1}{2}\frac{4\pi^2 k^2 m_e q_e^4}{n^2 h^2} = -\frac{1}{n^2}\left(\frac{2\pi^2 k^2 m_e q_e^4}{h^2}\right).$$

Again we can see that:

$$E_1 = -\left(\frac{2\pi^2 k^2 m_e q_e^4}{h^2}\right)$$

and hence

$$E_n = \frac{1}{n^2}E_1$$

remembering that E_1 has a negative value.

14.2 SAMPLE PROBLEM 2

CALCULATING THE ENERGIES OF ELECTRONS IN THE HYDROGEN ATOM

Given that the energy of an electron in the first stationary state of hydrogen is $E_1 = -2.179 \times 10^{-18}$ J, determine the energy in electron volts (eV) of an electron in the following stationary states of the hydrogen atom:

(a) the first stationary state ($n = 1$)
(b) the second stationary state ($n = 2$)
(c) the tenth stationary state ($n = 10$).

SOLUTION:

(a) We have been given this energy in joules so it is only a matter of converting to electron volts:

$$\begin{aligned} 1\,\text{eV} &= 1.602 \times 10^{-19}\,\text{J} \\ 2.179 \times 10^{-18}\,\text{J} &= \frac{2.179 \times 10^{-18}}{1.602 \times 10^{-19}}\,\text{eV} \\ &= 13.60\,\text{eV}. \end{aligned}$$

The energy of the first stationary state is −13.6 eV.

(b) The energy of an electron in the second stationary state, for which $n = 2$ is given by:

$$\begin{aligned} E_n &= \frac{1}{n^2}E_1 \\ E_2 &= \frac{1}{2^2}E_1 \\ &= \frac{-13.6}{4} \\ &= -3.4\,\text{eV}. \end{aligned}$$

(c) The energy of an electron in the tenth stationary state, for which $n = 10$ is given by:

$$\begin{aligned} E_n &= \frac{1}{n^2}E_1 \\ E_{10} &= \frac{1}{10^2}E_1 \\ &= \frac{-13.6}{100} \\ &= -1.36 \times 10^{-1}\,\text{eV}. \end{aligned}$$

Theoretical expression for wavelengths of the spectral lines of hydrogen

We are able to combine the expression for the energies of the stationary states with Bohr's second postulate to derive an expression for the energy differences between stationary states and, hence, the energies of the photons that may be emitted or absorbed by hydrogen.

We will consider the emission of a photon as an electron jumps from a higher energy initial state, E_i, to a lower energy final state, E_f.

The change in energy of the electron is:

$$\begin{aligned} \Delta E &= E_i - E_f \\ &= \frac{1}{n_i^2}E_1 - \frac{1}{n_f^2}E_1 \\ &= E_1\left(\frac{1}{n_i^2} - \frac{1}{n_f^2}\right). \end{aligned}$$

This is the energy of the emitted photon, hf.

We can now derive an expression for the frequency and wavelength of the photon.

$$hf = E_1\left(\frac{1}{n_i^2} - \frac{1}{n_f^2}\right)$$

$$f = \frac{-E_1}{h}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

$$\frac{c}{\lambda} = \frac{-E_1}{h}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

$$\frac{1}{\lambda} = \frac{-E_1}{hc}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

This equation is of the same form as Balmer's equation. If the value of $\frac{-E_1}{hc}$ is calculated, it agrees with the value of the Rydberg constant in Balmer's equation. (Remember that E_1 is a negative quantity and, hence, $-E_1$ is positive.)

Balmer's equation is an empirical equation. A theoretical equation derived from Bohr's model of the atom now agrees with the empirical equation. This is a major achievement and offers very strong support for the Bohr model.

14.2 SAMPLE PROBLEM 3

EMISSION OF PHOTONS FROM A HYDROGEN ATOM

(a) Given that the energy of the first stationary state of hydrogen is −13.60 eV, calculate the energy of the fourth stationary state of the hydrogen atom.
(b) Use this information to calculate the frequency of the photon emitted when an electron undergoes a transition from the state $n = 4$ to the state $n = 1$.
(c) Calculate the wavelength of the radiation emitted.

SOLUTION:

(a) The energy of the fourth stationary state is:

$$E_n = \frac{1}{n^2}E_1$$

$$E_4 = \frac{1}{4^2}E_1$$

$$= \frac{-13.6}{16}$$

$$= -0.85\ \text{eV}.$$

(b) The energy emitted by the photon will be:

13.60 eV − 0.85 eV = 12.75 eV.

$$\text{Energy of photon} = 12.75 \times 1.602 \times 10^{-19}\ \text{J}$$

$$f = \frac{E}{h}$$

$$= \frac{12.75 \times 1.602 \times 10^{-19}}{6.626 \times 10^{-34}}$$

$$= 3.083 \times 10^{15}\ \text{Hz}$$

(c) The wavelength will be calculated from:

$$\lambda = \frac{c}{f}$$
$$= \frac{3.00 \times 10^8}{3.083 \times 10^{15}}$$
$$= 9.73 \times 10^{-8} \text{ m.}$$

(Of course the wavelength could have been calculated directly from Balmer's equation.)

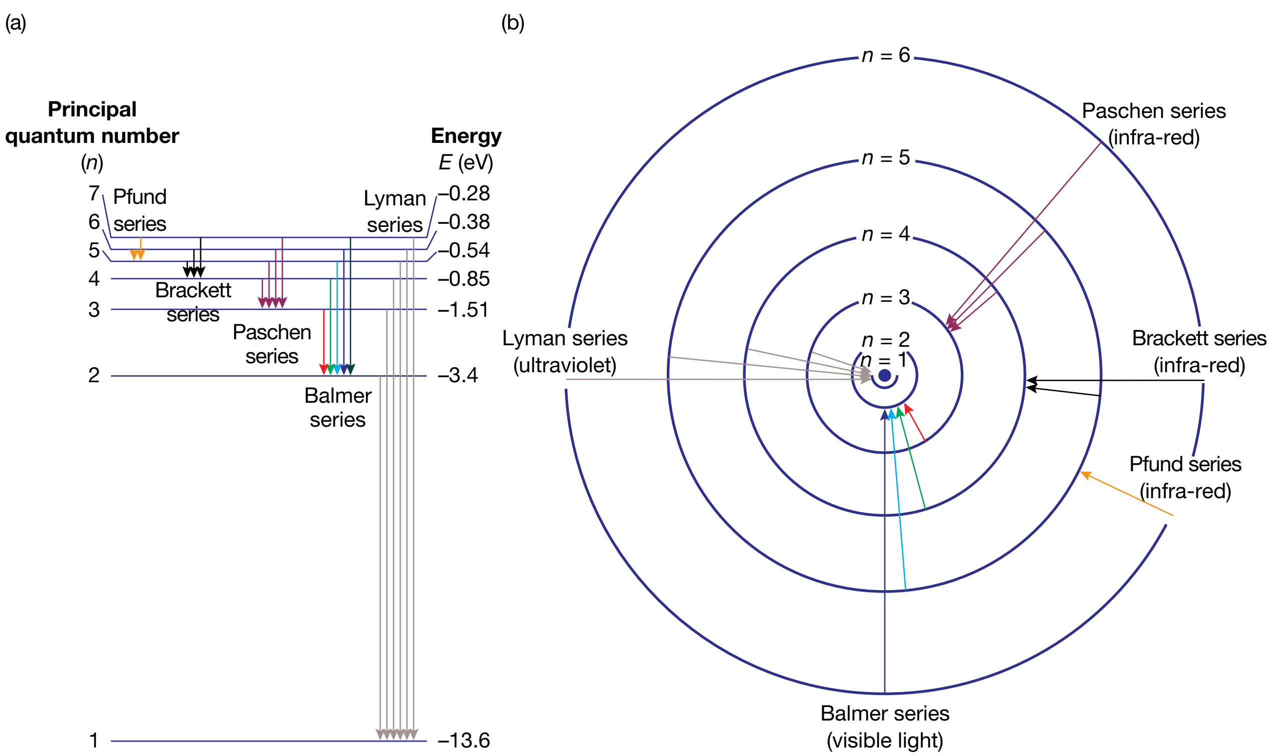
14.2.7 The hydrogen atom explained but with limitations

We are now able to calculate the wavelengths of the many spectral lines of the hydrogen atom. The original series of spectral lines was known as the Balmer series and contained the four spectral lines in the visible region of the spectrum. These lines correspond to electron jumps to the second lowest energy state, or first excited state, ($n = 2$) of the hydrogen atom.

The wavelengths of the spectral lines in other series can be calculated using Bohr's equation and are shown in Figure 14.8. The Paschen series of infra-red lines had already been discovered but other series of lines were found later and their wavelengths were in agreement with Bohr's theory. The series of lines in the ultraviolet and infra-red, named after their discoverers, are:

- Lyman series, discovered in 1916. These were ultraviolet lines with transitions to the **ground state** ($n = 1$). An electron has the lowest possible amount of energy when it is in the ground state.
- Paschen series, discovered in 1908. These were infra-red lines with transitions to the second **excited state** ($n = 3$). If it exists in a stationary state in which it has more energy, it is said to be in an excited state.

FIGURE 14.8 (a) Atomic energy level view of the spectral series of hydrogen (b) Electron orbit view of the spectral series of hydrogen. Note that the radii of the orbits of the electrons are not to scale.



- Brackett series, discovered in 1922. These were infra-red lines with transitions to the third excited state ($n = 4$).
- Pfund series, discovered in 1924. These were infra-red lines with transitions to the fourth excited state ($n = 5$).

Limitations of the Bohr model of the atom

The Bohr model takes the first step to introduce quantum theory to the hydrogen atom but it is only a first step. The model has the following limitations:

- It is not possible to calculate the wavelengths of the spectral lines of all other atoms.
- The Bohr model works reasonably well for atoms with one electron in their outer shell but does not work for any of the others.
- Examination of spectra shows that the spectral lines are not of equal intensity but the Bohr model does not explain why some electron transitions would be favoured over others.
- Careful observations with better instruments showed that there were other lines known as the hyperfine lines. There must be some splitting of the energy levels of the Bohr atom but the Bohr model cannot account for this.
- When a gas is excited while in a magnetic field, the emission spectrum produced shows a splitting of the spectral lines (called the Zeeman effect). Again, the Bohr model cannot account for this.
- Finally, the Bohr model is a mixture of classical physics and quantum physics and this, in itself, is a problem.

14.2 Exercise 2

1 Use the equation $E_n = \frac{1}{n^2}E_1$ where $E_1 = -13.6$ eV to calculate the energy of a photon emitted when a hydrogen atom undergoes a transition from the $n = 5$ state (4th excited state) to the $n = 2$ state (1st excited state). State your answer in both eV and in joules.

eBook*plus* RESOURCES

Try out this interactivity: Hydrogen emission spectra
Searchlight ID: int-0755

14.3 Matter waves

14.3.1 The wave behaviour of electrons

By the end of the nineteenth century, it was clear that light exhibited wave properties and could be very well modelled as consisting of waves. It was also firmly established, at that time, that matter could be modelled as consisting of particles. Early in the twentieth century, however, it was found that because of the photoelectric effect it was necessary for light to also be modelled as a particle. Was it possible that electrons, too, could exhibit wave phenomena as well as demonstrating particle behaviour?

Even though Bohr could calculate their energies, he could not explain why hydrogen electrons occupied only orbits whose energies were discrete. Why were they the only possible electron orbits? What was so special about them?

How did atoms make sure they emitted the right frequency to ensure they landed in another stationary state?

In fact Rutherford wrote to Bohr:

> 'Your ideas are very ingenious and seem to work out well ... There seems to me to be one grave difficulty in your hypothesis ... namely, how does an electron decide what frequency it is going to vibrate at when it passes from one stationary state into another? It seems to me that you would have to assume that the electron knows beforehand where it is going to stop.'

In 1923 French nobleman Louis de Broglie (1892–1987) suggested that matter also had a wavelength associated with it. He was intrigued by the fact that light exhibited both wavelike and particle-like properties, and on this basis proposed that matter may also exhibit wavelike properties. This work was done as his PhD thesis. De Broglie proposed that the wavelength of a particle, λ, is related to its momentum, p, according to the following equation:

$$\lambda = \frac{\mathrm{h}}{p} = \frac{\mathrm{h}}{mv}.$$

The constant h, Planck's constant, is related to the particle-like behaviour of light and has a value of 6.63×10^{-34} J s, or 4.15×10^{-15} eV s. The momentum of matter is given by the product of its mass and velocity.

We can appreciate why the wave properties of matter are difficult to observe. Let's calculate the **de Broglie wavelength** of a 70 kg athlete running at a speed of 10 m s^{-1}. The de Broglie wavelength is the wavelength associated with a particle or discrete piece of matter. Using the formula $\lambda = \frac{\mathrm{h}}{mv}$:

$$\begin{aligned}\lambda &= \frac{6.63 \times 10^{-34}\ \mathrm{J\ s}}{70\ \mathrm{kg} \times 10\ \mathrm{m\ s^{-1}}}\\ &= 9.5 \times 10^{-37}\ \mathrm{m}.\end{aligned}$$

This wavelength is much too small to allow for the ready observation of diffraction effects as an athlete runs through a narrow opening! However, for a particle with a small mass, such as an electron travelling at low speed, this is not the case. Electrons accelerated through a 100 V potential difference would have a speed of approximately 6.0×10^{6} m s^{-1}, and because the mass of an electron is 9.1×10^{-31} kg it would have a momentum of:

$$\begin{aligned}p &= mv\\ &= 9.1 \times 10^{-31}\ \mathrm{kg} \times 6.0 \times 10^{6}\ \mathrm{m\ s^{-1}}\\ &= 5.5 \times 10^{-24}\ \mathrm{kg\ m\ s^{-1}}\end{aligned}$$

The de Broglie wavelength for these electrons is:

$$\begin{aligned}\lambda &= \frac{6.63 \times 10^{-34}\ \mathrm{J\ s}}{5.5 \times 10^{-24}\ \mathrm{m\ s^{-1}}}\\ &= 1.2 \times 10^{-10}\ \mathrm{m}.\end{aligned}$$

This wavelength has the same order of magnitude as the spacing between atoms in many crystals. When the ratio of wavelength λ to slit width w, $\frac{\lambda}{w}$, is sufficiently large, say greater than $\frac{1}{10}$ for example, diffraction effects are readily observable.

The framework for testing to see if matter had an associated wavelength had now been constructed. Researchers could build an apparatus to fire a beam of electrons of specific energy and hence specific momentum and wavelength at a crystal and see if any diffraction effects appeared.

14.3 SAMPLE PROBLEM 1

(a) Calculate the de Broglie wavelength of a 10 g snail whose speed is 0.10 mm s^{-1}.
(b) How fast would an electron have to travel to have a de Broglie wavelength of 1 μm?

SOLUTION:

(a) The de Broglie wavelength is given by the expression:

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

Thus

$$\lambda = \frac{6.63 \times 10^{-34}}{10 \times 10^{-3} \times 0.10 \times 10^{-3}}$$
$$= 6.63 \times 10^{-33}\ \text{m},$$

keeping in mind that mass must be in kilograms and velocity in metres per second.

(b) The expression $\lambda = \frac{h}{mv}$ can be transposed to make v the subject. Thus $v = \frac{h}{m\lambda}$.

$$v = \frac{6.63 \times 10^{-34}}{9.1 \times 10^{-31} \times 1 \times 10^{-6}}$$
$$= 728.571$$
$$= 7.3 \times 10^{2}\ \text{m s}^{-1}\ \text{(to 2 significant figures)}$$

The speed of the electron is 7.3×10^{2} m s^{-1}.

14.3 Exercise 1

1 Which has the greater de Broglie wavelength: a proton ($m = 1.67 \times 10^{-27}$ kg) travelling at 2.0×10^{4} m s^{-1} or an electron ($m = 9.1 \times 10^{-31}$ kg) travelling at 2.0×10^{5} m s^{-1}?

2 If an electron is made to travel at a higher speed, what effect does this have on its de Broglie wavelength?

3 A research scientist is trying to produce a beam of electrons with a de Broglie wavelength of 1.0×10^{-11} m ($m = 9.1 \times 10^{-31}$ kg) using an electron gun to study a new molecule that is being proposed in a novel medical application. She would like to know the speed of an electron in this beam. Calculate the speed of an electron in this proposed beam.
Note: Diffraction patterns produced from this type of experiment are able to be used to determine the shape of such molecules. This is how we know the shape of the DNA molecule — by diffraction pattern analysis.

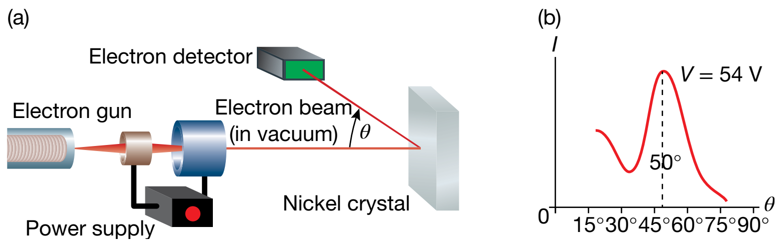
Finally, it is worth noting that the de Broglie wavelength associated with a piece of matter is inversely proportional to both the speed and mass. Hence, to create matter with large wavelengths, necessary for wave properties to manifest themselves, matter has to travel slowly and have little mass. Since electrons have a mass that is approximately $\frac{1}{1800}$ that of a proton or neutron, it is easier to detect the wave properties of electrons over those of other fundamental particles such as protons and neutrons.

14.3.2 Matter waves show themselves

De Broglie suggested conducting an experiment to confirm whether or not a beam of electrons could be diffracted from the surface of a crystal. The openings between atoms could be used as a diffraction grating in much the same way that X-rays were diffracted by thin crystals as suggested by Max von Laue in 1912. Clinton Davisson (1881–1958) and Lester Germer (1896–1971) directed a beam of electrons at a metal crystal in 1927, and the scattered electrons came off in regular peaks as shown in Figure 14.9.

This pattern is indicative of diffraction taking place with individual electrons as they scattered off the crystal surface. In fact, the wavelength determined from the diffraction experiments was exactly as predicted by the de Broglie wavelength formula. In this way, electrons were shown to have wavelike properties. Since then, protons, neutrons and, more recently, atoms have been shown to exhibit wavelike properties, but it begs the question: if matter can exhibit wave characteristics, what is it that is 'waving'? More technically, the question is what physical variable is it that has an amplitude and phase?

FIGURE 14.9 The Davisson and Germer experiment. (a) Electrons emitted from a heated filament are accelerated towards the crystal surface. The intensity of reflected electrons is recorded as the angle of the detector is changed. (b) Electron intensity as a function of angle.



14.3 SAMPLE PROBLEM 2

What would be the dimensions of the array of slits required to observe diffraction of 60 g tennis balls travelling at 30 m s^{-1}? What about electrons travelling at 3.0×10^6 m s^{-1}?

SOLUTION:

To observe diffraction effects, the size of the opening needs to be of the same order of magnitude or smaller than the wavelength of the waves. We can see below that the de Broglie wavelength of the tennis ball is of the order of 10^{-34} m and the electron of the order of 10^{-10} m.

The de Broglie wavelength of:

the tennis ball

$$\lambda = \frac{6.6262 \times 10^{-34}\ \text{J s}}{0.060\ \text{kg} \times 30\ \text{m s}^{-1}} = 3.7 \times 10^{-34}\ \text{m}$$

the electron

$$\lambda = \frac{6.6262 \times 10^{-34}\ \text{J s}}{9.109 \times 10^{-31}\ \text{kg} \times 3.0 \times 10^{6}\ \text{m s}^{-1}} = 2.4 \times 10^{-10}\ \text{m}$$

The distances between atoms in a crystal are of the order of 10^{-10} m, so we could observe diffraction and interference when these electrons are scattered from a crystal. It is not surprising that we never observe diffraction and interference effects with tennis balls, due to the extremely small wavelength, 10^{-34} m, that they have.

14.3 SAMPLE PROBLEM 3

What voltage is required to accelerate electrons to a speed of 3.0×10^6 m s^{-1}?

SOLUTION:

To accelerate electrons to a speed of 3.0×10^6 m s^{-1}, we need to calculate the work done by a voltage V.

$$\Delta E_{\text{k electron}} = \frac{1}{2} m_e v^2$$

$$\Delta E_{\text{k electron}} = -\Delta E_{\text{p electron}}$$

$$= q_e V$$

where
q_e is the *magnitude* of the charge of the electron.

$$
\begin{aligned}
\Rightarrow V &= \frac{m_e v^2}{2q_e} \\
&= \frac{9.109 \times 10^{-31}\ \mathrm{kg} \times (3.0 \times 10^{6}\ \mathrm{m\ s^{-1}})^2}{2 \times 1.6 \times 10^{-19}\ \mathrm{C}} \\
&= +26\ \mathrm{V}
\end{aligned}
$$

So, only 26 V is required to accelerate an electron to $3 \times 10^{6}\ \mathrm{m\ s^{-1}}$.

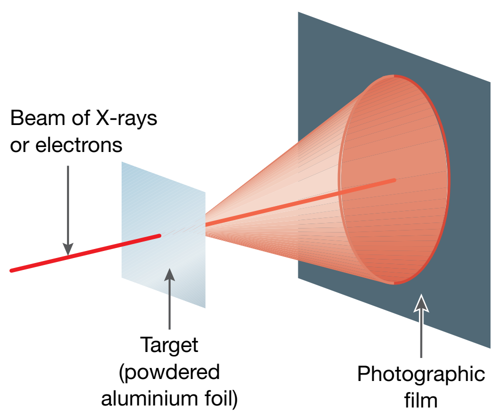
14.3.3 Electrons through foils

Intense, creative interest in fundamental physics ran in the Thomson family. Remember, it was J.J. Thomson whose ingenious experiment yielded the measurement of the charge-to-mass ratio of the electron. At that time there was no doubt that electrons were extremely well modelled as particles. However, G.P. Thomson, son of J.J., continued the exploration of the wave properties of electrons. He fired electrons through a thin polycrystalline metallic foil (see Figure 14.10). The electrons had a much greater momentum than those used by Davisson and Germer. They were able to penetrate the foil and produce a pattern demonstrating diffraction of the electrons by the atoms of the foil — further evidence for wavelike behaviour of electrons. The polycrystalline nature of the foil results in a series of rings of high intensity. A single crystal would produce a pattern of spots. Thomson used identical analysis techniques to those used for diffraction of X-rays through foils to confirm the de Broglie relationship.

Both Thomsons were awarded Nobel prizes — J.J. in 1897 for measuring a particle-like characteristic of electrons, and G.P. in 1937, together with C.J. Davisson, for demonstrating their wave properties.

Just as light requires a wave model and a particle model to interpret and explain how it behaves, so too does matter: it behaves like a particle in the sense that work can be done on it to increase its kinetic energy under the action of forces, but matter can also be made to diffract through sufficiently narrow openings and around obstacles. This requires a wave model and the de Broglie wavelength is used to determine the extent of matter's wave behaviour. It appears we need both a particle and a wave model for both light and matter. Electrons passed through a voltage V acquire a kinetic energy E_k equal to qV. Since they have kinetic energy, they also possess momentum and, according to de Broglie, a wavelength. We can determine a useful relationship between the de Broglie wavelength of an electron (λ) and the accelerating voltage (V) used.

FIGURE 14.10 Diffraction of X-rays and electrons by polycrystalline foils.



By equating the kinetic energy of the electron (E_k) to the work done by an accelerating voltage acting on an electron ($q_e V$), we get:

$$E_k = \frac{1}{2} m_e v^2 = q_e V$$

$$m_e v^2 = 2q_e V$$

$$m_e^2 v^2 = 2m_e q_e V.$$

The left-hand side is just the square of the momentum of the electron, and hence by taking the square root of both sides:

$$p = \sqrt{2m_e q_e V} \qquad \text{or} \qquad p = \sqrt{2m_e E_k},$$

remembering that E_k is equal to $q_e V$.

Since the de Broglie wavelength λ is given by $\frac{h}{p}$, it follows that:

$$\lambda = \frac{h}{\sqrt{2m_e q_e V}}$$

for a given accelerating voltage V, or

$$\lambda = \frac{h}{\sqrt{2m_e E_k}}$$

when the kinetic energy E_k of the electron in joules is known.

14.3 SAMPLE PROBLEM 4

Some of the X-rays used in G.P. Thomson's experiment had a wavelength of 7.1×10^{-11} m. Confirm that the 600 eV electrons have a similar wavelength.

SOLUTION:

Electrons of energy 600 eV have passed through a voltage equal to 600 V; thus, their energy is $1.6 \times 10^{-19} \times 600$ J. From this, their de Broglie wavelength can be determined. Use the relationship:

$$\lambda = \frac{h}{\sqrt{2m_e E_k}}.$$

Thus:

$$\lambda = \frac{6.63 \times 10^{-34}}{\sqrt{2 \times 9.1 \times 10^{-34} \times 1.6 \times 10^{-19} \times 600}}$$
$$= 5.0 \times 10^{-11} \text{ m}.$$

This is a similar value to the 7.1×10^{-11} m wavelength of the X-rays.

14.3 SAMPLE PROBLEM 5

Consider a photon and an electron that both have a wavelength of 2.0×10^{-10} m.

(a) Calculate the momentum of the photon and the electron. What do you notice?
(b) Calculate the energy of the photon and the electron. What do you notice?
(c) Summarise what you have found concerning the momentum and energy of a photon and an electron with the same wavelength.

SOLUTION:

(a) The momentum of the photon and the electron are governed by the same equation, namely $p = \frac{h}{\lambda}$. Hence, both the photon and the electron will have the same momentum because they have the same associated wavelength.

Thus:

$$p = \frac{6.63 \times 10^{-34}}{2.0 \times 10^{-10}}$$
$$= 3.3 \times 10^{-24} \text{ N s.}$$

We notice here that both the photon and the electron have the same momentum.

(b) To determine the energy of an object from its momentum, we now have to ask if it is a photon or an object with mass. The relations are different. For the photon, $E = pc$. Thus:

$$E = 3.3 \times 10^{-24} \times 3.0 \times 10^{8}$$
$$= 9.9 \times 10^{-16} \text{ J or } 6.2 \text{ keV.}$$

For the electron, however, $E = \frac{p^2}{2m}$. Thus:

$$E = \frac{(3.3 \times 10^{-24})^2}{2 \times 9.1 \times 10^{-31}}$$
$$= 6.0 \times 10^{-18} \text{ J or } 37 \text{ eV.}$$

The electron has substantially less kinetic energy than the photon, even though they have the same momentum.

(c) Light and matter with the same wavelength will have the same momentum, and vice versa. However, when photons and electrons have the same momentum, they will not necessarily have the same energy. In the problem above, the photon has substantially more energy than the electron.

14.3 Exercise 2

1 Consider a photon and an electron that both have a wavelength 1.0×10^{-10} m.
  (a) Calculate the momentum of the photon and the electron. What do you notice?
  (b) Calculate the energy of the photon and the electron. What do you notice?

2 A photon and an electron both have energy 2.0×10^{-16} J.
  (a) Calculate the momentum of the photon and the electron. What do you notice?
  (b) Use your results in (a) to find the wavelength of the photon and the electron. What do you notice?

3 An electron gun with an accelerating voltage of 28 V is applied to electrons emitted from a tungsten filament to make a beam of electrons all having the same energy.
  (a) What is the energy in both eV and joules that these electrons acquire? Assume the kinetic energy of the electrons emitted from the tungsten filament is zero.
  (b) Use your result to part (a) to determine the momentum of an electron in this beam.
  (c) Calculate the de Broglie wavelength of an electron in this beam.
  (d) If the accelerating voltage is increased, what effect would this have on the de Broglie wavelength of an electron in this beam?

14.3.4 Atoms and standing waves

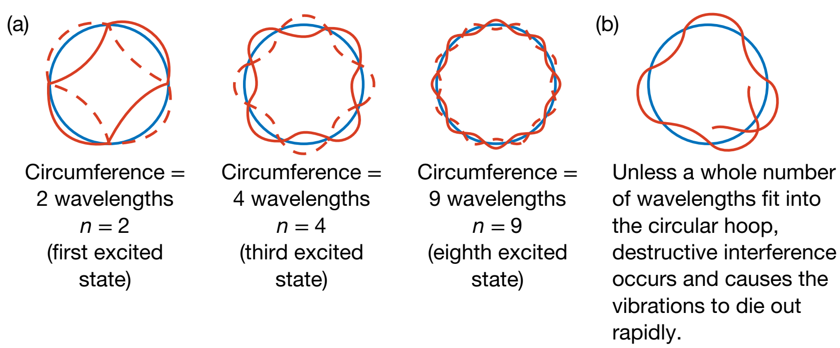
Individual electrons act like waves when they are diffracted by atoms in crystals. Do electrons *in* the atoms also exhibit wavelike properties? They certainly do! Thinking of electrons behaving like waves solved the puzzle of stationary states. This wave model for electrons that are bound within atoms also neatly explained why atoms absorb and emit photons of only particular frequencies, and provided the answers to Rutherford's questioning of the Bohr model of the atom. In essence, only waves whose de Broglie wavelength multiplied by an integer $n\lambda$ set equal to the circumference of a traditional electron orbit are allowed to exist due to these waves being the only ones able to constructively interfere to produce a standing wave. De Broglie speculated about the electron in a hydrogen atom displaying wavelike behaviour in 1924. A complete description of the hydrogen atom awaited a more sophisticated mathematical treatment called quantum mechanics. The fundamentals of this model were developed by Erwin Schrödinger and Werner Heisenberg later in the 1920s.

Louis de Broglie's picture

Louis de Broglie pictured the electron in a hydrogen atom travelling along one of the allowed orbits around the nucleus, together with its associated wave. In de Broglie's mind, the circumference of each allowed orbit contained a *whole number* of wavelengths of the electron-wave so that it formed a standing wave around the orbit. Thus, $n\lambda = 2\pi r$ or $\lambda = \frac{2\pi r}{n}$ fixes the allowed wavelength. An electron-wave whose wavelength was slightly longer, or shorter, would not join onto itself smoothly. It would quickly collapse due to destructive interference. Only orbits corresponding to standing waves would survive. This is shown below. The concept is identical to the formation of standing waves on stringed instruments.

It is worth noting that the standing waves produced on a stringed instrument of length l have a series of possible wavelengths $\lambda_n = \frac{2l}{n}$ where n is a positive integer (1, 2, 3 and so on). This series of wavelengths is called a harmonic series. At this level of physics, which is only an introduction to the conceptual nature of quantum mechanics, the harmonic series provides for a series of associated momenta that are discrete in value. This in turn provides for a series of energy states that are also discrete. This connection is in complete agreement with the observation of emission and absorption spectra. When you pluck a guitar string, only certain frequencies are produced. Likewise, when you energise an atom, only certain energy levels are able to be sustained, resulting in the emission of well-defined frequencies of light in the form of individual photons.

FIGURE 14.11 A model of the atom showing the electron as a standing wave.



In de Broglie's model of the atom, electrons are viewed as standing waves. It is this interpretation that provides a reasonable explanation for the emission spectra of atoms. It answers Rutherford's remark to Bohr (see Section 14.3.1). When a guitar string is plucked, how does it know what frequencies to vibrate at? The answer is: the frequencies that equate to the standing waves with wavelengths compatible with the length of the string.

Electrons viewed as standing waves can exist only in stable orbits with precise or discrete wavelengths. This implies that the electrons can have only discrete quantities of momentum. This in turn implies that the electrons can have only discrete amounts of energy. Energy transitions that are made by electrons occur in jumps from one high-energy standing wave to another standing wave of lower energy. In this way the emission spectra and, hence, absorption spectra can be understood as arising from transitions between quantised energy levels due to electrons having a wavelike character.

It's a consistent story — light displays both wave and particle behaviour and so do electrons and all other forms of matter. The two models are complementary. You observe behaviour consistent with wave properties or particle properties, but not the two simultaneously. Remember how William Bragg expressed it: 'On Mondays, Wednesdays and Fridays light behaves like waves, on Tuesdays, Thursdays and Saturdays like particles, and like nothing at all on Sundays'. This delicate juggling of the two models by both light and matter is known as **wave–particle duality**. Wave–particle duality describes light as having characteristics of both waves and particles. This duality means that neither the wave model nor the particle model adequately explains the properties of light on its own.

There have been many conceptual hurdles for physicists in arriving at this amazingly consistent view of the interaction between light and matter. Their guiding questions always kept them probing for the evidence. Observations and careful analysis gave them the answers. Imagination, creativity and ingenuity were vital in their search for a more complete picture of light and matter.

We now know that both light and matter can exhibit both wavelike and particle-like behaviour, depending on the types of experiments performed. For example, when light strikes a material object, it transfers energy as if it is a particle (the photoelectric effect), but when light passes through a narrow opening or a pair of slits, it acts as if it is a wave. Likewise, matter can have work done on it via well-understood forces accelerating it, but matter can also be diffracted when it passes through a crystal, producing diffraction patterns similar to those of X-rays. Also, the behaviour of electrons within atoms can only be understood by treating them as a type of wave phenomena.

A more detailed model for the seemingly paradoxical result of both wavelike and particle-like behaviour for both light and matter was developed in the 1910s and 1920s. The model is called quantum mechanics, and in it wave and particle behaviour for both light and matter are unified successfully.

14.3.5 Schrödinger and the development of quantum mechanics

So far we have encountered quantum theory as applied to the Bohr model of the atom. That model, however, suffers from not being a complete quantum theory. It is basically a classical model with some quantum ideas superimposed on it. This 'old quantum theory' reached a peak in 1922 but was eventually replaced by a complete quantum theory now known as quantum mechanics.

The first three 'quantum numbers' are the principal quantum number, n from the Bohr model, the angular momentum quantum number, l, and the magnetic quantum number, m.

Heisenberg's uncertainty principle

Pauli used Bohr's idea of shells of electrons and in 1925 realised that if he introduced a fourth quantum number, he could explain the maximum number of electrons in each shell. The fourth quantum number was associated with 'spin'. The maximum number of electrons in each shell corresponded to the number of different sets of quantum numbers available for each shell. Pauli's exclusion principle states that no two electrons can have the same set of quantum numbers.

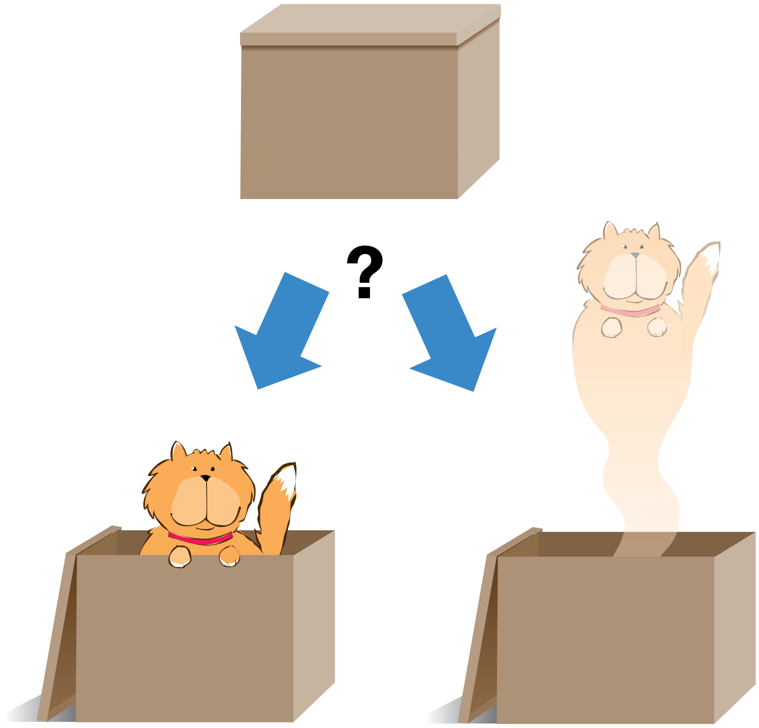
Pauli's exclusion principle provided the reason for electrons in atoms being arranged in shells with the maximum number of electrons being 2, 8, 18, 32, 18, 8 from the first to the sixth shell.

De Broglie's work on the wave nature of particles might not have received wide publicity if it had not been brought to the attention of Einstein. In 1925, Erwin Schrödinger read a comment by Einstein on de Broglie's work that referred to it as more than a mere analogy. Schrödinger then set about trying to restore some of the familiar concepts of waves to quantum theory. He eventually derived equations that looked like the equations used to describe real waves and it seemed that he had managed to bring quantum ideas back towards a much more comfortable formulation associated in some way with classical physics.

Heisenberg did not like Schrödinger's approach. Heisenberg did not see how continuous waves could be used to describe the discontinuous behaviour of an electron jumping from one state to another. Many papers started appearing on Schrödinger's wave mechanics and very few on Heisenberg's matrix mechanics. Schrödinger did not like Heisenberg's non-visual interpretation or the use of matrices. Most physicists of the time preferred Schrödinger's approach to Heisenberg's approach.

However, it was not long before Schrödinger demonstrated that the two different approaches were simply different versions of the same thing. He showed that Heisenberg's matrices could be generated in Schrödinger's theory and Schrödinger's waves could be produced from Heisenberg's matrices.

FIGURE 14.12 Schrödinger's cat could be considered to be both alive and dead at the same time until the box is actually opened and the observation made.



Schrödinger later spent time with Bohr and was most disappointed to find that his 'waves' were not real waves at all. Max Born showed that they were associated with the probability of finding an electron at a particular location.

Schrödinger tried unsuccessfully to restore 'common sense' to quantum theory, but his name has become associated with 'Schrödinger's cat', which brings a quantum nature to a large-scale system. The idea of Schrödinger's cat is that a cat is locked in a box with a deadly cyanide pellet. The release of the cyanide pellet will be triggered by a radioactive decay that might occur at some time. If the box is opened and the cat is observed, it will be alive if the pellet has not been released and dead if it has been released. However, the quantum mechanical view is that if the box is still closed and the cat is not observed, it will be a quantum state that is the superposition of its two possible states; it will be in a state in which it is both dead and alive. Opening the box to observe the cat is said to 'collapse the wave function' into either the dead state or the alive state.

In late 1926, Heisenberg showed that uncertainty is an inherent property of quantum mechanics and that there are pairs of quantities that cannot be determined simultaneously. If we know the accurate position of a particle, say an electron, then you cannot know its momentum accurately. This is because any method that you use to observe the position of the particle will transfer energy to the particle and, hence, change its momentum. Also, the methods that allow the measurement of the momentum of a particle will change its position. This is represented by Heisenberg's uncertainty principle:

$$\Delta x \times \Delta p \geq \frac{\mathrm{h}}{4\pi}$$

where Δx and Δp are the inherent uncertainties in position and momentum and h is Planck's constant.

TRY THIS!

Material: A blindfold, a ruler and a ping-pong ball.
This may help you to get a better grip on the uncertainty principle! Blindfold a volunteer and give them a ruler. Now, place the ping-pong ball at a random location on the bench and get the volunteer to attempt to locate the exact position of the ping-pong ball using only the end of the ruler *without moving the ball!*
(*Note:* Be kind to your volunteer while they are blindfolded!)

The Dirac equation

Paul Dirac extended quantum mechanics and derived the Dirac equation, which added relativity to quantum theory. It predicted correctly the spin of electrons (which is a relativistic effect). It also predicted the existence of a particle similar to an electron but with a positive charge. The anti-electron or positron was observed by Carl Anderson in 1932.

Dirac discovered that the equations of quantum mechanics have the same structure as the equations of classical physics and that the equations of classical physics can be obtained from quantum mechanics by using very large quantum numbers or setting Planck's constant to zero.

14.3.6 What does it all mean?

The complete quantum theory came about after the breakthrough of Heisenberg and Schrödinger, who independently in 1925 and 1926 discovered different forms of the same theory. Heisenberg introduced the uncertainty principle and Bohr completed the theory with his principle of complementarity.

By the Solvay Conference of October 1927, the old quantum theory had been replaced. At this conference, Schrödinger presented a paper on his wave function theory but he declined to discuss the interpretation of the wave functions (which Born interpreted as being related to the probability of finding an electron in a certain location). The theory is now called quantum mechanics, and Bohr's ideas along with Heisenberg's uncertainty principle and Born's probability interpretation became known as the Copenhagen interpretation.

At the Solvay Conference, Einstein raised his first public objections to quantum mechanics, and he was to continue to debate with Bohr this interpretation of quantum mechanics. Einstein never accepted that quantum mechanics was a 'complete' theory, and the Copenhagen interpretation is still considered obscure by some physicists today. It is no wonder that Bohr made his famous statement, 'Anyone who is not shocked by quantum theory has not understood it'.

We have not even scratched the surface of quantum mechanics. We have seen that there were major problems with the ideas of the original quantum theory and that in the process of overcoming those problems, a new theory was developed that required a modification of our ideas about the physical world.

In this strange new theory there is no such thing as a particle or a wave but rather there is a wave-particle duality and making an accurate observation of one property means that another property cannot be measured accurately. We have not studied the mathematics of either the matrix mechanics of Heisenberg or the wave mechanics of Schrödinger and we have not studied the probability interpretation of Born. We are therefore not in a position to see why quantum theory and in particular the Copenhagen interpretation is as shocking as Bohr suggests.

A deeper study of quantum mechanics leads to an atomic world that is fuzzy and nebulous and in which, according to Bohr and the Copenhagen interpretation, nothing actually exists until it is observed. The clockwork world of Newton becomes a world of quantum uncertainty where nothing is predictable. Even worse, events can occur without having a cause and quantum particles can suddenly pop into existence.

Yet, despite all the problems of interpretation, quantum mechanics is an incredibly successful theory. Quantum mechanics helps us explain and control the properties of metals, insulators, semiconductors and superconductors. The inventors of the transistor acknowledge the part quantum theory played in their discovery. That discovery led to the development of ever more powerful computers and microcomputers that have led to a revolution in communications and information. Lasers and masers are quantum devices. Quantum mechanics explains the structure of the atom and nucleus as well as mechanical and thermal properties of solids. Quantum mechanics gave chemistry a firm base and explained chemical bonding. The new areas of molecular biology and genetic engineering have arisen from quantum chemistry. In astrophysics, the processes that occur in stars can be explained by quantum mechanics and even our theories regarding such exotic objects as black holes are based on quantum mechanics. There has even been the suggestion that our universe began as a 'quantum fluctuation'. Perhaps in the not too distant future, quantum computing will become a reality and quantum computers will be vastly more powerful than today's computers.

Quantum mechanics is a theory that has changed the world and our view of it.

eBook*plus* RESOURCES

Watch this eLesson: De Broglie wavelength
Searchlight ID: med-0199

Watch this eLesson: The momentum of photons and matter
Searchlight ID: med-0423

Explore more with this weblink: The atomic lab: electron interference

14.4 Review

14.4.1 Summary

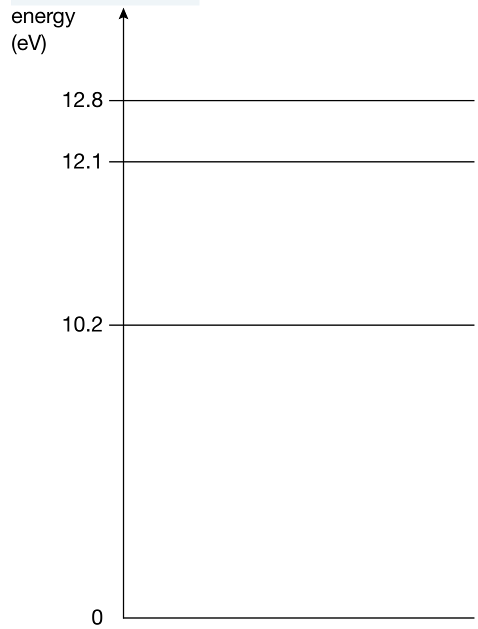
- The scattering of alpha particles through large angles by very thin gold foils led Rutherford to propose that an atom consisted of a very small, dense, positively charged nucleus. Electrons were in orbit about the nucleus at distances very large compared to the dimensions of the nucleus.
- A major problem with the Rutherford model was that it did not account for any properties of the electrons in the atom, in particular how the electrons could be accelerating without emitting electromagnetic radiation.
- Bohr extended the Rutherford model by formulating two postulates that enabled him to apply the quantum ideas of Planck and Einstein to the Rutherford atom.
- Bohr's postulates enabled him to describe an atom in which electrons existed in stable 'stationary states' where they did not emit electromagnetic radiation. The transition of an electron from one stationary state to another would be accompanied by the emission or absorption of a quantum of electromagnetic radiation or a photon.
- Using his model of the atom, Bohr was able to derive a theoretical expression for the wavelengths of the spectral lines of hydrogen, which was in agreement with Balmer's empirical formula.
- To account for emission spectra, Neils Bohr proposed a radical model where electrons within atoms have stable orbits but only discrete energy levels are allowed.
- While successful in explaining the wavelengths of the spectral lines in the hydrogen spectrum, Bohr's model failed to account for the relative intensities of the lines, the existence of the hyperfine structure of the lines or for the splitting of spectral lines when the excited gas was in a magnetic field. Bohr's model was also a strange mixture of classical physics and quantum physics.
- When an atom jumps from a high energy level, E_{initial}, to a lower energy level, E_{final}, resulting in a difference, ΔE, a photon of light is emitted with frequency, f, according to the equation $\text{h}f = \Delta E$. Hence, the observation of emission spectra having precise frequencies is evidence for atoms having discrete energy levels.
- The best model for atoms having discrete energy levels is to interpret electrons in atoms as behaving as a standing wave. The allowable standing waves are known as orbitals.
- In 1924 Louis de Broglie suggested that electrons may exhibit wave properties under suitable conditions. He proposed a diffraction experiment using a beam of electrons and a crystal to act as a diffraction grating.
- The de Broglie wavelength, λ, can be determined from the momentum, p, according to the equation $\lambda = \frac{\text{h}}{p}$. Remember also that the momentum of a particle is given by $p = mv$, where m is the mass and v is the speed.
- In 1927 Clinton Davisson and Lester Germer established the wavelike behaviour of electrons when they performed a diffraction experiment. Not only did they observe diffraction effects, they also established that the wavelength of the electrons in the beam was consistent with Louis de Broglie's prediction.

- The theory now known as quantum mechanics was introduced by Heisenberg and developed by Bohr, Schrödinger and others.
- In particular, Schrödinger's wave equation describing the quantum behaviour of both light and matter became the cornerstone for modern quantum mechanics in the same way that Newton's laws of motion became the foundation of classical physics.

14.4.2 Questions

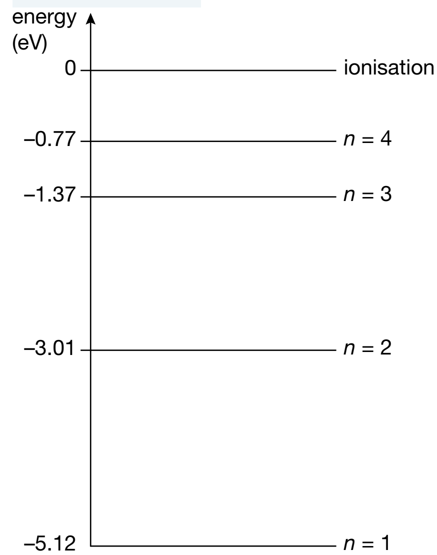
1. List the strengths and weaknesses of the Rutherford model of the atom.
2. There are two common ways of depicting the energy levels of an atom. In one method the ground state is taken to be zero energy, and in the other method the ionisation energy is taken to be zero. The first excited state of mercury atoms is known to be 4.9 eV above the ground state, the second excited state is 6.7 eV, the third excited state is 8.8 eV, and the ionisation energy is 10.4 eV above the ground state. Using the second method, where the ionisation energy is taken as 0 eV, give the energies of the ground state and the first three excited states. *Note:* Your values will be negative numbers, and a drawing of the energy level diagram will assist you.
3. The ground state and the first three excited states of hydrogen are shown in Figure 14.13. An emission spectrum of hydrogen gas shows many different spectral lines.
   (a) Copy the diagram and label the ground state and first three excited states.
   (b) Draw arrows to represent all six possible transitions that may occur when hydrogen atoms in states lower than the fourth excited state emit a photon of light.
   (c) Calculate the energy of each of the possible six photons.
   (d) Determine the wavelength of the photon having the least and greatest energy in your answer to part (c).

FIGURE 14.13



4. When sodium chloride (common salt) is placed in a flame, the flame glows bright gold. Figure 14.14 shows some of the energy levels of a sodium atom.
   (a) On a copy of the diagram, label the ground state of the atom, and the first excited state.
   (b) Draw arrows to represent the change in energy of atoms in the ground state that absorb energy during collisions with other particles in the flame.
   (c) Calculate the wavelength of light emitted by these atoms as they return to the ground state in a single jump. Which energy change is responsible for the yellow glow?

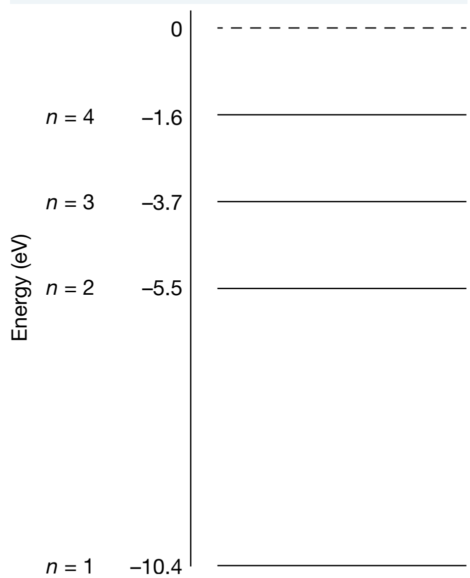
FIGURE 14.14



5. Use Balmer's equation to calculate the wavelength of the radiation emitted from an excited hydrogen atom when an electron undergoes a transition from the state $n = 5$ to:
   (a) the state $n = 1$ (b) the state $n = 2$ (c) the state $n = 3$.

6. (a) Calculate the wavelengths of the lines of the Balmer series corresponding to transitions from the states $n = 8$, $n = 10$, $n = 12$.
   (b) What trend do you notice in the wavelengths as the value of n increases?
7. The radius of the orbit of an electron in the ground state of the hydrogen atom is 5.3×10^{-11} m. Calculate the radius of the orbit of an electron when it is in each of the following states:
   (a) the state $n = 2$
   (b) the state $n = 3$
   (c) the state $n = 4$.
8. (a) State which photon, red or blue, has the higher frequency.
   (b) State which photon, red or blue, has the longer wavelength.
   (c) State which photon, red or blue, has the higher energy.
9. If the atoms in a sample of hydrogen were all in the state $n = 5$, how many different spectral lines could possibly be produced by the gas as the electrons returned to the ground state?
10. Given that $E_1 = -13.6\,\text{eV}$, $E_2 = -3.40\,\text{eV}$, $E_3 = -1.51\,\text{eV}$, $E_4 = -0.85\,\text{eV}$, $E_5 = -0.54\,\text{eV}$, calculate the wavelengths of:
    (a) the first two lines in the Lyman series
    (b) the first two lines in the Balmer Series
    (c) the first two lines in the Paschen series.
11. (a) What is the wavelength of the longest wavelength spectral line of the Pfund series?
    (b) What is the wavelength of the shortest wavelength line of the Pfund series?
12. The 'series limit' is the term applied to the shortest wavelength spectral line in each of the spectral series of hydrogen.
    (a) What value of n_i would be used to calculate the wavelength of the series limit?
    (b) Calculate the series limit for the Lyman, Balmer and Paschen series of hydrogen.
    (c) How many electron volts of energy would be carried by a photon corresponding to the series limit of the Lyman series?
13. Figure 14.15 is an energy level diagram for energies of the stationary states in atoms of a gas, Q.

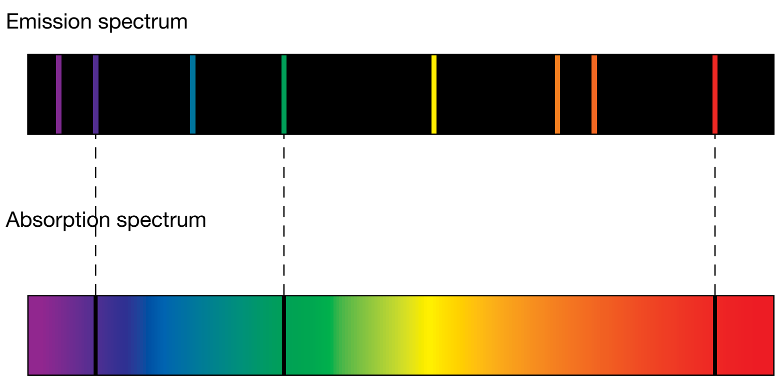
FIGURE 14.15 The energy level diagram for the gas, *Q*.



(a) i. Determine the energy of the photon emitted when an electron in the state $n = 3$ undergoes a transition to the state $n = 2$.
ii. Determine the frequency and wavelength of this photon.
(b) Determine the wavelength of the photon absorbed by this gas when an electron undergoes a transition from the state $n = 1$ to the state $n = 4$.

14. The emission spectrum of a particular gas has eight bright lines in the visible region as shown in Figure 14.16. The absorption spectrum of the same gas has only three lines in the visible region as shown.
(a) Explain why each of the absorption lines corresponds to one of the emission lines.
(b) Explain why there is not a corresponding absorption line for five of the emission lines.

FIGURE 14.16 The emission and absorption spectra for a particular gas.



15. An absorption spectrum is produced when the atoms in a cool gas absorb energy from white light passing through the gas. These excited atoms then re-emit the energy and return to low energy states. How can this re-emission occur but there still be dark lines in the absorption spectrum?
16. Balmer predicted accurately the wavelengths of the visible spectral lines and invisible spectral lines of hydrogen that had not been detected. Bohr did the same about thirty years later. Explain why Bohr's prediction is considered more important than that of Balmer.
17. What evidence supports the idea that the electron energies in the hydrogen atom are discrete?
18. If electrons in hydrogen atoms obeyed the rules of classical mechanics instead of those of quantum mechanics, would the hydrogen atoms produce a line spectrum or a continuous spectrum? Explain your answer.
19. Explain why each element has its own characteristic spectrum.
20. When de Broglie was examined for his PhD, his thesis was first thought by his examiners to bear little relationship to reality.
(a) What did de Broglie predict that made it seem to be unrelated to reality?
(b) What did de Broglie suggest could be observed to support his prediction?
21. If a proton and an electron are travelling with equal velocities, which has the longer de Broglie wavelength?
22. If one electron travels twice as fast as another electron, which one has the greater wavelength?
23. (a) If an electron travelling at 1.0×10^4 m s^{-1} was accelerated to 2.0×10^4 m s^{-1}, what would be the ratio of its new wavelength to its original wavelength?
(b) If an electron travelling at 1.0×10^8 m s^{-1} was accelerated to 2.0×10^8 m s^{-1}, would it change its wavelength by the same amount as the electron in part (a)? Explain your answer.
24. (a) Calculate the de Broglie wavelength of an electron in a TV set that hits the screen with a velocity of one-tenth of the velocity of light.
(b) With what velocity would you roll a ball of mass 0.1 kg if it is to have the same de Broglie wavelength as the electron in part (a)?

25. A neutron emitted when a uranium-235 nucleus undergoes fission may have an energy of about 1 MeV. A 'thermal' neutron that would be captured by a uranium-235 nucleus in a nuclear reactor would have an energy of about 0.02 MeV.
    (a) Calculate the wavelength of a 1 MeV neutron.
    (b) Calculate the wavelength of a 0.02 MeV neutron.
26. Construct a time line to include the following physicists and indicate their contribution to the transition from classical mechanics to the development of quantum mechanics: J.J. Thomson, Rutherford, Bohr, Balmer, de Broglie, Heisenberg, Schrödinger.

eBook*plus* RESOURCES

Complete this digital doc: Investigation: The spectrum of hydrogen
Searchlight ID: doc-26602

PRACTICAL INVESTIGATIONS

Investigation 14.1: The spectrum of hydrogen

Aim

To observe the spectral lines of hydrogen, measure their wavelengths and compare these values with the theoretical values.

Apparatus

hydrogen spectral tube and power supply spectroscope

Theory

According to the Bohr model of the hydrogen atom, when an electron jumps from a higher energy state to a lower energy state, it will emit a photon. When an electron jumps to the state $n = 2$ from any of the states from $n = 3$ to $n = 6$, the emitted photon will be in the visible region of the spectrum.

A spectroscope can be used to measure the deviation of the spectral lines. The wavelengths of the spectral lines can then be calculated.

The wavelengths of the spectral lines will be given by:

$$\lambda = \frac{d \sin \theta}{n}$$

where
λ = wavelength
θ = angle of deviation
d = distance between lines on the grating
n = order of spectra.

The theoretical values of the wavelengths, based on Bohr's theory of the hydrogen atom, can be calculated after the energies of the states $n = 2$ to $n = 6$ have been calculated.

The energy of the ground state, $n = 1$ is -13.6 eV. The energies of the other states are given by $E_n = \frac{E_1}{n^2}$.

Method

In this experiment, the hydrogen spectral tube is switched on and the radiation viewed through a spectroscope.

Setting up the hydrogen spectral tube

Different types of spectral tube and power supply may be used but we will describe a special power supply that is designed for spectral tubes. The spectral tube can be clamped in place on a vertical metal rod mounted on top of the power supply. (The rod is maintained at Earth potential and is safe to touch when the power supply is switched on.)

Some hydrogen spectral tubes are very faint and this makes measurement of the spectral lines very difficult. Hydrogen spectral tubes should probably be replaced fairly regularly as they tend to become fainter over time.

The spectroscope

The spectroscope consists of two tubes, one of which can be rotated around a small central table. One tube, the fixed one, is a collimator and the moveable one is a telescope. A small prism or a diffraction grating can be mounted on the small table.

There is an adjustable narrow slit at the front of the collimator. The collimator is set up to shine parallel rays of light onto the diffraction grating or prism. (For the remainder of this practical investigation, we will assume that a diffraction grating is being used. We will assume that the information about the number of lines per metre is provided. It is possible to calibrate a grating, and a procedure to do this is included in the last section of the method.)

The light that passes through the diffraction grating deviates through an angle that depends on the wavelength of the light and the number of lines per metre ruled on the diffraction grating.

The telescope is rotated around the table and the image of the narrow slit is observed at different angles for the different wavelengths of light. These angles can be measured, usually with the help of a vernier scale fixed to the telescope.

Setting up the spectroscope

Setting up the spectroscope involves two parts; adjusting the telescope for parallel light rays and then adjusting the collimator to produce parallel light rays.

There should be fine cross-wires visible in the eyepiece of the telescope. These cross-wires should be in sharp focus and an adjustment of the eyepiece in its holder may have to be made if they are not sharply focused.

The telescope should be pointed at a distant object and the focus adjusted using the objective lens of the telescope, lens L_3, until the image of the distant object is sharply focused. (In fact any object outside should be far enough away.) The telescope is then aligned with the collimator and the lens on the collimator is adjusted until the slit is seen sharply focused.

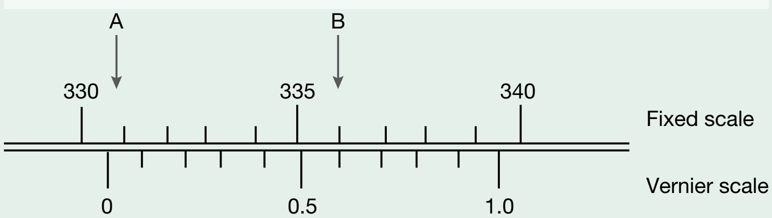
A light source, possibly a brighter spectral tube than the hydrogen tube, can now be set up in front of the slit and the slit width adjusted until narrow spectral lines can be viewed when the telescope is rotated to the appropriate position.

By clamping the telescope and then using the fine adjustment, it should be possible to align the cross-wires visible in the eyepiece with the spectral line. A measurement can then be made.

Reading a vernier scale

A vernier scale has ten lines on the moveable scale in the space of nine lines on the fixed scale. This enables an extra decimal place to be determined. This extra digit corresponds to the position of the line on the vernier-moveable scale that aligns with any one of the lines on the fixed scale.

FIGURE 14.17 Reading a vernier scale. The position of the zero mark on the vernier scale, indicated by arrow A, is just less than 331. The line on the vernier scale that matches a line on the fixed scale is 0.6, as indicated by arrow B. Therefore, the reading is $330.6°$.



Measuring the wavelengths of the spectral lines of hydrogen

The lines will probably be quite faint and it will probably be necessary to have the apparatus in a darkened room to observe the lines clearly. The most difficult part is aligning the spectral lines with the cross-wires. If the room is completely dark, it will be impossible to see the cross-wires. A small amount of field illumination is necessary to be able to see the cross-wires.

There should be no problem with making the measurement for the straight through position. There should be sufficient light coming directly through the slit to make locating the image of the slit on the cross-wires quite easy.

Record this value and then record the reading of as many of the spectral lines as possible. (If it is possible to measure any of the spectral lines of the second order spectrum it is worth doing so.)

Calibration of diffraction grating

If necessary, the diffraction grating could be calibrated using a sodium vapour spectral tube. Set up this tube and observe the angle to the very bright orange line in the first order spectrum of sodium ($n = 1$). This line is really a double line, the wavelengths of the lines being 589.0 nm and 589.6 nm.

You can use the information in the equation $\lambda = \dfrac{d \sin\theta}{n}$ to calculate d.

Results

Record your results in a table similar to the table below and calculate the wavelengths of the spectral lines.

The number of lines per centimetre or perhaps even the number of lines per inch is probably supplied with the diffraction grating. It will be necessary to convert this to lines per metre and d is the inverse of this value.

Record the reading of the straight through position θ_0.

SPECTRAL LINE COLOUR	POSITION θ	ANGLE $\theta - \theta_0$	ORDER OF SPECTRA (n)	WAVELENGTH
Faint violet			1	
Violet			1	
Blue-Green			1	
Red			1	
			2	
			2	
			2	
			2	

Analysis

1. The energy of the ground state of hydrogen is $E_1 = -13.6$ eV.
   The energies of the other states are given by
   $$E_n = \frac{E_1}{n^2}.$$
   Determine the energy, in electron volts, of the energy states $n = 2, 3, 4, 5,$ and 6.
2. Draw an energy level diagram and calculate the energies (in electron volts) of photons emitted when an electron jumps to the $n = 2$ state from each of the four higher energy states.
3. Convert these values from electron volts to joules and calculate the wavelengths of these photons.
   Use:
   $$E = \mathrm{h}f = \mathrm{h}\frac{\mathrm{c}}{\lambda}$$
   $$\lambda = \frac{\mathrm{hc}}{E}$$
   where
   $$1\ \mathrm{eV} = 1.6 \times 10^{-19}\ \mathrm{J}$$
   $$\mathrm{h} = 6.602 \times 10^{-34}\ \mathrm{J\,s}.$$
4. Compare the values of the wavelengths calculated above with the values determined from the measurements of the angles.

Questions

1. How accurate do you consider your determination of the wavelengths of the spectral lines? Aside from any difficulty with aligning the spectral lines with the cross-wires, you are restricted to measuring the angle to the nearest $0.1°$. Consider how a change in angle of $0.1°$ will alter your calculations.
2. Taking into account the expected accuracy of your observations, do you consider that your results are in agreement with the theoretical values of the wavelengths of these four spectral lines of hydrogen?

TOPIC 15
Properties of the nucleus

15.1 Overview

15.1.1 Module 8: From the universe to the atom

Properties of the nucleus

Inquiry question: How can the energy of the atomic nucleus be harnessed?

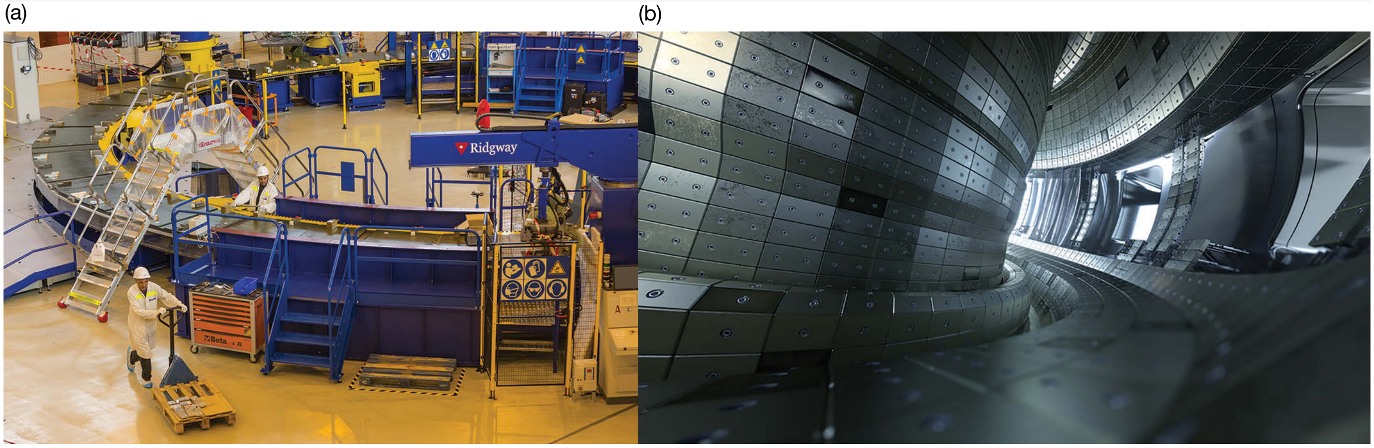
Students:

- analyse the spontaneous decay of unstable nuclei, and the properties of the alpha, beta and gamma radiation emitted (ACSPH028, ACSPH030)
- examine the model of half-life in radioactive decay and make quantitative predictions about the activity or amount of a radioactive sample using the following relationships:
  - $N_t = N_0 e^{-\lambda t}$
  - $\lambda = \dfrac{\ln(2)}{t_{\frac{1}{2}}}$

  where N_t = number of particles at time t, N_o = number of particles present at $t = 0$, $\lambda =$ decay constant, $t_{1/2}$ = time for half the radioactive amount to decay (ACSPH029)
- model and explain the process of nuclear fission, including the concepts of controlled and uncontrolled chain reactions, and account for the release of energy in the process (ACSPH033, ACSPH034)
- analyse relationships that represent conservation of mass-energy in spontaneous and artificial nuclear transmutations, including alpha decay, beta decay, nuclear fission and nuclear fusion (ACSPH032)
- account for the release of energy in the process of nuclear fusion (ACSPH035, ACSPH036)
- predict quantitatively the energy released in nuclear decays or transmutations, including nuclear fission and nuclear fusion, by applying: (ACSPH031, ACSPH035, ACSPH036)
  - the law of conservation of energy
  - mass defect
  - binding energy
  - Einstein's mass–energy equivalence relationship ($E = mc^2$)

FIGURE 15.1 The International Fusion Energy Organization (ITER) is developing the tokamak, which is an experimental machine designed to harness the energy of fusion. (a) The coils winding facility (b) The tokamak fusion reaction chamber.

(a) (b)



15.2 Radioactivity

15.2.1 Naturally occurring radioactivity — the spontaneous decay of unstable nuclei

For over a hundred years, humans have burned fossil fuels to meet their energy demands, a practice that is driving climate change. Is it possible that our knowledge of atomic nuclei could be used to solve this dilemma, using naturally occurring radioactive materials, to provide us with the energy we need without polluting our world or changing its climate?

Henri Becquerel discovered radioactivity in 1896 when he was studying the radiation emitted from **phosphorescent** substances that had previously been exposed to sunlight. Becquerel found by accident that a salt of uranium, potassium–uranyl sulfate, continuously emitted radiation regardless of whether or not it had been exposed to sunlight. This radiation penetrated matter, passing through black paper (opaque to light) and causing a photographic plate to become darkened. It seemed to be similar in nature to X-rays, which had been recently discovered by Wilhelm Röntgen (1845–1923).

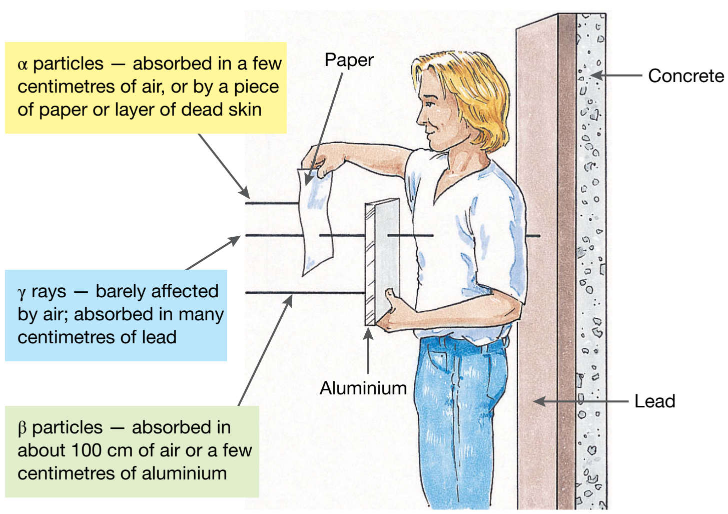
In 1898, Rutherford showed that there were two components (alpha and beta rays) of the radiation discovered by Becquerel, and in 1900 Paul Villard (1860–1934) discovered the third component (gamma rays).

The properties of alpha, beta and gamma radiation can be summarised as follows.

Penetrating power

Figure 15.2 shows that the penetrating power is lowest for alpha particles, which can be stopped by a sheet of paper or a few centimetres of air. Beta particles will be stopped by many metres of air or a sheet of aluminium about a centimetre thick. Gamma rays may pass through a few centimetres of lead or many metres of concrete before being stopped.

FIGURE 15.2 The relative penetrating powers of alpha, beta and gamma radiation.



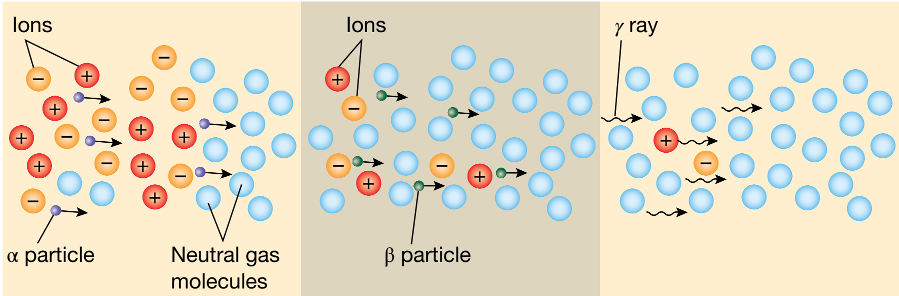
Ionising power

As might be expected, the ionising power is the inverse of the penetrating power. Alpha particles interact most strongly with matter and hence have a low penetrating power and high ionising power. The ionising power of beta particles is lower and that of gamma rays is very low (see Figure 15.3).

Deflection by a magnetic field

The paths of the different types of radiation through a magnetic field proved to be harder to observe, but once detected they indicated that alpha particles were positively charged, beta particles were negatively charged and gamma rays were neutral.

FIGURE 15.3 Ionisation caused by radioactive emissions passing through gas.



(a) Alpha particles cause intense ionisation in a gas.

(b) Beta particles ionise a gas much less than alpha particles do.

(c) Gamma rays ionise a gas even less than beta particles.

FIGURE 15.4 The paths of alpha, beta and gamma rays through magnetic fields.

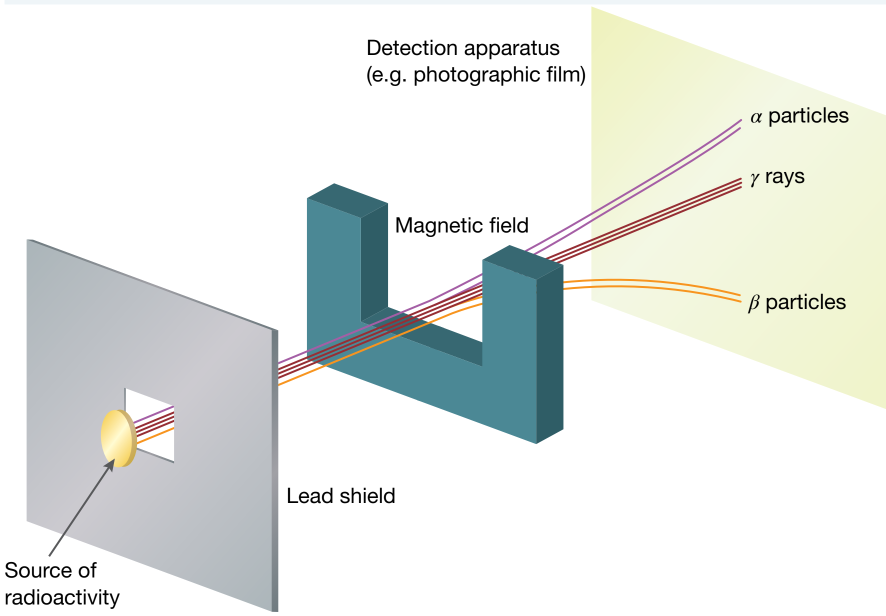


TABLE 15.1 Summary of properties and identities of alpha, beta and gamma radiation.

Type of radiation	Penetrating power	Ionising power	Path through magnetic field	Nature of radiation (in today's terms)
Alpha	Very low	Very high	Curved path of positive charge	Helium nucleus
Beta	High	Moderate	Curved path of negative charge	Electron
Gamma	Extremely high	Very low	Not deflected	High energy photon

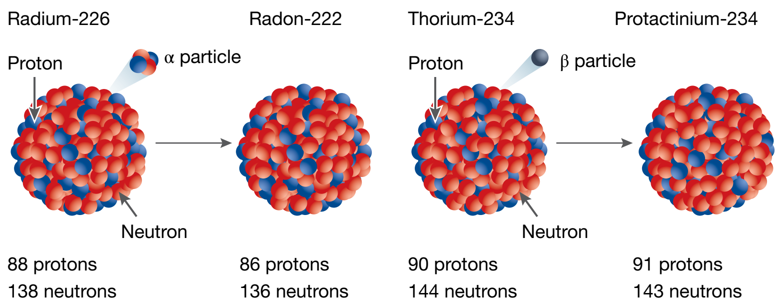
15.2.2 Naturally occurring radioactivity explained

In 1902, Rutherford and the English chemist Frederick Soddy (1877–1956) published a paper proposing that the emission of radioactivity was the result of 'radioactive transformation'. When a radioactive atom emitted an alpha particle or beta particle, the atom split into two. The alpha or beta particle was emitted and what remained was a heavy leftover part that was chemically and physically different from the parent atom (see Figure 15.5).

This 'transformation', 'disintegration', 'decay' or '**transmutation**' was responsible for turning one element into another.

After the discovery of the nucleus, the transmutation was identified as the emission of alpha or beta particles *from the nucleus*.

FIGURE 15.5 The release of an alpha particle from an atomic nucleus is called alpha decay. Similarly, the release of a beta particle is called beta decay. In each case, the number of protons in the nucleus changes and hence the nucleus transmutes into a different element.



PHYSICS FACT

Writing nuclear equations

The formulas for nuclei are written in the form ${}^{A}_{Z}X$ where X is the symbol for the element, A is the mass number (number of protons plus neutrons) and Z is the atomic number (number of protons).

The term **nuclide** is used to denote a nucleus characterised by particular values of Z and A. If a group of nuclides share the same atomic number but have different mass numbers, they are referred to as **isotopes** of that element.

In any nuclear reaction, the sum of the mass numbers before the reaction must be equal to the sum of the mass numbers after the reaction. The sum of the atomic numbers before the reaction must likewise be equal to the sum of the atomic numbers after the reaction. This can be slightly complicated by the fact that if a beta decay is involved, the electron is assigned an atomic number of negative 1. It may be necessary to look up a periodic table to determine the element formed if not all the information is supplied.

The equations for the transmutations associated with some common examples of alpha decay and beta decay are:

$${}^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\text{He}$$

$${}^{234}_{90}\text{Th} \rightarrow {}^{234}_{91}\text{Pa} + {}^{0}_{-1}\text{e}$$

We can see that alpha decay reduces the atomic number by two and the mass number by four, and beta decay increases the atomic number by one and leaves the mass number unchanged.

15.2 SAMPLE PROBLEM 1

Write the nuclear equation for the beta decay of lead-209 into bismuth-209.

SOLUTION:

$$^{209}_{82}\text{Pb} \rightarrow {}^{209}_{83}\text{Bi} + {}^{0}_{-1}\text{e}$$

15.2 SAMPLE PROBLEM 2

It is possible to bombard beryllium-9 with alpha particles and force it to form a new substance. In doing so, a neutron is ejected. Write the nuclear reaction for this process and identify the new substance.

SOLUTION:

$$^{9}_{4}\text{Be} + {}^{4}_{2}\text{He} \rightarrow {}^{12}_{6}\text{C} + {}^{1}_{0}\text{n}$$

In a nuclear reaction, the superscripts (mass numbers) and subscripts (atomic numbers) must balance on the left and right sides of the equation. By doing this, the new substance is identified as carbon-12.

15.2 Exercise 1

1 Write the equations for:
- (a) alpha decay of americium-241
- (b) β^- decay of platinum-197

15.3 The model of half-life in radioactive decay

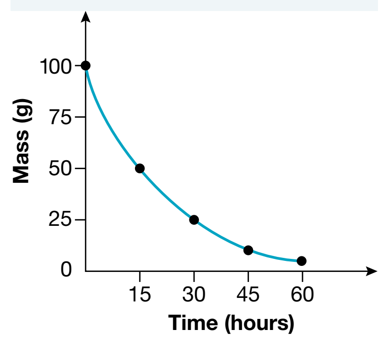
15.3.1 Half-life of a radioactive isotope

Not all radioactive isotopes decay at the same rate. The rate of decay is measured by the **half-life** of the radioisotope. This is the time taken for half the radioactive material to decay.

Half-lives can vary significantly. Uranium-238 has a half-life of 4.5×10^9 years whereas polonium-218 has a half-life of only 1.5×10^{-4} seconds.

A radioisotope with a very long half-life is unsuitable for medical diagnosis as it lingers in the patient after all necessary measurements are taken. This can pose a danger to the patient and people in close contact because of the radiation emitted. On the other hand, if the half-life is too short, the radioisotope either loses its useful radiation before measurements can be taken or has to be administered in a dangerously large dose. Radioisotopes with half-lives ranging from several minutes to days are used for medical diagnosis.

FIGURE 15.6 The radioactive decay of sodium-24.



The decay of a radioisotope can be plotted on a graph from which the half-life can be read. In the graph in Figure 15.6, we see that there is initially 100 g of sodium-24. From the graph, the mass has halved to 50 g after 15 hours. The half-life $\left(t_{\frac{1}{2}}\right)$ of sodium-24 is therefore 15 hours.

15.3 SAMPLE PROBLEM 1

A 20 mg sample of iodine-123 is to be used as a radioactive tracer in the body. The half-life of the iodine-123 is 13 hours.

(a) How long will it take for 17.5 mg to decay?
(b) Calculate how much iodine-123 will remain after 26 hours.

SOLUTION:

(a) In 1 half-life, 10 mg of iodine-123 will decay. This will leave 10 mg iodine-123.
In the second half-life, 5 mg iodine-123 will decay, leaving 5 mg iodine-123.
In the third half-life, 2.5 mg iodine-123 will decay. Altogether, 17.5 mg (10 + 5 + 2.5 mg) iodine will have decayed in 3 half-lives or 39 hours.
(b) 26 hours is 2 half-lives (2×13 hours).
After 1 half-life, 10 mg of iodine-123 will decay, leaving 10 mg iodine-123.
After 2 half-lives, 5 mg iodine-123 will decay, leaving 5 mg iodine-123.
5 mg iodine-123 will remain after 26 hours.

15.3 SAMPLE PROBLEM 2

A radioisotope sample has a half-life of 10.0 minutes.

(a) Calculate the time it will take the activity to drop from 8.0 MBq (mega becquerels) to 4.0 MBq.
(b) Calculate the time it will take for its activity to be 1.0 MBq.

SOLUTION:

(a) When half the sample has decayed, the activity will also halve. This assumes that the atoms formed are not radioactive. Hence the time needed to reduce the activity to 4.0 MBq is one half-life, or 10.0 minutes.
(b) Halving the activity each half-life means 3 half-lives have passed before the activity is 1.0 MBq. The time taken is 30.0 minutes.

The relationship graphed in Figure 15.6 can be expressed mathematically by first noting that the rate of decay of a certain number of atoms is proportional to the number of atoms present. This can be expressed using a derivative function:

$$\frac{dN}{dt} = -\lambda N \ \left(\text{where } \lambda \text{ is known as the decay constant, units s}^{-1}\right)$$

This equation can be integrated to give

$$N_t = N_0 e^{-\lambda t}$$

where N_0 = the number of atoms present at time $t = 0$ (that is, at the start), and
N_t = the number of atoms present at time t.
The same expression can be used for the decay rates R:

$$R_t = R_0 e^{-\lambda t}$$

From this expression, it is possible to derive the following expression for half-life $t_{\frac{1}{2}}$:

$$t_{\frac{1}{2}} = \frac{\ln 2}{\lambda} \approx \frac{0.6931}{\lambda}$$

This formula can also be rearranged to give $\lambda = \dfrac{\ln\ 2}{t_{\frac{1}{2}}}$ so that the decay constant can be calculated if the half-life is known.

15.3 SAMPLE PROBLEM 3

The activity of a radioactive source is initially 600 counts per minute but drops to 100 after 6 hours. Determine the half-life and decay constant for this source.

SOLUTION:

$$N_t = N_0 e^{-\lambda t}$$

$$100 = 600 e^{-6\lambda}$$

$$\frac{100}{600} = e^{-6\lambda}$$

$$\ln\left(\frac{100}{600}\right) = -6\lambda$$

$$\Rightarrow \lambda = \frac{\ln\left(\frac{100}{600}\right)}{-6}$$

$$= 0.299\ \mathrm{s}^{-1}$$

And hence the half-life can be found using $t_{\frac{1}{2}} = \dfrac{0.6931}{\lambda}$

$$= \frac{0.6931}{0.299}$$

$$= 2.32\ \text{hours}$$

15.3 Exercise 1

1 A 20 mg sample of iodine-123 is to be used as a radioactive tracer in the body. The half-life of iodine-123 is 13 hours.
   (a) How long will it take for 17.5 mg to remain undecayed?
   (b) Calculate how much iodine-123 will remain after 24 hours.

2 A radioisotope sample has a half-life of 10.0 minutes. Calculate the time it will take the activity to drop from 8.0 MBq (mega becquerels) to
   (a) 4.0 MBq (b) 3.0 MBq (c) 2.0 MBq

3 The activity of a radioactive source is initially 350 counts per minute but drops to 100 after 4 hours. Determine the decay constant and half-life for this source.

eBook*plus* RESOURCES

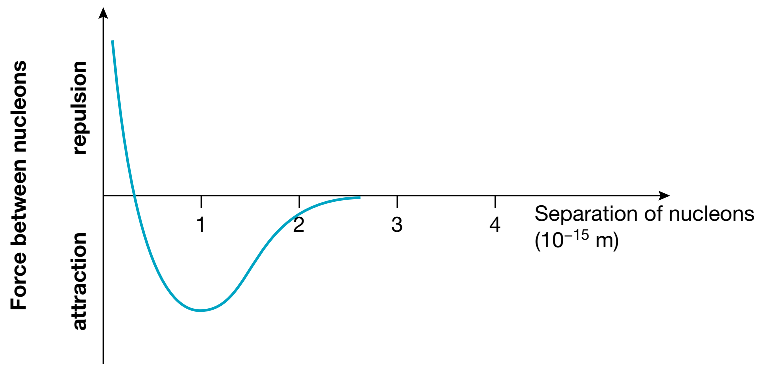
Watch this eLesson: Nuclear stability and radiation
Searchlight ID: eles-2518

15.4 Nuclear stability

15.4.1 What holds the nucleus together?

The force that holds electrons around a nucleus is called an electrostatic force. Electrostatic forces increase as charges move closer together. Electrostatic attraction exists between unlike charges; electrostatic repulsion exists between like charges. So, it seems strange that the positive charges inside a nucleus don't repel each other so strongly that the nucleus splits apart. In fact, two protons do repel each other when they are brought together, but in the nucleus they are so close to each other that the force of repulsion is overcome by an even stronger force — **strong nuclear force**. While the strong nuclear force is, as its name suggests, a very strong force, it is able to act over only incredibly small distances. Inside a nucleus, the nucleons are sufficiently close that the pull of the strong nuclear force is much greater than the push of the protons repelling each other, and therefore the nucleus remains intact.

FIGURE 15.7 The graph shows how the strong nuclear force between two nucleons varies with the separation of the nucleons.

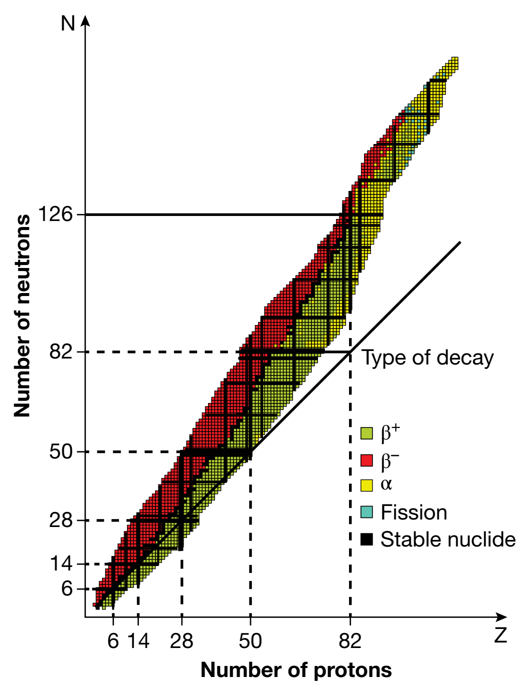


15.4.2 The proton to neutron ratio

The stability of any nucleus depends on the number of protons and neutrons. For small nuclei to be stable, the number of protons must roughly equal the number of neutrons. As the number of protons increases, however, more neutrons are needed to maintain stability.

In larger nuclei, the cumulative repulsion of the protons makes the nucleus less stable. Neutrons do not have a charge and so do not add to that repulsion, but they do still share the strong nuclear force. Hence, in order to be stable, it becomes necessary for larger nuclei to possess more neutrons than protons. This is an increasing trend with an increasing number of protons.

FIGURE 15.8 This graph shows which nuclei are stable (black) and which are unstable (other).

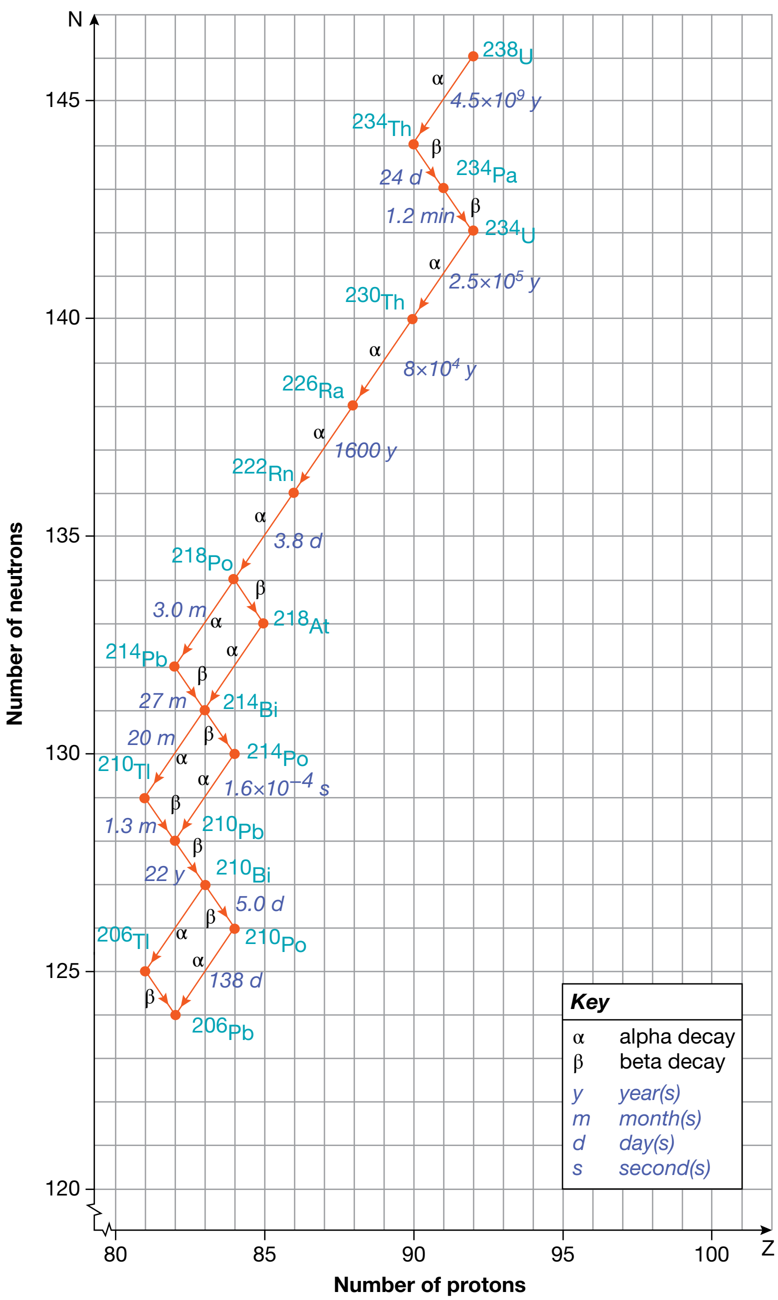


15.4.3 Radioactive decay series

In its 'quest' to become stable, an isotope may have to pass through many stages. As a radioactive isotope decays, the daughter nucleus is often radioactive itself. When this isotope decays, the resulting nucleus may also be radioactive. This sequence of radioisotopes is called a **decay chain** or decay series.

Uranium-238 undergoes 14 radioactive decays before it finally becomes the stable isotope lead-206. Two other decay chains, one starting with uranium-235 and another with thorium-232, also end with a stable isotope of lead. Another decay chain once passed through uranium-233, but this chain is almost extinct in nature now due to its shorter half-lives.

FIGURE 15.9 Radioactive decay series of uranium-238. The half-life is given beside each decay.



15.4 SAMPLE PROBLEM 1

The decay series starting with uranium-238 proceeds by alpha decay and beta decay until the stable isotope of lead-206 is reached.

(a) How many alpha decays are involved in this series?

(b) How many beta decays are involved in this series?

SOLUTION:

(a) As alpha decay is the only decay that reduces the mass number, and each alpha decay causes a decrease in mass number of four, there must be $\frac{238 - 206}{4} = 8$ alpha decays.

(b) The atomic number of uranium is 92 and that of lead is 82. As there are eight alpha decays, these would reduce the atomic number by 16. However, it decreased only by 10. Hence, there must have been six beta decays, each one increasing the atomic number by one

15.4 Exercise 1

1 Referring to Figure 15.8, identify each of the following nuclides, whether or not they are stable, which nuclear decay will bring them closer to stability and explain your reasoning:

$^{28}_{14}X$, $^{42}_{14}X$, $^{56}_{28}X$, $^{78}_{28}X$, $^{100}_{50}X$, $^{222}_{90}X$

2 Referring to Figure 15.9, write out the three successive nuclear decays that lead from Pb-210 to Pb-206.

15.5 Mass defect and binding energy of the nucleus

15.5.1 Mass defect

Most nuclear reactions involve an energy change — either an input (endothermic) or an output (exothermic) of energy. Some reactions, such as those used in nuclear bombs and nuclear power reactors, can release huge amounts of energy.

The key to the large energy involved in nuclear reactions is the fact that mass and energy are equivalent and are linked by Einstein's relationship, $E = mc^2$. The other important fact is that the mass of any nucleus *is not the sum* of the masses of its constituent protons and neutrons. The difference between the mass of a nucleus and the total mass of its constituent nucleons is called the **mass defect** of the nucleus. Rather than define mass in kilograms, it is usual to use atomic mass units for the masses of nuclei. The conversion factor is:

1 atomic mass unit, $u = 1.661 \times 10^{-27}$ kg.

The masses of protons, neutrons and electrons in atomic mass units are:

mass of a proton, $m_p = 1.007\,276$ u
mass of a neutron, $m_n = 1.008\,665$ u
mass of an electron, $m_e = 0.000\,548\,580$ u.

The mass of a deuterium atom, an atom of the isotope of hydrogen, with a neutron as well as a proton in its nucleus, is 2.014 102 u. Therefore, the mass of a deuterium nucleus is $2.014\,102 - 0.000\,549 = 2.013\,553$ u (the mass of the atom − the mass of the electron).

The total mass of an isolated proton and an isolated neutron would be $1.007\,276 + 1.008\,665 = 2.015\,941$ u.

If this proton and neutron combined to form a deuterium nucleus, they would have to lose 2.015 941 − 2.013 553 = 0.002 388 u.

The mass defect of deuterium is 0.002 388 u and if a proton and a neutron combined, energy equivalent to a mass of 0.002 388 u would be released.

If more nucleons could be added to build bigger nuclei, energy would be released and the total mass defect would increase.

15.5.2 Binding energy

If we now tried to do just the opposite, that is, to split our deuterium nucleus into an isolated proton and neutron, we would find that it was not possible. There would not be sufficient mass for an isolated proton and neutron to exist. If we really wanted to accomplish the separation, we would have to provide the missing mass, the mass defect of the deuterium nucleus. Somehow, we would have to supply energy to the deuterium nucleus and have that energy converted into mass. The exact amount of energy that would have to be converted into mass would be the energy equivalent of the mass defect. We call this energy the **binding energy** of the deuterium nucleus.

It is possible to convert the mass defect in atomic mass units to a mass in kilograms and then use $E = mc^2$ to find the energy in joules that would be released. This energy in joules can then be converted to an energy in MeV. However, it is much easier to use the standard conversion factor where the energy equivalent of a mass of 1 u is 931.5 MeV. On data sheets this is stated as $1\ \text{u} = 931.5\ \text{MeV}\ \text{c}^{-2}$.

The mass defect of the deuterium nucleus was 0.002 388 u. The equivalent energy is 0.002 388 × 931.5 = 2.224 MeV. The binding energy of a deuterium nucleus is 2.224 MeV.

If a proton and neutron combined to form a deuterium nucleus, 2.224 MeV of energy would be released:

$$^{1}_{1}\text{p} + {}^{1}_{0}\text{n} \rightarrow {}^{2}_{1}\text{H} + 2.224\ \text{MeV}.$$

If we wanted to split a deuterium nucleus into an isolated proton and neutron, we would have to supply 2.224 MeV of energy that could be converted into mass.

Since binding energy must be supplied to a nucleus to separate it into its constituent parts, this missing energy gives stability to a nucleus. The measure that is most closely related to nuclear stability is **average binding energy per nucleon**.

This gives a measure of how strongly an average nucleon is bound to a particular nucleus. The graph of average binding energy per nucleon against mass number (Figure 15.10) shows that the most stable nuclei have a mass number of about 50 to 60. The most stable nucleus is, in fact, iron-56.

We can see from Figure 15.10 that if we were able to join together light nuclei, we would produce nuclei with a higher average binding energy per nucleon and, hence, energy would be released. This is the process of nuclear fusion. Some atomic masses for light nuclides are given in Table 15.2.

Also, if we were able to take a large mass number nucleus and split it in two, we would produce two nuclei with higher average binding energy per nucleon than the original nucleus. Again, energy would be released. This is the process of nuclear fission.

FIGURE 15.10 A graph of average binding energy per nucleon plotted against mass number.

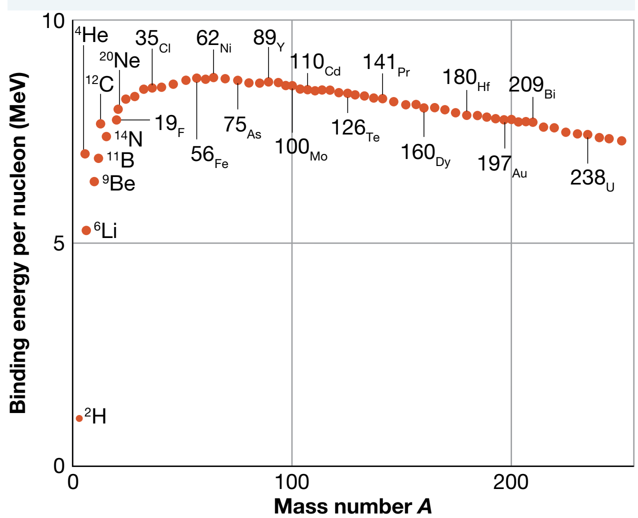


TABLE 15.2 Atomic masses for some light nuclides.

Element and isotope	Neutron number, N	Atomic number, Z	Atomic mass (u)	Mass number, A
Hydrogen ($^{1}_{1}H$)	0	1	1.007 825	1
Deuterium ($^{2}_{1}H$)	1	1	2.014 102	2
Tritium ($^{3}_{1}H$)	2	1	3.016 049	3
Helium ($^{3}_{2}He$)	1	2	3.016 029	3
Helium ($^{4}_{2}He$)	2	2	4.002 603	4
Lithium ($^{6}_{3}Li$)	3	3	6.015 121	6
Lithium ($^{7}_{3}Li$)	4	3	7.016 003	7
Beryllium ($^{9}_{4}Be$)	5	4	9.012 182	9
Boron ($^{10}_{5}B$)	5	5	10.012 937	10
Boron ($^{11}_{5}B$)	6	5	11.009 305	11
Carbon ($^{12}_{6}C$)	6	6	12.000 000	12
Carbon ($^{13}_{6}C$)	7	6	13.003 355	13
Nitrogen ($^{14}_{7}N$)	7	7	14.003 074	14
Nitrogen ($^{15}_{7}N$)	8	7	15.000 109	15
Oxygen ($^{16}_{8}O$)	8	8	15.994 915	16
Oxygen ($^{17}_{8}O$)	9	8	16.999 131	17
Oxygen ($^{18}_{8}O$)	10	8	17.999 160	18

15.5 SAMPLE PROBLEM 1

The mass of a helium atom is 4.002 603 u.

(a) Calculate the mass defect of the helium nucleus.
(b) Calculate the total binding energy of the helium nucleus.
(c) Calculate the average binding energy per nucleon of helium.

SOLUTION:

(a) The total mass of the constituents of a helium atom (two protons, two neutrons and two electrons) is:

$$2\,(1.007\,276 + 1.008\,665 + 0.000\,549) = 4.032\,980\ \text{u}$$

$$\begin{aligned}\text{Mass defect} &= 4.032\,980 - 4.002\,603\\ &= 0.030\,377\ \text{u}\end{aligned}$$

(b)
$$\begin{aligned}\text{Binding energy} &= \text{Mass defect} \times 931.5\\ &= 28.30\ \text{MeV}\end{aligned}$$

(c)
$$\begin{aligned}\text{Average binding energy per nucleon} &= \frac{28.30}{4}\\ &= 7.08\ \text{MeV per nucleon}\end{aligned}$$

Energy change in nuclear reactions

When we considered the formation of a deuterium nucleus from a proton and neutron, we were really dealing with the nuclear reaction:

$${}_{1}^{1}\text{H} + {}_{0}^{1}\text{n} \rightarrow {}_{1}^{2}\text{H} + \text{energy.}$$

We know that energy was released because the product nucleus had less mass than the two reacting nucleons.

We can treat any nuclear reaction in the same way. If the mass of the products is less than the mass of the reacting nuclei, energy will be released. The energy released will be the energy equivalent to the decrease in mass.

15.5 SAMPLE PROBLEM 2

A possible fission reaction for uranium-235 is given below. Find the energy (in MeV) released when one uranium-235 nucleus undergoes such a fission.

$${}_{92}^{235}\text{U} + {}_{0}^{1}\text{n} \rightarrow {}_{57}^{139}\text{La} + {}_{42}^{95}\text{Mo} + 2\left({}_{0}^{1}\text{n}\right) + 7\left({}_{-1}^{0}\text{e}\right)$$

Atomic masses:

$${}^{139}\text{La} = 138.8061\ \text{u}$$

$${}^{95}\text{Mo} = 94.9057\ \text{u}$$

$${}^{235}\text{U} = 235.0439\ \text{u}$$

SOLUTION:

$$\begin{aligned}\text{Total mass of products} &= 235.0439 + 1.008\,665\\ &= 236.0526\ \text{u (to four decimal places)}\end{aligned}$$

$$\begin{aligned}\text{Total mass of products} &= 138.8061 + 94.9057 + 2 \times 1.008\,665 + 7 \times 0.000\,549\\ &= 235.7330\ \text{u (to four decimal places)}\end{aligned}$$

$$\begin{aligned}\text{Decrease in mass} &= 236.0526 - 235.7330\\ &= 0.3196\ \text{u}\end{aligned}$$

$$\begin{aligned}\text{Energy released} &= 0.3196 \times 931.5\\ &= 297.7\ \text{MeV}\end{aligned}$$

15.5 Exercise 1

1 Calculate the binding energy of ${}_{6}^{12}\text{C}$, which has a nuclear mass of 11.996 706 amu.
Data: proton mass = 1.007 276 amu; neutron mass = 1.008 665 amu

2 The isotope ${}_{7}^{12}\text{N}$ has a nuclear mass of 12.014 770 amu. Calculate its:
(a) mass defect
(b) binding energy
(c) binding energy per nucleon.

3 Element X has an atomic number of 29, a mass number of 64 and a nuclear mass of 63.913 843 amu.
(a) What is the identity of element X?
(b) Calculate its binding energy.

eBook*plus* RESOURCES

Try out this interactivity: Making nuclei
Searchlight ID: int-6393

15.6 Spontaneous transmutations

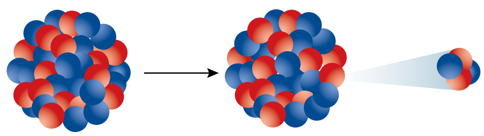
15.6.1 Alpha decay

Spontaneous transmutations occur naturally, such as alpha and beta decay. We will look closer at each of these reactions now to see how the mass and energy changes.

During alpha decay, an unstable nucleus ejects an alpha particle, consisting of two protons and two neutrons (much the same as a helium nucleus). In doing so, the parent nucleus becomes a different, more stable, daughter nucleus.

The number of protons in the nucleus determines the elemental name of the atom. The daughter nucleus is therefore of a different element. For example, uranium-238 decays by emitting an α particle. The uranium-238 atom contains 92 protons and 146 neutrons ($238 - 92 = 146$). It emits an α particle, with two protons and two neutrons. The original nucleus is left with four less nucleons: 90 protons ($92 - 2 = 90$) and 144 neutrons ($146 - 2 = 144$). As the daughter nucleus now has 90 protons, it is called thorium and has the symbol Th. This particular isotope of thorium has 234 nucleons (90 protons and 144 neutrons) and is more correctly written as thorium-234.

FIGURE 15.11 An alpha decay.



The information in the previous paragraph can be written much more effectively in symbols. This is called the **decay equation**:

$${}^{238}_{92}\mathrm{U} \rightarrow {}^{234}_{90}\mathrm{Th} + {}^{4}_{2}\mathrm{He} + \text{energy}$$

$$\textit{or}\ {}^{238}_{92}\mathrm{U} \rightarrow {}^{234}_{90}\mathrm{Th} + \alpha + \text{energy}$$

The ejected α particle is relatively slow and heavy compared with other forms of nuclear radiation. The particle travels at 5–7% of the speed of light: roughly 2×10^7 metres each second. Every object that moves has a form of kinetic energy, or energy of motion. Because the α particle is moving, it has kinetic energy. That energy is written into the decay equation.

The energy can be calculated by first finding the mass lost during the decay. That mass can then be converted into energy.

$$\text{Mass of } {}^{238}_{92}\mathrm{U} = 238.050\,79\ \mathrm{u} = \text{initial mass}$$

$$\text{Mass of } {}^{234}_{90}\mathrm{Th} = 234.043\,63\ \mathrm{u}$$

$$\text{Mass of } {}^{4}_{2}\mathrm{He} = 4.002\,60\ \mathrm{u}$$

$$\therefore \text{final mass} = 238.046\,23\ \mathrm{u}$$

$$\text{Mass defect} = 238.050\,79 - 238.046\,23 = 0.004\,56\ \mathrm{u}$$

$$\therefore \text{energy released} = 0.004\,56 \times 931.5 = 4.25\ \mathrm{MeV}$$

Note that the masses provided are *atomic* masses and so the electrons in the electron shells are included. In this calculation, it is not necessary to take these into account as they automatically cancel out.

15.6.2 Beta decay

Two types of β decay are possible. The β^- particle is a fast-moving electron that is ejected from an unstable nucleus. The β^+ particle is a positively charged particle with the same mass as an electron, and is called a **positron**. Positrons are mostly produced in the atmosphere by cosmic radiation, but some nuclei do decay by β^+ emission.

In β^- decay, an electron is emitted from inside the nucleus. Since nuclei do not contain any electrons, this might seem strange, but it is in fact true! There is no change whatsoever to the electrons in the shells surrounding the nucleus.

Some very interesting changes take place inside a nucleus to enable it to emit an electron. One of the neutrons in the nucleus transforms into a proton and an electron. The proton remains in the nucleus and the electron is emitted and called a β^- particle:

$${}^{1}_{0}\text{n} \rightarrow {}^{1}_{1}\text{p} + {}^{0}_{-1}\text{e}$$

The resulting daughter nucleus has the same number of nucleons as the parent, but one less neutron and one more proton.

An example of β^- decay is the decay of thorium-234. This isotope is the result of the α decay of uranium-238. The nucleus is more stable than it was before the emission of the α particle, but could become more stable by emitting a β^- particle. During this second decay, the mass number of the nucleus is unchanged (234). The number of protons, however, increases by one when a neutron changes into a proton and an electron. There are now 91 (90 + 1) protons in the nucleus, so the atom must be called protactinium-234. The decay equation is written as:

$${}^{234}_{90}\text{Th} \rightarrow {}^{234}_{91}\text{Pa} + \beta^- + \text{energy}$$

or $${}^{234}_{90}\text{Th} \rightarrow {}^{234}_{91}\text{Pa} + {}^{0}_{-1}\text{e} + \text{energy.}$$

In β^+ decay, the positron is also emitted from inside the nucleus. In this case, strange as it may seem, the proton changes into a neutron and a positron with the neutron staying in the nucleus.

$${}^{1}_{1}\text{p} \rightarrow {}^{1}_{0}\text{n} + {}^{0}_{+1}\text{e} + \text{energy}$$

The resulting nucleus has one less proton, but the same number of nucleons. An example of β^+ decay is sodium-22 decaying to neon-22:

$${}^{22}_{11}\text{Na} \rightarrow {}^{22}_{10}\text{Ne} + {}^{0}_{+1}\text{e} + \text{energy}$$

Beta particles are very light when compared to alpha particles. They travel at a large range of speeds — from that of an alpha particle up to 99 per cent of the speed of light. Just like α particles, β particles are deflected by electric and magnetic fields.

15.6 SAMPLE PROBLEM 1

Write down the complete decay equation in each case:

$${}^{234}_{92}\text{U} \rightarrow {}^{230}_{90}\text{Th} + ? + \text{energy} \qquad [1]$$

$${}^{210}_{82}\text{Pb} \rightarrow {}^{210}_{83}\text{Bi} + ? + \text{energy} \qquad [2]$$

$${}^{11}_{6}\text{C} \rightarrow {}^{11}_{5}\text{B} + ? + \text{energy} \qquad [3]$$

SOLUTION:

In [1] the number of particles in the nucleus has decreased by 4, while the number of protons has decreased by 2. This implies that an α particle, or helium nucleus, has been released. The full equation is:

$${}^{234}_{92}\text{U} \rightarrow {}^{230}_{90}\text{Th} + {}^{4}_{2}\text{He} + \text{energy.}$$

Equation [2] cannot show an α emission, as the mass number remains constant. The atomic number has increased, indicating that a proton has been formed, and therefore β^- decay has occurred. The equation becomes:

$${}^{210}_{82}\text{Pb} \rightarrow {}^{210}_{83}\text{Bi} + {}^{0}_{-1}\text{e} + \text{energy.}$$

In equation [3], the mass number stays the same, but there is one less proton, so it must be β^+ decay. The equation becomes:

$${}^{11}_{6}\text{C} \rightarrow {}^{11}_{5}\text{B} + {}^{0}_{+1}\text{e} + \text{energy.}$$

15.6.3 Neutrinos

Alpha particles are emitted with particular energies that are unique to the host nucleus, whereas beta particles are emitted with any energy up to a maximum. When examining decay reactions, scientists found that not all of the energy was accounted for. The possible explanations were:

- the Law of Conservation of Energy, one of the foundations of physics, did not apply in nuclear processes
- a second particle, as yet undetected, was emitted. This idea was proposed in 1930 by Wolfgang Pauli, who said the particle must have no charge, as all the charge was accounted for, and have negligible mass.

Enrico Fermi named the particle 'neutrino', from the Italian for 'little neutral one'. Fermi incorporated Pauli's suggestion into a theory of β decay that not only built on Dirac's work but also derived a mathematical relationship between the half-life of a particular decay and the maximum energy of the emitted β particle. This relationship matched the experimental data, which was convincing evidence for the existence of the neutrino, although it was not actually detected until 1956.

The neutrino has the symbol v, which is the Greek letter *nu*. The complete β^- decay process is:

$$^{1}_{0}\text{n} \rightarrow {}^{1}_{1}\text{p} + {}^{0}_{-1}\text{e} + \bar{v}$$

The word 'energy' in β decay equations should be replaced by v in β^+ decay and $\bar{v}$ in β^- decay ($\bar{v}$ represents an antineutrino). For example:

$$^{234}_{90}\text{Th} \rightarrow {}^{234}_{91}\text{Pa} + {}^{0}_{-1}\text{e} + \bar{v}$$

15.6 SAMPLE PROBLEM 2

Calculate the kinetic energy of the beta particle in the beta negative decay of Pb-210 in sample problem 1.

$$^{210}_{82}\text{Pb} \rightarrow {}^{210}_{83}\text{Bi} + {}^{0}_{-1}\text{e} + \bar{v}$$

Data table

Object	Mass (amu)
$^{210}_{82}\text{Pb}$	209.984 19
$^{210}_{83}\text{Bi}$	209.984 12
$^{0}_{-1}\text{e}$	0.000 55

Note that the masses stated for Bi-210 and Pb-210 are *atomic* masses.

SOLUTION:

In this calculation, the electron masses included in the atomic masses will need to be considered.

$$\begin{aligned}
\text{Initial mass} &= \text{mass of } {}^{210}_{82}\text{Pb} - 82 \text{ electrons} = 209.984\,19 - (82 \times 0.000\,55) \\
&= 209.939\,09 \text{ amu} \\
\text{Final mass} &= (\text{mass of Bi-210} - 83 \text{ electrons}) + \text{mass of electron (mass of the neutrino is} \\
&\quad \text{negligible in this problem)} \\
&= \text{mass of } {}^{210}_{83}\text{Bi} - 82 \text{ electrons} \\
&= 209.984\,12 - (82 \times 0.000\,55) = 209.939\,02 \text{ amu} \\
\text{Mass loss} &= 209.939\,09 - 209.939\,02 = 0.000\,07 \text{ amu}
\end{aligned}$$

$$\text{Energy released} = 0.000\,07 \times 931.5 = 0.065 \text{MeV}$$

15.6 SAMPLE PROBLEM 3

Calculate the kinetic energy of the beta particle in the beta positive decay of C-11 in sample problem 1.

$${}^{11}_{6}\text{C} \rightarrow {}^{11}_{5}\text{B} + {}^{0}_{+1}\text{e} + v$$

Data table

Object	Mass (amu)
${}^{11}_{6}\text{C}$	11.011 43
${}^{11}_{5}\text{B}$	11.009 31
${}^{0}_{+1}\text{e}$	0.000 55

Note that the masses stated for C-11 and B-11 are *atomic* masses.

SOLUTION:

As in the last problem, the electron masses included in the atomic masses will need to be considered.

$$\begin{aligned}\text{Initial mass} &= \text{mass of } {}^{11}_{6}\text{C} - 6\,\text{electrons} = 11.011\,43 - (6 \times 0.000\,55)\\ &= 11.008\,13\,\text{amu}\\ \text{Final mass} &= (\text{mass of } {}^{11}_{5}\text{B} - 5\text{ electrons}) + \text{mass of positron (ignore the negligible mass of the neutrino)}\\ &= 11.009\,31 - (4 \times 0.000\,55) = 11.007\,11\,\text{amu}\end{aligned}$$

$$\text{Mass loss} = 11.008\,13 - 11.007\,11 = 0.001\,02\,\text{amu}$$

$$\text{Energy released} = 0.001\,02 \times 931.5 = 0.950\,\text{MeV}$$

15.6 Exercise 1

Data table for these exercises

Object	Mass (amu)
${}^{45}_{20}\text{Ca}$	44.956 19
${}^{30}_{15}\text{P}$	29.978 31
${}^{30}_{14}\text{Si}$	29.973 77
${}^{4}_{2}\text{He}$	4.002 60
${}^{0}_{-1}\text{e}$	0.000 55
${}^{0}_{+1}\text{e}$	0.000 55

1 The isotope of ${}^{45}_{20}\text{Ca}$ decays to form a new isotope and releases a beta particle and an antineutrino in the process. Identify the new isotope.

2 What particle and kinetic energy is released when phosphorus-30 decays to form silicon-30?

15.7 Artificial nuclear transmutations

15.7.1 Artificial nuclear transmutations, including alpha decay and beta decay

Alpha and beta decay are natural examples of nuclear transformations. The numbers of protons and neutrons in the nucleus change during these processes. Artificial nuclear transformations are also possible. These are done either to investigate the structure of the nucleus or to produce specific radioisotopes for use in medicine or industry. The first artificial transformation was made by Ernest Rutherford, who fired alpha particles at nitrogen atoms to produce an isotope of oxygen.

$$^{14}_{7}\text{N} \rightarrow {}^{4}_{2}\text{He} + {}^{17}_{8}\text{O} + {}^{1}_{1}\text{H}$$

This result raised the intriguing possibility of achieving the alchemist's dream of changing lead into gold. Although prohibitively expensive, it appears to be theoretically possible.

The building of particle accelerators in the early 1930s enabled charged particles such as protons and alpha particles to be fired at atoms as well as alpha particles, but with the advantage that their energy could be pre-set. The limitation of both these particles is that since they are positively charged, they have to be travelling at very high speed to overcome the repulsion of the positively charged nucleus. This problem was overcome with the discovery of the neutron in 1932. The neutron, which has no net charge, can enter the nucleus at any speed. Both protons and neutrons are used today to produce radioisotopes. Particle accelerators firing positive ions produce neutron-deficient radioisotopes such as thallium-201 ($t_{\frac{1}{2}} = 73$ days), which is used to show damaged heart tissue, and zinc-65 ($t_{\frac{1}{2}} = 244$ days), which is used as a tracer to monitor the flow of heavy metals in mining effluent. Neutrons from nuclear reactors produce neutron-rich radioisotopes such as iridium-192 ($t_{\frac{1}{2}} = 74$ days), which is used to locate weaknesses in metal pipes, and iodine-131 ($t_{\frac{1}{2}} = 8.0$ days), which is used in the diagnosis and treatment of thyroid conditions.

Other examples of artificial transmutations are the nuclear reactions that have been used to yield huge amounts of energy — nuclear fission (splitting nuclei to form lighter elements) and nuclear fusion (joining nuclei to form heavier elements). Nuclear fission has been used in bombs and power plants. Nuclear fusion involves much greater amounts of energy and has also been used for bombs; however, science has not yet successfully controlled this reaction for power production. A multinational experimental reactor called ITER is being constructed in southern France (see Figure 15.1) to develop this process. Success in this project could solve many of the world's energy problems.

15.7 SAMPLE PROBLEM 1

$^{9}_{4}\text{Be}$ is bombarded with alpha particles to form a new isotope and a neutron. What is the isotope formed?

SOLUTION:

The nuclear reaction takes the form:

$$^{9}_{4}\text{Be} + {}^{4}_{2}\text{He} \rightarrow {}^{A}_{Z}\text{X} + {}^{1}_{0}\text{n}$$

As the sum of the superscripts on the left and the right of the equation must balance, we can find the value of A as follows:

$$9 + 4 = A + 1$$
$$\therefore A = 12$$

The same process can be used to find Z:

$$2 + 4 = Z + 0$$
$$\therefore Z = 6$$

Z is the atomic number and by referring to a periodic table, it can be seen that element 6 is carbon. Hence the new isotope formed is $^{12}_{6}C$.

15.7 Exercise 1

1. What new isotope is formed when a nucleus of the isotope $^{14}_{7}N$ absorbs an alpha particle? Assume no other particles are emitted.
2. Plutonium-239 is a fissile material (able to undergo nuclear fission) and can be 'bred' from U-238, a non-fissile form of uranium that makes up 99.3% of naturally occurring uranium. If placed around a nuclear reactor core, it can absorb neutrons from the core to form U-239. This has a short half-life of 23.5 minutes and decays by beta negative decay to form Np-239, which in turn has a half-life of 2.36 days. The neptunium-239 then undergoes beta negative decay to form Pu-239. An isotope such as U-238 that is capable of forming fissile material is called a 'fertile' material.
   Construct the sequence of three nuclear equations that represents the process described.

15.7.2 Artificial nuclear transmutations: nuclear fission

In 1934, Enrico Fermi investigated the effect of firing neutrons at uranium. The products had half-lives different from that of uranium. He thought that he had made new elements with atomic numbers greater than 92. Others repeating the experiment got different half-lives. In 1939, Otto Hahn and Fritz Strassmann chemically analysed the samples and found barium, which has atomic number 56, indicating that the nuclei had split.

Lise Meitner and Otto Frisch called this process 'fission' and showed that neutrons could also initiate fission in thorium and protactinium. Further chemical analysis revealed a range of possible fission reactions, each with a different combination of **fission fragments** including bromine, molybdenum or rubidium (which have atomic numbers around 40), and antimony, caesium or iodine (which have atomic numbers in the 50s). Cloud chamber photographs showing two heavy nuclei flying off in opposite directions confirmed that fission had occurred. Meitner and Frisch also calculated from typical binding energies that the fission of one uranium-235 nucleus would produce about 200 MeV of energy, mainly as kinetic energy of the fission fragments. This is a huge amount of energy to be released by one nucleus, as can be seen when it is compared to the burning of coal in power plants. Each atom of carbon used in coal burning releases only 10 eV of energy — about 20 million times less than the energy released in the fission of uranium-235.

Also in 1939, Frederic Joliot and his team confirmed that two or three fast neutrons were emitted with each fission reaction. This allowed for the possibility of a chain reaction, which could potentially release enormous amounts of energy.

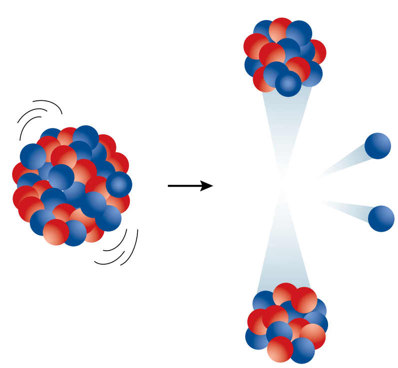
Some possible equations for the fission of uranium-235 set off by the absorption of a neutron are:

$$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{236}_{92}U \rightarrow ^{148}_{57}La + ^{85}_{35}Br + 3^{1}_{0}n + \text{energy}$$

$$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{236}_{92}U \rightarrow ^{141}_{56}Ba + ^{92}_{36}Kr + 3^{1}_{0}n + \text{energy}$$

$$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{236}_{92}U \rightarrow ^{140}_{54}Xe + ^{94}_{38}Sr + 2^{1}_{0}n + \text{energy}$$

FIGURE 15.12 A nuclear fission reaction.



The data in Figure 15.10 on binding energies and Einstein's equation $E = mc^2$ can be used to calculate the amount of energy released in each of the fission reactions above.

Knowing the binding energies of the nuclides involved allows the energy yield of the reactions to be calculated.

FIGURE 15.13 The neutrons produced by a fission reaction can go on to produce a chain reaction.

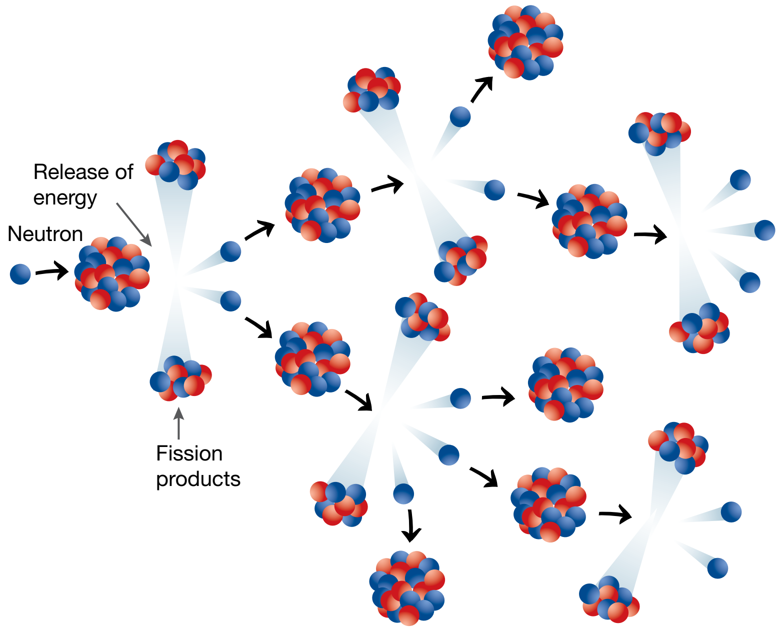
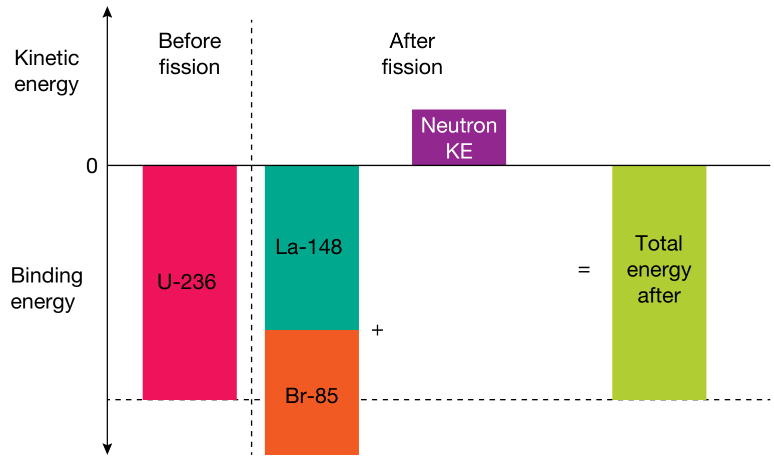


TABLE 15.3 Masses and binding energies of atoms.

Nucleus	Symbol	Mass (kg)	Total binding energy (MeV)
Uranium-236	${}^{236}_{92}U$	$3.919\,629 \times 10^{-25}$	1790.415 039
Lanthanum-148	${}^{148}_{57}La$	$2.456\,472 \times 10^{-25}$	1213.125 122
Bromine-85	${}^{85}_{35}Br$	$1.410\,057 \times 10^{-25}$	737.290 649
Barium-141	${}^{141}_{56}Ba$	$2.339\,939 \times 10^{-25}$	1173.974 609
Krypton-92	${}^{92}_{36}Kr$	$1.526\,470 \times 10^{-25}$	783.185 242
Xenon-140	${}^{140}_{54}Xe$	2.323453×10^{-25}	1160.734 009
Strontium-94	${}^{94}_{38}Sr$	$1.559\,501 \times 10^{-25}$	807.816 711
Neutron	${}^{1}_{0}n$	$1.674\,924 \times 10^{-27}$	

Speed of light, c, = $2.997\,924\,58 \times 10^{8}$ m s^{-1}. 1 MeV=$1.602\,176 \times 10^{-13}$ joules.

FIGURE 15.14 Graph of energies in a fission reaction. The sum of the binding energies of La-148 and Br-85 is greater than the binding energy of U-236. The difference is released as kinetic energy of the neutrons and the fission fragments. The total energy after fission is the same as the energy before.



15.7 SAMPLE PROBLEM 2

Answer the following questions about the fission of uranium-236 producing lanthanum-148 and bromine-85. Use Table 15.3 for data on masses and binding energies.

(a) What is the difference between the binding energy of the uranium-236 nucleus and the sum of the binding energies of the two fission fragments?

(b) What is the difference between the mass of the uranium-236 nucleus and the sum of the masses of all the fission fragments, including neutrons?

(c) Use $E = mc^2$ to calculate the energy equivalent of this mass difference in joules and MeV.

SOLUTION:

(a) From above, we know that the equation for this fission is:

$${}^{236}_{92}\text{U} \rightarrow {}^{148}_{57}\text{La} + {}^{85}_{35}\text{Br} + 3{}^{1}_{0}\text{n} + \text{energy}$$

Binding energy of uranium-236 = 1790.415 039 MeV

Sum of binding energies of fragments = 1213.125 122 + 737.290 649 MeV
= 1950.415 771 MeV

Energy difference = 1950.415 771 − 1790.415 039 MeV
= 160.000 732 MeV

(b) Mass of uranium-236 = $3.919\,629 \times 10^{-25}$ kg

Sum of masses of fragments = $2.456\,472 \times 10^{-25} + 1.410\,057 \times 10^{-25} + 3 \times 1.674\,924 \times 10^{-27}$ kg
$= 3.916\,777 \times 10^{-25}$ kg

Mass difference = $0.002\,852 \times 10^{-25}$ kg

(c) Energy difference in joules = mc^2
$= 0.002\,852 \times 10^{-25} \times (2.997\,924\,58 \times 10^{8})^2$
$= 2.563\,250 \times 10^{-11}$ J

Energy difference in MeV = $2.563\,250 \times 10^{-11} \div 1.602\,176 \times 10^{-13}$
= 159.985 545 MeV

The two answers are effectively identical. The slight difference in the two answers is due to rounding errors, because of the different powers of 10 in the data values.

15.7 Exercise 2

1 It is difficult to know exactly what products will be formed from the fission of uranium as this large nucleus generally breaks into two fragments that are roughly the same size, but not exactly. Does this affect the energy output of the fission reaction? Following are two possible fission reactions of U-235. Calculate mass loss and compare the subsequent energy released by each reaction.

(a) ${}^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{140}_{54}\text{Xe} + {}^{94}_{38}\text{Sr} + 2{}^{1}_{0}\text{n}$

(b) ${}^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{141}_{55}\text{Cs} + {}^{93}_{37}\text{Rb} + 2{}^{1}_{0}\text{n}$

Data table for this exercise

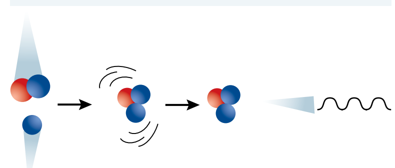
Object	Mass (amu)
${}^{235}_{92}\text{U}$	235.043 93
${}^{140}_{54}\text{Xe}$	139.921 64
${}^{94}_{38}\text{Sr}$	93.915 36
${}^{141}_{55}\text{Cs}$	140.920 04
${}^{93}_{37}\text{Rb}$	92.922 04
${}^{1}_{0}\text{n}$	1.008 67

15.7.3 Artificial nuclear transmutations: nuclear fusion

Nuclear fusion is the process of joining two smaller nuclei together to form a larger more stable nucleus. This was first observed by Australian physicist Mark Oliphant in 1932 when he was working with Ernest Rutherford. He was searching for other isotopes of hydrogen and helium. Heavy hydrogen (one proton and one neutron) was already known, but Oliphant discovered tritium (one proton and two neutrons) and helium-3 (with only one neutron). In his investigation he fired a fast heavy hydrogen nucleus at a heavy hydrogen target to produce a nucleus of tritium plus an extra neutron. This fusion reaction was to become the basis of the hydrogen bomb, but Oliphant was interested only in the structure of the nucleus and did not realise the energy implications. For fusion to occur more extensively, high temperatures and pressures are needed, such as those that exist inside the Sun or in a fission bomb explosion.

The Sun's core has a temperature of more than 15 million K, just perfect for fusion to occur! Inside the Sun, hydrogen nuclei fuse together to form helium. As helium is more stable than hydrogen, the excess nuclear energy is released. This energy is emitted from the nuclei as γ radiation, and is eventually received on Earth as light and heat.

FIGURE 15.15 Nuclear fusion.



Fusion reactions also take place in other stars. Stars that are bigger than the Sun have such severe conditions that larger, more stable nuclei such as silicon and magnesium can be produced from the fusion of smaller nuclei. A star about 30 times more massive than the Sun would be needed to produce conditions that would enable the formation of iron by fusing smaller nuclei.

Our Sun

The chain of events occurring in the Sun is quite complex. The major component of the Sun is ${}^{1}_{1}\text{H}$, that is, nuclei consisting of only one proton and no neutrons. When collisions occur between ${}^{1}_{1}\text{H}$ nuclei, they fuse together in an unusual way. One of the protons is changed into a neutron (in much the same way as a neutron is changed into a proton and an electron during β^- decay). This forms a ${}^{2}_{1}\text{H}$ nucleus. The by-products of this process are a positron and a neutrino.

Positrons are produced when some artificially produced isotopes undergo radioactive decay. Positrons are the opposite of electrons; they have the same mass, but carry a positive charge. When a positron and an electron collide, they immediately annihilate each other. The only thing that remains of either is a gamma ray. Neutrinos are produced when protons change into neutrons and vice versa. They have no charge, are considered massless and travel at close to the speed of light. Fifty trillion neutrinos from the Sun pass through the human body every second.

When a ${}_{1}^{1}\text{H}$ nucleus and a ${}_{1}^{2}\text{H}$ nucleus collide, they form a more stable ${}_{2}^{3}\text{He}$ nucleus, and release the extra energy as a γ ray. If two ${}_{2}^{3}\text{He}$ nuclei collide, they complete the process of turning hydrogen into helium. The collision results in the formation of a ${}_{2}^{4}\text{He}$ nucleus and two ${}_{1}^{1}\text{H}$ nuclei. Again, energy is released. The energy released during nuclear reactions inside the Sun provides energy for life on Earth.

This is the sequence of nuclear equations that occur in the Sun to convert hydrogen to helium:

$${}_{1}^{1}\text{H} + {}_{1}^{1}\text{H} \rightarrow {}_{1}^{2}\text{H} + {}_{1}^{0}\text{e} + v$$

$${}_{1}^{2}\text{H} + {}_{1}^{1}\text{H} \rightarrow {}_{2}^{3}\text{He} + \gamma$$

$${}_{2}^{3}\text{He} + {}_{2}^{3}\text{He} \rightarrow {}_{2}^{4}\text{He} + {}_{1}^{1}\text{H} + {}_{1}^{1}\text{H}$$

15.7 SAMPLE PROBLEM 3

In the final reaction above, two helium-3 nuclei collide to produce a helium-4 nucleus and two hydrogen-1 nuclei, that is, two protons. Use the data in the table below to calculate:

(a) the difference between the binding energy of the helium-4 nucleus and sum of the binding energies of the two helium-3 nuclei

(b) the difference between the sum of masses of the helium-4 nucleus and the two protons, and mass of two helium-3 nuclei

(c) the energy equivalent of this mass difference in joules and MeV.

TABLE 15.4

Particle	Symbol	Mass (kg)	Total binding energy (MeV)
Helium-3	${}_{2}^{3}\text{He}$	$5.022\,664 \times 10^{-27}$	7.718 058
Helium-4	${}_{2}^{4}\text{He}$	$6.665\,892 \times 10^{-27}$	28.295 673
Proton	${}_{1}^{1}\text{p}$ or ${}_{1}^{1}\text{H}$	$1.678\,256 \times 10^{-27}$	

SOLUTION:

(a) Binding energy of helium-4 nucleus $= 28.295\,673$ MeV

Binding energy of two helium-3 nuclei $= 2 \times 7.718\,058 = 15.436\,116$ MeV

Difference $= 28.295\,673 - 15.436\,116 = 12.859\,557$ MeV

(b) $$\text{Mass before fusion} = 2 \times 5.022\,664 \times 10^{-27} = 10.045\,328 \times 10^{-27}\ \text{kg}$$

$$\begin{aligned}\text{Mass after fusion} &= 6.665\,892 \times 10^{-27} + (2 \times 1.678\,256 \times 10^{-27})\\ &= 10.022\,404 \times 10^{-27}\ \text{kg}\end{aligned}$$

$$\text{Mass difference} = 0.022\,924 \times 10^{-27}\ \text{kg}$$

(c) $$\begin{aligned}\text{Energy equivalent (in joules)} &= mc^2\\ &= 0.022\,924 \times 10^{-27} \times (2.997\,924\,58 \times 10^{8})^2\\ &= 2.060\,306 \times 10^{-12}\ \text{joules}\end{aligned}$$

$$\begin{aligned}\text{Energy equivalent (in MeV)} &= 2.060\,306 \times 10^{-12}\,\text{J} \div 1.602\,176 \times 10^{-13}\,(\text{MeV J}^{-1})\\ &= 12.859\,426\ \text{MeV}\end{aligned}$$

15.7 Exercise 3

1 It has been claimed that the fusion of 1 kg of deuterium could keep a 100 W lamp burning for 30 000 years. We shall test this claim using the following reaction:

$${}^{2}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + {}^{1}_{0}\mathrm{n}$$

(a) Calculate the mass of the reactants.
(b) Calculate the mass of the products.
(c) Determine the mass loss in kg.
(d) Convert this mass loss to energy in J.
(e) Use this information to determine the energy yield from the fusion of 1 kg of deuterium.
(f) Identify the energy in J required to run a 100 W lamp for 1 s.
(g) Calculate the energy in J required to run a 100 W lamp for 30 000 years.
(h) Compare your answers to parts (e) and (g). Is there any truth to the claim?

Data table for this exercise

Object	Mass (kg)
${}^{2}_{1}\mathrm{H}$	$3.344\,49 \times 10^{-27}$
${}^{3}_{2}\mathrm{He}$	$5.008\,23 \times 10^{-27}$
${}^{1}_{0}\mathrm{n}$	$1.674\,93 \times 10^{-27}$

2 The process of nuclear fusion occurring in the Sun involves a number of steps but can be summarised in the equation

$$4\left({}^{1}_{1}\mathrm{H}\right) \rightarrow {}^{4}_{2}\mathrm{He} + 2\left({}^{0}_{+1}\mathrm{e}\right) + 2\nu$$

How much energy in MeV is released in this process?

Data table for this exercise

Object	Mass (amu)
${}^{1}_{1}\mathrm{H}$	1.007 83
${}^{4}_{2}\mathrm{He}$	4.002 60
${}^{0}_{+1}\mathrm{e}$	0.000 55

15.8 Review

15.8.1 Summary

- Nuclear radiation is emitted from the nucleus of unstable atoms (radioisotopes) that are striving to become more stable.
- There are four types of radiation: α, β^-, β^+ and γ radiation.
- α particles are released during α decay. α particles are slow-moving particles that are equivalent to a helium nucleus and can be represented as ${}^{4}_{2}\mathrm{He}$. After α decay, the mass number of the daughter nucleus is four less than that of the parent nucleus and the atomic number is two less.

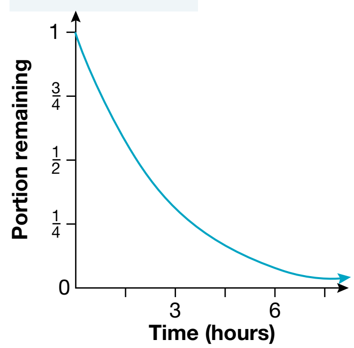
- β^- particles are released in β^- decay. β^- particles are high-speed electrons released from the nucleus when a neutron transforms into a proton and an electron. After β^- decay, the mass number of the daughter nucleus is the same as that of the parent nucleus, but the atomic number is one more than that of the parent nucleus.
- β^+ particles are released in β^+ decay. β^+ particles are high-speed positrons emitted from the nucleus when a proton transforms into a neutron. The atomic number of the daughter nucleus is one less than the parent nucleus; the mass number remains the same.
- γ radiation is electromagnetic radiation that is emitted when an excited nucleus becomes more stable. γ rays are emitted during α and β decay.
- In all nuclear transformations, atomic and mass numbers are conserved.
- Half-life is the time for half of a group of unstable nuclei to decay. It is different for every isotope. The shorter the half-life of an isotope, the greater the activity, that is, the greater the number of decays per second. Activity decreases over time as less and less of the isotope remains. Activity is measured in becquerel (Bq).
- Isotopes may pass through a sequence of decays in order to become stable. Such a sequence is called a decay chain, or decay series.
- The force that holds nucleons together in a nucleus of an atom is called a strong nuclear force. It acts over a very short distance and is strong enough to overcome the electrostatic force of repulsion that exists between the protons of a nucleus.
- The nuclei of different atoms have varying degrees of stability. The binding energy of a nucleus is the energy required to completely separate a nucleus into individual nucleons. Therefore the binding energy is a measure of the stability of a nucleus. Iron is the most stable of all nuclei.
- In order for a nucleus to become more stable, it emits energy called nuclear energy. The amount of energy released is related to the size of the difference between the mass of a nucleus and the mass of the individual nucleons.
- Fusion reactions generally occur between nuclei smaller than iron. Fusion occurs in our Sun, where it converts hydrogen nuclei into helium nuclei and releases large amounts of energy.
- Fission reactions occur when a nucleus is split into smaller, more stable fission fragments.

15.8.2 Questions

1. How many protons and neutrons are in the following atoms?
   (a) $^{66}_{30}\text{Zn}$ (b) $^{230}_{90}\text{Th}$ (c) $^{45}_{20}\text{Ca}$ (d) $^{31}_{14}\text{Si}$
2. Write the symbols for isotopes containing the following nucleons:
   (a) two neutrons and two protons
   (b) seven protons and 13 nucleons
   (c) 91protons and 143 neutrons
3. Write the elemental name and the number of protons and neutrons in each of the following:
   (a) Au-197 (b) $^{210}_{83}\text{Bi}$ (c) $^{210}_{82}\text{Pb}$
4. Explain why it is possible to have two atoms of different elements with the same number of nucleons.
5. From where in an atom are α and β particles and γ rays emitted?
6. In each of the following, determine the type of decay that has occurred:
   (a) $^{14}_{6}\text{C} \rightarrow \, ^{14}_{7}\text{N} + \text{X} + \text{energy}$
   (b) $^{90}_{38}\text{Sr} \rightarrow \, ^{90}_{39}\text{Y} + \text{X} + \text{energy}$
   (c) $^{238}_{92}\text{U} \rightarrow \, ^{234}_{90}\text{Th} + \text{X} + \text{energy}$
7. Write a decay equation to show the α decay of:
   (a) radium-226 (b) polonium-214 (c) americium-241

8. Write a decay equation to show the β^- decay of:
   (a) cobalt-60 (b) strontium-90 (c) phosphorus-32
9. Write a decay equation to show the γ decay of excited magnesium-24.
10. Complete the following decay equations.
   (a) α decay: ${}^{Z}_{A}X \rightarrow {}^{?}_{?}D + ? +$ energy
   (b) β^- decay: ${}^{Z}_{A}X \rightarrow {}^{?}_{?}E + ? +$ energy
   (c) γ decay: ${}^{Z}_{A}X \rightarrow ? + ?$
   (d) ${}^{27}_{13}Al + {}^{2}_{1}H \rightarrow {}^{?}_{?}? + {}^{1}_{0}n$
   (e) ${}^{22}_{11}Na + {}^{4}_{2}He \rightarrow {}^{?}_{?}? + {}^{1}_{1}H$
11. Draw a small decay chain graph, similar to that given for uranium-238 in Figure 15.9, for the β^- decay of yttrium-90. (*Hint:* There is only one decay.)
12. How is β^- decay of a nucleus possible when a nucleus does not contain electrons?
13. How many α particles are released by one atom of uranium-238 as it becomes lead-206? How many β particles are released? (*Hint*: Look at the change in the proton number and the change in the nucleon number.) Check your answer by using Figure 15.9.
14. Repeat question 13 for the following decay series.
   (a) Uranium-235 to lead-207
   (b) Thorium-232 to lead-208
15. In a decay chain, radium-226 emits two α particles, then one β^- particle. What is the element at the end of this sequence and what is its atomic mass?
16. Bismuth-212 has two possible decay modes: an α decay followed by a β^- decay, or a β^+ decay followed by an α decay. The first mode happens about 36% of the time. Will the two modes produce different final results? Explain.
17. What is the half-life of the substance represented in the graph below?

FIGURE 15.16



18. Sketch a decay curve for technetium, which has a half-life of 6 hours.
19. Assume the half-life of carbon-14 is 5730 years. If you had 1 g of carbon-14, how many years would it take for one-eighth of it to remain?
20. The artificial isotope ${}^{15}_{8}O$ is used in medical diagnosis. It has a half-life of 120 seconds. If a doctor requires 1 g of the isotope at exactly 2 pm, how many grams must be delivered to the room 30 minutes earlier? (*Hint*: How much will be needed at 1.58 pm? How much will be needed at 1:56 pm? At 1:54 pm? Can you see a pattern?)
21. Americium-241 is an alpha emitter with a half-life of 432.2 years. It is used in smoke detectors because when the smoke absorbs the α particles, the current drops and the alarm is triggered. The label on the smoke detector says it contains 0.20 micrograms of americium-241 with an activity of 27.0 kBq.

(a) Determine the activity of the americium-241 after 0, 1, 2, 3 and 4 half-lives.
(b) Plot the data and draw a smooth graph (assuming the half-life is 400 years).
(c) Use your graph to estimate the activity after:
    i. 100 years
    ii. 50 years.
(d) What is the implication of your answers to part (c) for the lifetime of the smoke detector?
(e) Write down the decay equation for americium-241 and do an internet search to determine the decay chain.

22. The activity of a radioactive sample drops from 8.0 kBq to 1.0 kBq in 6.0 hours. What is its half-life?
23. Cobalt-60 has a half-life of 5.3 years. If a sample has an activity of 250 GBq (2.5×10^{11} disintegrations per second), what will the activity be in 21.2 years?
24. Explain how individual nucleons are held together in a nucleus, given that like charges repel.
25. (a) Define the terms 'fusion' and 'fission'.
    (b) Which of these reactions occurs in our Sun?
26. Explain why *splitting* uranium-235 nuclei releases energy, but *joining* hydrogen atoms also releases energy.
27. Use the graph of binding energy per nucleon (see Figure 15.10) to estimate the amount of energy released when a uranium-235 nucleus is split into barium-141 and krypton-92. Think carefully about the number of significant figures in your answer. How well does your answer agree with the measured value of 200 MeV?
28. Why is energy released in the process of fusing two small nuclei together?
29. Neutrons are considered to be ionising radiation. Research how neutrons are able to produce ions.
30. Why are neutrons good at initiating nuclear reactions?
31. In what form does the energy released from a nuclear fusion reaction appear?

Use the following table to help answer questions 32, 33 and 34.

TABLE 15.5

Nucleus	Symbol	Mass(kg)	Total binding energy (MeV)
Plutonium-240	$^{240}_{94}Pu$	$3.986\,187 \times 10^{-25}$	1813.454 956
Strontium-90	$^{90}_{38}Sr$	$1.492\,953 \times 10^{-25}$	782.631 470
Barium-147	$^{147}_{56}Ba$	$2.439\,896 \times 10^{-25}$	1204.158 203
Uranium-234	$^{234}_{92}U$	$3.886\,341 \times 10^{-25}$	1778.572 388
Zirconium-95	$^{95}_{40}Zr$	$1.575\,985 \times 10^{-25}$	821.139 160
Tellurium-136	$^{136}_{52}Te$	$2.257\,006 \times 10^{-25}$	1131.440 918
Neutron	$^{1}_{0}n$	$1.674\,924 \times 10^{-27}$	
Proton	$^{1}_{1}p$ or $^{1}_{1}H$	$1.673\,533 \times 10^{-27}$	
Hydrogen-2	$^{2}_{1}H$ or $^{2}_{1}D$	$3.344\,494 \times 10^{-27}$	2.224 573
Hydrogen-3	$^{3}_{1}H$ or $^{3}_{1}T$	$5.008\,267 \times 10^{-27}$	8.481 821
Helium-4	$^{4}_{2}He$	$6.646\,480 \times 10^{-27}$	28.295 673
Lithium-6	$^{6}_{3}Li$	$9.988\,344 \times 10^{-27}$	31.994 564

32. A plutonium-239 nucleus absorbs a neutron to become plutonium-240, which splits to form strontium-90, barium-147 and 3 neutrons.
    (a) What is the difference between the binding energy of the plutonium-240 nucleus and the sum of the binding energies of the two fission fragments?
    (b) What is the difference between the mass of the plutonium-240 nucleus and the sum of the masses of all the fission fragments, including neutrons?
    (c) Use $E = mc^2$ to calculate the energy equivalent of this mass difference in joules and MeV.

33. A uranium-233 nucleus absorbs a neutron to become uranium-234, which splits to form zirconium-95, tellurium-136 and 3 neutrons.
   (a) What is the difference between the binding energy of the uranium-234 nucleus and the sum of the binding energies of the two fission fragments?
   (b) What is the difference between the mass of the uranium-234 nucleus and the sum of the masses of all the fission fragments, including neutrons?
   (c) Use $E = mc^2$ to calculate the energy equivalent of this mass difference in joules and MeV.
34. A fusion reactor could not feasibly use the same reactions as the Sun. A reactor on Earth would have to use a different reaction, preferably a one-step reaction with only two reactants. Three possible reactions for a terrestrial fusion reactor are displayed below; there are many others.
   (a) ${}^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{0}\text{n}$
   (b) ${}^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{1}\text{H} + {}^{1}_{1}\text{H}$
   (c) ${}^{2}_{1}\text{H} + {}^{6}_{3}\text{Li} \rightarrow {}^{4}_{2}\text{He} + {}^{4}_{2}\text{He}$

   Using data from the table above, calculate:
   i. the difference between the binding energy of the products and the sum of the binding energies of the reactants
   ii. the difference between the sum of masses of the products and of the reactants
   iii. the energy equivalent of this mass difference in joules and MeV.

eBook*plus* RESOURCES

Complete this digital doc: Numerical model of a decay series
Searchlight ID: doc-0045

Complete this digital doc: Investigation: Radioactive decay
Searchlight ID: doc-16179

Complete this digital doc: Investigation: Chain reaction with dominoes
Searchlight ID: doc-17060

PRACTICAL INVESTIGATIONS

Investigation 15.1 Radioactive decay

Aim

In this investigation, the radioactive decay of a source will be analysed. The number of undecayed nuclei will be measured as time elapses. The 'source' is the Saunders 'Magic' Source box, which simulates the decay of a radioactive substance over time. When it is switched on and source B is selected, the display shows 1×10^{20} atoms. When the start button is pressed, the source begins to decay and the display shows the *number of nuclei remaining in their original form*.

Apparatus

- Saunders 'Magic' Source
- 12 V power supply
- 2× electrical leads.

Method

Begin the experiment by starting the source and read out the display every 15 seconds. The display represents the number of nuclei remaining in their original form, that is, the number of nuclei that have NOT yet decayed. Use of the HOLD button on the box enables the reading to be frozen while the box continues on. Continue with the experiment for a total time of 300 seconds. Record your results in a suitable table.

Graph your results with *time* on the *x*-axis and *no. remaining* on the *y*-axis.

Draw a smooth line of best fit through your data points. (It does not have to pass through all points.)

Results

Determine 4 values from your graph of the half-life of the source.

Half-life experiments with short-lived radioisotopes can be done in the classroom with a number of radioactive sources:

- protactinium-234, which is a decay product of uranium. The production of the 'cow' uses uranyl nitrate and a common organic solvent.
- barium-137 from a caesium-137 generator
- thallium-208 from thorium-232 'cow'
- radon-220 from a thorium hydroxide generator.

Investigation 15.2 Chain reaction with dominoes

Aim

Try to model the figures using dominoes.

Apparatus

- dominoes
- stopwatch.

Method

If a sufficient number of dominoes is available, try to measure the amount of time it takes to knock over all the dominoes in a controlled chain reaction and in an uncontrolled chain reaction. What does a longer time indicate, in terms of energy and power, for a real fission chain reaction?

FIGURE 15.17

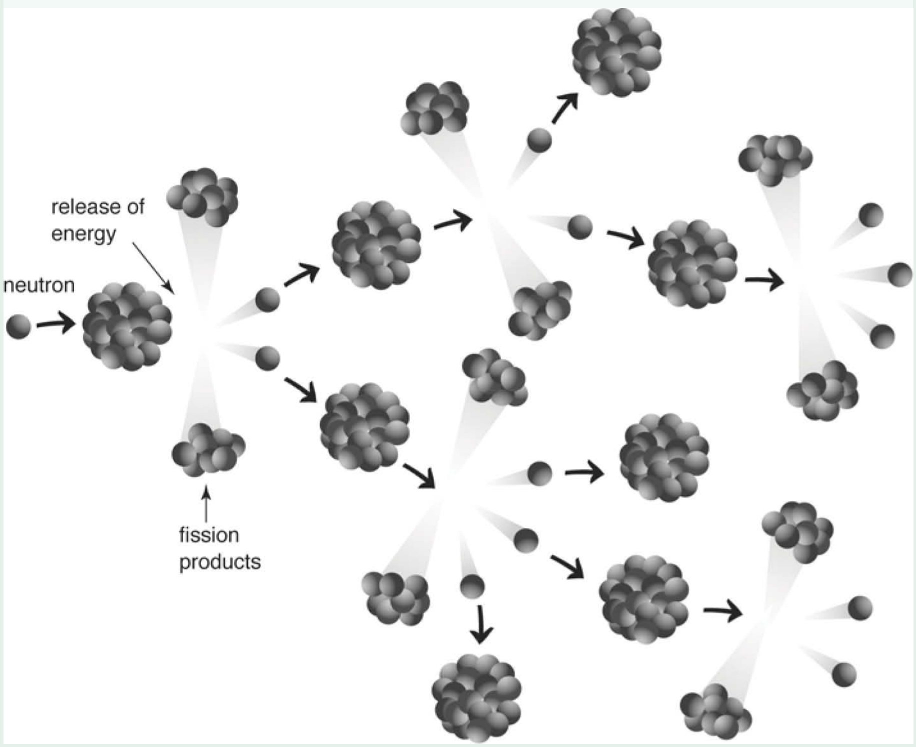
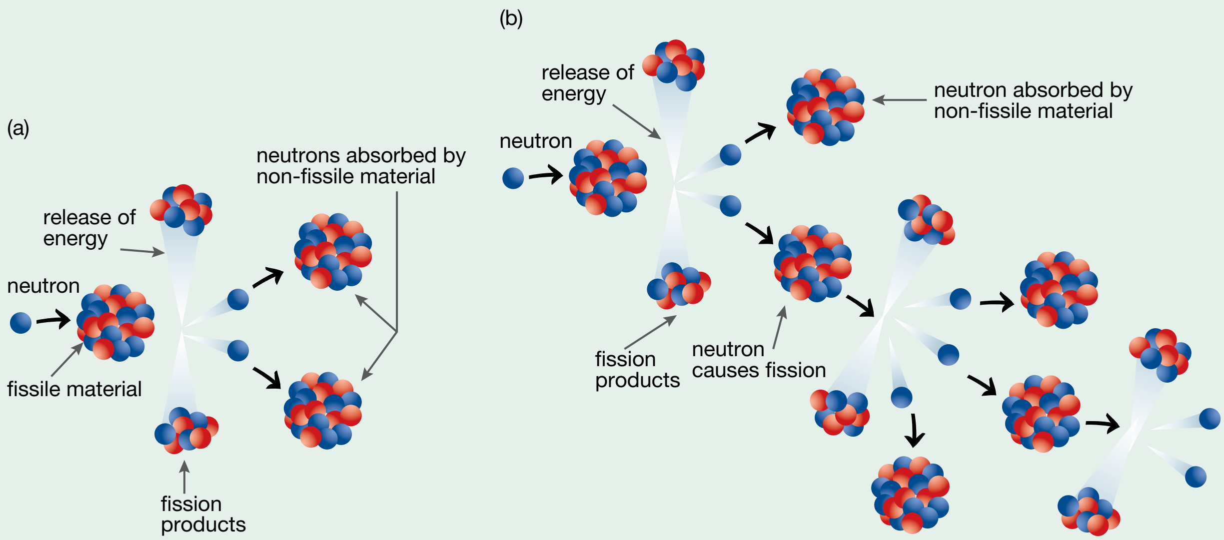


FIGURE 15.18


(a)
neutrons absorbed by non-fissile material
release of energy
neutron
fissile material
fission products
(b)
release of energy
neutron absorbed by non-fissile material
neutron
fission products
neutron causes fission

TOPIC 16 Deep inside the atom

16.1 Overview

16.1.1 Module 8: From the universe to the atom

Deep inside the atom

Inquiry question: How is it known that human understanding of matter is still incomplete?

Students:

- analyse the evidence that suggests:
  - that protons and neutrons are not fundamental particles
  - the existence of subatomic particles other than protons, neutrons and electrons
- investigate the standard model of matter, including:
  - quarks, and the quark composition hadrons
  - leptons
  - fundamental forces (ACSPH141, ACSPH142)
- investigate the operation and role of particle accelerators in obtaining evidence that tests and/or validates aspects of theories, including the Standard Model of Matter (ACSPH120, ACSPH121, ACSPH122, ACSPH146)

FIGURE 16.1 An aerial view of the Fermi National Accelerator Laboratory (Fermilab) at Batavia, Illinois, in the USA. The main accelerator, 6.28 km in circumference, is clearly visible as the circle in the top half of the photograph. The circle in the lower half is the main injector ring. Some of the other accelerator and storage rings are just visible near the main building on the very left. Many important discoveries have been made at Fermilab, with possibly the greatest being the discovery of the top quark in 1995.

16.2 The discovery of subatomic particles

16.2.1 Cosmic rays

Natural radioactivity was first discovered when beta particles exposed photographic plates. One of the other technologies used to investigate radioactivity was the gold leaf electroscope, a device that shows the presence of electric charge. A charged electroscope slowly loses charge due to the ions in the air produced by radioactive elements in the Earth's crust.

It was thought that this effect would decrease with height above the ground. However, in 1909 it was found that the intensity of radiation was greater on top of the Eiffel Tower than at ground level. Subsequent balloon flights revealed that the radiation intensity continued to increase with height, suggesting that it may originate from space. As a result, the term 'cosmic rays' was coined to describe this mysterious radiation.

Initially, cosmic rays were called 'rays' because they were thought to be like light. However, even though they are now known to be particles, the name has stuck. Further investigation over the following decades showed that the particles entering the Earth's atmosphere were mainly protons. The particles seemed to come from beyond the solar system from all points of the sky. Indeed, now they are thought to originate in supernovae and the centres of galaxies.

They are also extremely fast and energetic. The energy of these protons is 40 million times the energy of the protons in the Large Hadron Collider used to produce the Higgs boson. When these protons with their massive energy hit an atom in the upper atmosphere, they cause a cascade of successive collisions that produces a shower of charged particles and gamma rays at the Earth's surface.

FIGURE 16.2 A charged gold leaf electroscope loses charge over time to the airborne ions produced by natural radiation.

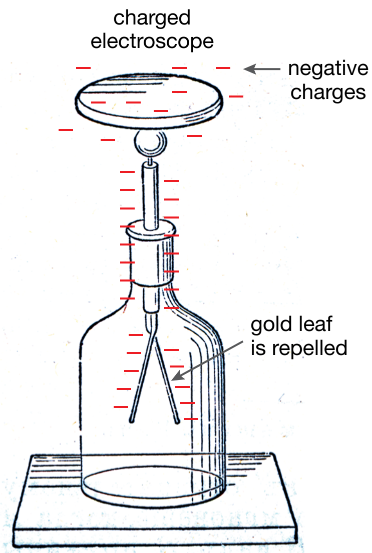
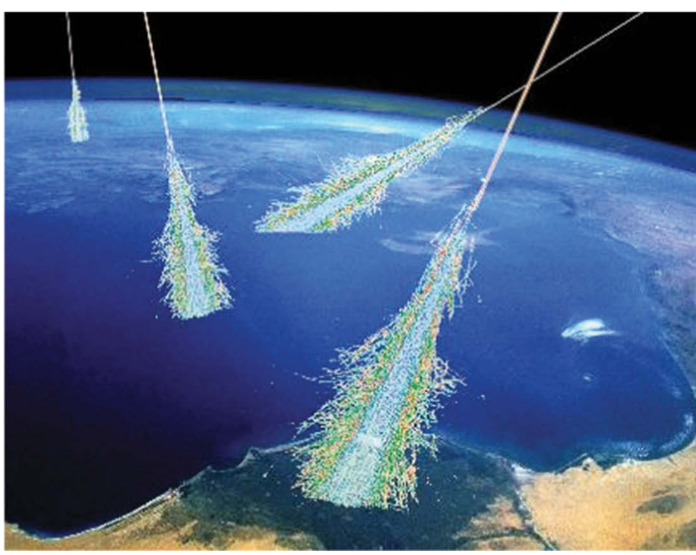
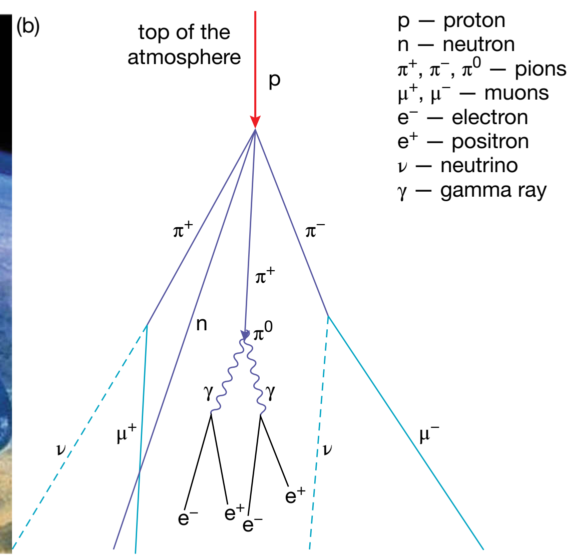


FIGURE 16.3 (a) A cosmic ray shower hitting the Earth's atmosphere (b) The high energy protons entering the upper atmosphere produce showers of charged particles and gamma rays that reach the Earth's surface.

(a)



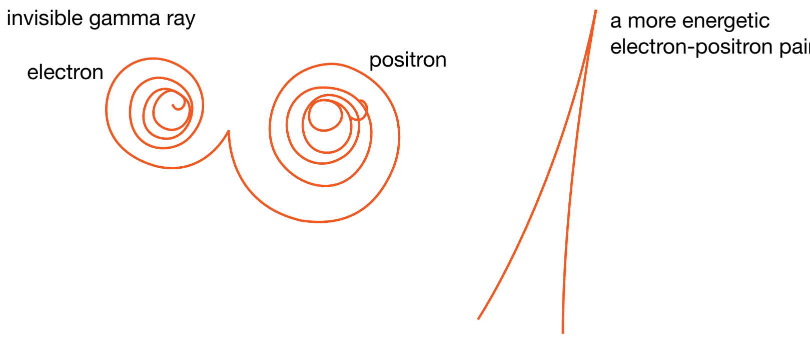
(b)



On average, cosmic rays contribute about 16% of your exposure to ionising radiation from natural sources. This exposure increases the more you fly in a plane and the higher you fly.

In 1933 Carl Anderson was investigating the charged particles in cosmic ray showers and observed a particle that had the same mass as the electron, but with a positive charge. He had discovered a new particle, the positron.

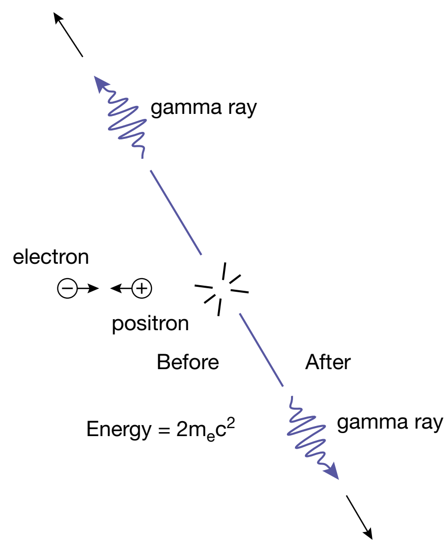
FIGURE 16.4 A gamma ray produces an electron and a positron. The spiralling is due to the slow loss of energy as the track is created.



The chamber Anderson used to detect this charged particle was placed in a strong magnetic field so that a positively charged particle would curve one way and a negatively charged particle would curve the other way. In this experimental set-up, an incoming gamma ray collides with a nucleus; the energy of the gamma ray is converted into mass (recall that $E = mc^2$) but, because charge needs to be conserved, two particles of opposite sign are produced.

The reverse process is also possible. An electron and a positron, or indeed, any particle and its antiparticle, can collide and annihilate each other, producing two gamma rays.

FIGURE 16.5 A particle and antiparticle annihilate each other with their mass, producing two gamma rays.



16.2.2 Explaining the strong nuclear force

Also in the 1930s, Hideki Yukawa was seeking an explanation for the properties of the strong nuclear force that exists between particles inside the nucleus. It was known that this force had a very short range, with each proton or neutron attracted only to its near neighbours, not the whole nucleus. Yukawa suggested that a previously unobserved particle acted as 'glue' between pairs of protons in the nucleus, as well as between other pairings. To fit the known features of the strong force, he determined the properties of this unobserved particle.

He said it should:

- be about 200 times the mass of the electron
- have the same charge size as the electron
- come in two types: positive and negative
- have a very short half-life of about a millionth of a second
- interact very strongly with nuclei.

In 1936 Carl Anderson's group found such a particle in cosmic ray showers. This particle was named the **muon**. However, while the muon satisfied the first four points above, it became apparent that its interaction with nuclei was very weak, so the muon was not a good candidate to explain the strong nuclear force.

A few years later, Cecil Powell investigated cosmic ray showers at high altitude in the Pyrenees and the Andes mountain ranges. These observations were higher up in the cascade of collisions that cosmic rays set off when they hit the atmosphere. At this altitude, Powell found another particle that better fit the needs of the strong nuclear force. This particle is called the π-meson or pion.

Shortly after Powell's discovery, the pion was also detected in the laboratory when carbon nuclei were bombarded with high-energy alpha particles.

After this time, most new particles were found in laboratories using particle accelerators. In the years that followed, even more particles were discovered, so that by the 1970s over 200 subatomic particles had been identified.

16.2.3 So many particles!

In the nineteenth century, chemists had to make sense of the large array of chemical elements. The periodic table was the result, with gaps for yet to be found elements. In the late twentieth century, physicists needed to find some pattern among the particles that would bring a similar form of order to subatomic particles.

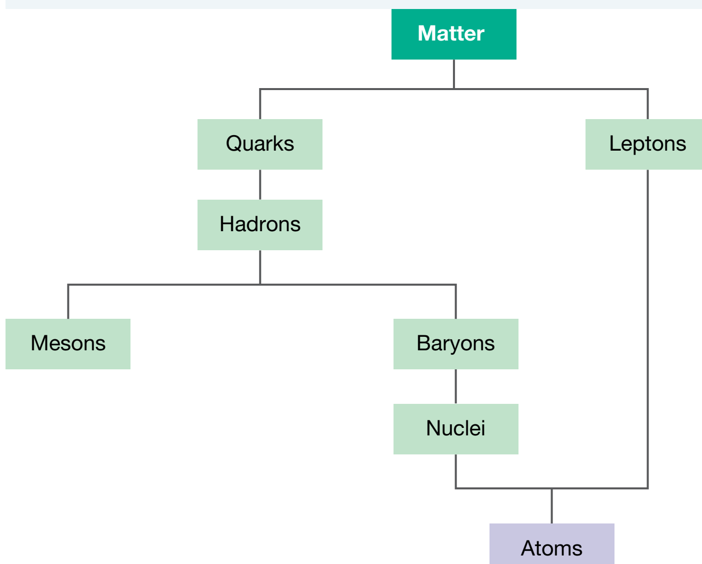
TABLE 16.1 Comparison of the discovery of chemical elements and subatomic particles.

Discovery of elements		Discovery of subatomic particles	
Time	**Progress**	**Time**	**Progress**
Late eighteenth century	About 30 elements known	By 1920	2 known (p and e^-)
Mid nineteenth century	About 60 elements known; Mendeleev produces periodic table with gaps predicting properties of unknown elements	By 1940 By 1950 By 1960	4 more discovered (n, e^+, μ^+, μ^-) 2 more found (π^+, π^-) Several more particles discovered; quark model proposed, predicting new particles
Early twentieth century	92 elements found to fill gaps	By 1970	Predicted particles found

The periodic table initially grouped the elements by common properties, for example, metals and non-metals. Similarly, the subatomic particles can be divided into groups. The two groups are leptons and hadrons, with hadrons made up of two subgroups.

When Enrico Fermi was asked about the names of some of these particles, he made his famous response: 'If I could remember the names of all these particles I would have been a botanist'. While we sympathise with Fermi's view, we have to look at some of the terms that are collectively assigned to different groups of particles.

FIGURE 16.6 The families of subatomic particles and their relationship to matter and atoms.



Leptons

Leptons are the simplest and lightest of the subatomic particles. The different types of leptons are shown in Table 16.2.

TABLE 16.2 Leptons.

Name	Symbol	Mass	Charge	First observed	Half-life
Electron	e^-		Negative	1869	Stable
Muon	μ^-	About 200 × mass of electron	Negative	1936	10^{-6}s
Tau	τ^-	About 277 × mass of electron	Negative	1977	10^{-13}s
Electron neutrino	υ_e	Negligible	Neutral	1956	Stable
Muon neutrino	υ_μ	Negligible	Neutral	1962	Stable
Tau neutrino	υ_τ	Negligible	Neutral	2000	Stable

Leptons are fundamental particles, that is, they have no internal structure, although muons and tau particles decay into electrons. The neutrinos accompany any interaction of their heavier partner.

The electron is found in atoms and determines the chemical properties of elements.

The muon decays to an electron according to the equation:

$$\mu^- \rightarrow e^- + \overline{\upsilon_e} + \upsilon_m$$

The $\overline{v_e}$ particle is an anti-electron neutrino. The bar above the symbol indicates that it is an antiparticle.

Muons surprisingly have industrial uses. They are more penetrating than X-rays and gamma rays, and they are non-ionising, so they are safe for humans, plants and animals. Their better penetrating power means that, for example, they can be used to investigate cargo containers for shielded nuclear material. Muons have also been used to look for hidden chambers in the pyramids. Muon detectors were used at the Fukushima nuclear complex to determine the location and amount of nuclear fuel still inside the reactors that were damaged by the Japanese tsunami in 2011.

The tau particle was discovered some time later than the muon. The unusual feature of this particle is that it decays into two pions, which are discussed later. The decay equation is

$$\tau^- \rightarrow \pi^- + \pi^0 + \upsilon_\tau$$

The negative pion, π^-, then decays into an electron, while the neutral pion, π^0, decays to two gamma rays.

Each of the six leptons has an antiparticle. For the electron, the antiparticle is the positron. The anti-muon and the anti-tau, like the positron, are also positively charged.

Hadrons

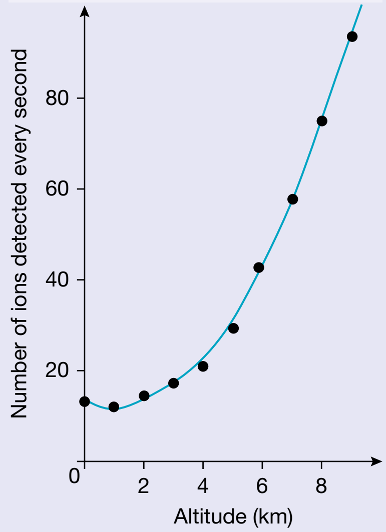
Hadrons are distinctive because they are much heavier than the leptons but, much more importantly, they all have an internal structure. Hadrons are made up of different combinations of particles called quarks. We will learn more about quarks in Section 16.4. Hadrons that are a combination of two quarks are called **mesons**. The other hadrons are combinations of three quarks and are called **baryons**.

There are over 60 different types of mesons, including the pion mentioned earlier. They play a role in nuclear interactions, but have very short half-lives, so they are very difficult to detect. Each meson also has an antiparticle.

Baryons include the proton and neutron as well as about 70 other different particles. Only the proton and neutron are stable, with all other baryons having extremely short half-lives. Each baryon also has its own antiparticle.

16.2 Exercise 1

FIGURE 16.7 The number of charged particles from cosmic rays varies by altitude above the Earth's surface.



1 In the graph shown in Figure 16.7, the number of ions detected initially decreases with height. Then, about 1 km above the Earth's surface, it increases quite rapidly. Suggest a reason why there is a strong reading at the Earth's surface that then decreases with height.

2 Cosmic rays are often described as cosmic ray 'showers'. Why is this word an appropriate description?

3 What aspect of the more energetic pair of tracks do you think indicates that the electron and positron are moving faster?

4 A possible reaction for the formation of the pion from a carbon nucleus and alpha particle is that the alpha particle and carbon nucleus join with the pion being emitted. Complete the nuclear equation below by determining the values of X and Y, and the symbol for the chemical element, Z. Note: the pion has a charge of +1.

$${}^{4}_{2}\text{He} + {}^{12}_{6}\text{C} \rightarrow {}^{X}_{Y}\text{Z} + {}^{0}_{1}\pi$$

5 What is the subatomic particle equivalent of Mendeleev's periodic table?

16.3 The tools of particle physicists

16.3.1 Particle detectors

Many of the most important discoveries in physics made between fifty and one hundred years ago were made, sometimes by accident, by a single physicist working with a very simple apparatus. Enrico Fermi was awarded a Nobel prize for the work completed after he put a rough piece of paraffin in the path of the neutrons that he was using to irradiate a sample. Rutherford and his co-workers used simple scintillation detectors to observe the scattering of alpha particles. This method of detection enabled them to count the alpha particles. However, despite the significant results achieved with such simple apparatus, better detectors that would provide information such as the charge and energy of the particles were required.

The quest for better detectors saw the use of the cloud chamber and the development of the bubble chamber. In recent times, these were superseded by larger and more complex multicomponent detectors. Examples of these are the detectors used at the high-energy accelerator facilities, such as CERN, Fermilab and Brookhaven.

Cloud chambers

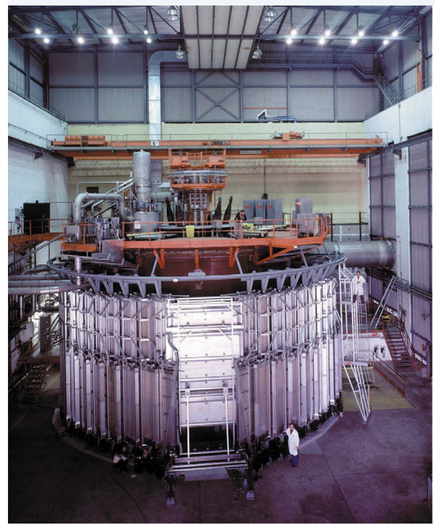
The cloud chamber was invented by C.T.R. Wilson before the end of the nineteenth century, but not used to detect particles until about 1910. It remained in use until about 1960.

A cloud chamber contains a supersaturated vapour. As ionising radiation passes through the vapour, fine droplets of vapour form on the ions produced by the radiation, leaving a visible vapour trail showing the path of the particle. If the chamber is in a magnetic field, the path of a charged particle will be curved, with the direction of the curve indicating the charge of the particle.

Bubble chambers

In 1952, Donald Glaser invented the bubble chamber. It is claimed that the observation of bubbles in glasses of beer played a significant part in the invention.

FIGURE 16.8 The 3.7 m bubble chamber at CERN. Before being dismantled in 1984, this bubble chamber was used for over six million photographs.



The bubble chamber has a similar principle of operation to the cloud chamber except that the bubble chamber contains a superheated liquid (a superheated liquid exists in the liquid state at a temperature above its normal boiling point). Propane and pentane were used in early bubble chambers and hydrogen in later ones. When ionising radiation passes through the liquid, localised boiling occurs on the ions and leaves a trail of bubbles. Bubble chambers were much better detectors than cloud chambers because of the greater density of the substance in the chamber. A 10 cm bubble chamber was approximately equivalent to a 10 m cloud chamber.

Modern detectors

Detectors in use at large nuclear research facilities, such as CERN, Fermilab and Brookhaven, are now larger and more complex than bubble chambers. In typical high-energy experiments performed at these facilities, multicomponent detectors are used to record what may be millions of events and to store them on computer for later analysis.

The function of a detector is to record the trajectory, energy and momentum of the particles produced in a collision 'event'. If two beams of particles, perhaps protons and antiprotons, with similar energies collide head-on, the particles produced could travel in any direction and a large cylindrical detector is used. Such a detector commonly has four different regions — an inner tracking chamber surrounded in turn by an electromagnetic calorimeter, then a hadronic calorimeter and finally a muon chamber.

The inner tracking chamber contains a gas and, as the charged particles produced in the collision event traverse this chamber, they produce ions. The ions may be collected on thin metal wires and produce a small electrical pulse. Once the presence of the ions has been detected, the tracks of the particles that produced the ions can be deduced. Many very short-lived particles do not leave the tracks but they may decay into particles that do.

The calorimeters are made of dense materials that absorb the energy of the particles, interleaved with sensitive detector materials. The different materials are segmented and it is possible to determine where a particle was finally absorbed. The electromagnetic calorimeter is optimised to measure the energy and positions of electrons and photons that interact via the electromagnetic force. The hadronic calorimeter is optimised to measure the energy and positions of hadrons that interact via the strong force.

Only muons (and neutrinos) are able to pass through the two inner calorimeters. Any charged particle that reaches the outermost calorimeter must be a muon. Neutrinos, of course, continue without interacting with any part of the detector.

FIGURE 16.9 A simplified end-on view of a cylindrical detector that might be used for a colliding beam experiment.

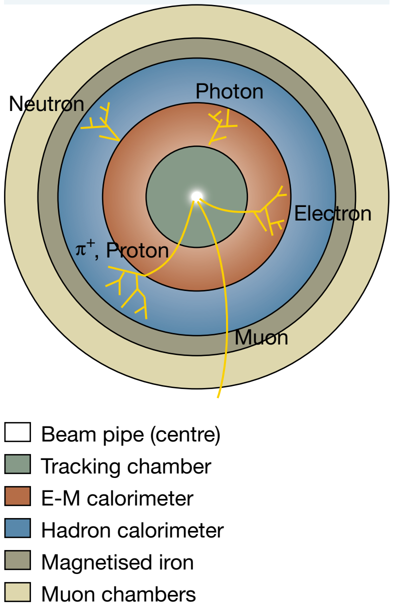
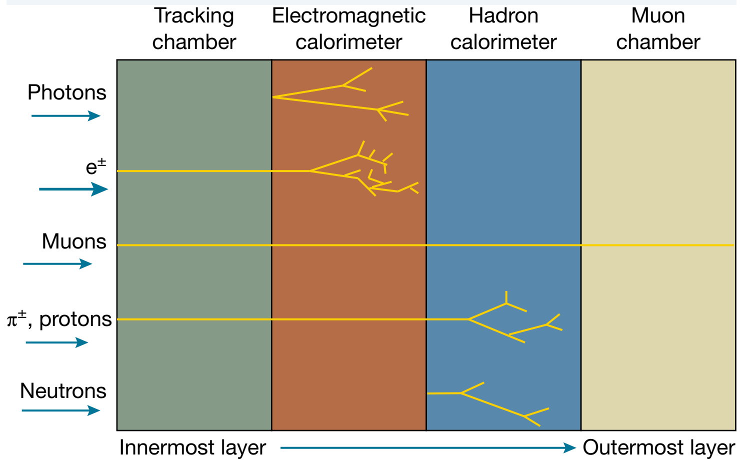


FIGURE 16.10 The passage of particles through the different sections of a multicomponent detector.



16.3.2 Particle accelerators

Early particle physicists used alpha-particle sources that were naturally occurring alpha-particle emitters. Some of these produced alpha particles of much higher energy than others. When alpha particles were used to induce artificial radioactivity, it soon became apparent that particles with even higher energies would be more useful.

The quest for higher energy particles saw the development of a variety of particle accelerators. The higher energy particles from the particle accelerators were used to bombard nuclei and produce a wide variety of new particles.

A very simple accelerator is the electron gun in the tube of a television set. Electrons are accelerated across a large potential difference and then directed at the screen of the television set. Some of the early particle accelerators were similar to this in that they were single-stage electrostatic accelerators. Higher energies were possible when the particles were accelerated many times, such as in a linear accelerator. The development of cyclic accelerators such as the cyclotron saw large energies possible with smaller devices but, as we will see, modern accelerators have become very large and complex devices.

The first particle accelerators

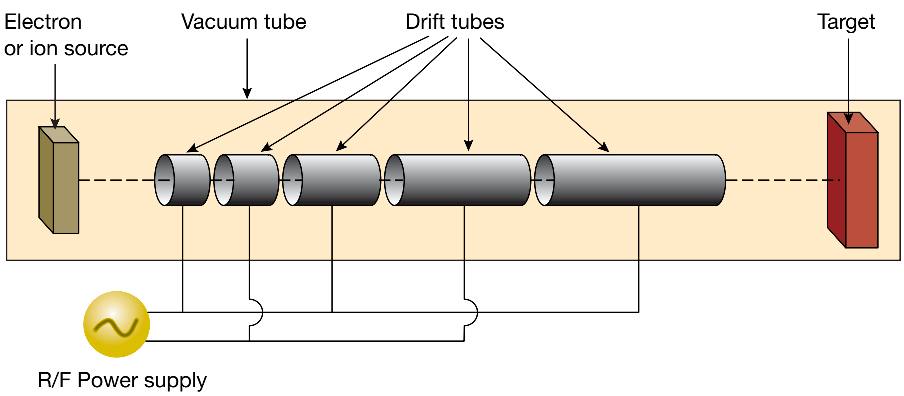
An early particle accelerator that was used to accelerate protons to 770 keV was developed by John D. Cockroft and Ernest T.S. Walton at the Cavendish Laboratory in 1932. It was an electrostatic machine that gave the protons a single high-energy 'kick'. Another electrostatic accelerator is the Van de Graaff generator that you have probably encountered in your science studies in the classroom. Large versions were capable of reaching 1.5 MeV or higher.

These accelerators have since been improved as it became apparent that many more important discoveries could be made with particles of a higher energy, and new accelerators were developed. However, despite these new accelerators, the earlier accelerators are still found to be useful. It is interesting to note that, at Fermilab, the initial step or pre-acceleration is provided by a Cockcroft–Walton accelerator, and, at Brookhaven, the initial acceleration of the Relativistic Heavy Ion Collider (RHIC) is provided by tandem Van de Graaff accelerators.

Linear accelerators

The most famous linear accelerator is at the Stanford Linear Accelerator Center (SLAC). Charged particles are fired through a three-kilometre-long evacuated tube. The charged particles pass through one cylindrical electrode and are then accelerated by an electric field as they pass through a gap before encountering another electrode. This process is repeated and the particles increase their energy. Of course, the alternating accelerating potential has to keep in step with the particles and this requires the cylindrical electrodes to become longer and longer. Eventually it becomes impractical to add extra stages to a linear accelerator. At SLAC, electrons were accelerated to a velocity very close to that of light and had an energy of 20 GeV. After modifications were completed in 1987, the SLAC was able to produce 50 GeV electrons.

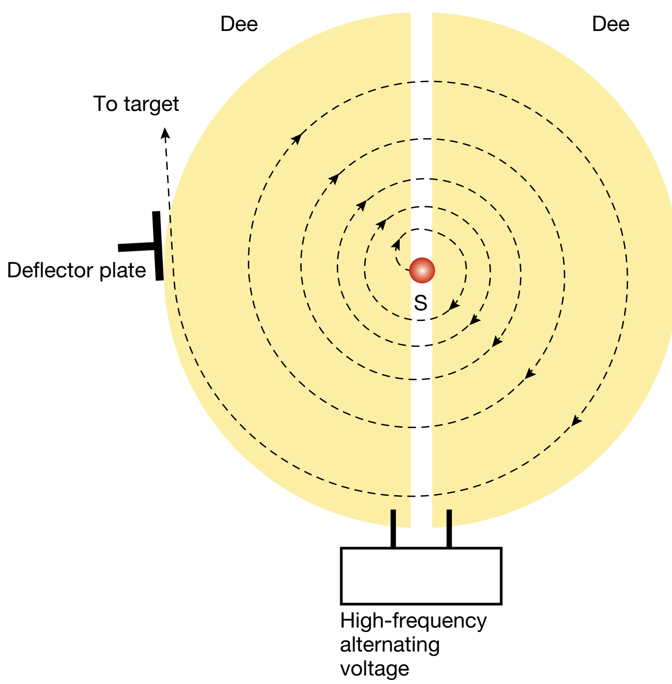
FIGURE 16.11 A linear accelerator.



Cyclotrons

Like a linear accelerator, a cyclotron is able to give a charged particle many 'kicks' as it passes through the electric fields between hollow cylinders of non-magnetic metal (called the 'dees') of the cyclotron. Again, the particles move through an evacuated region. The whole apparatus lies between the poles of a large magnet. Therefore, the particles move in circular paths, with the radii of the paths increasing each time the particle gains energy as it passes through the gap between the dees. When the particles reach the limit of the magnetic field, they are deflected into a target. Very high energy cyclotrons are not possible for a number of reasons. Eventually size would become prohibitive and also, as the particles reached very high velocities, the relativistic increase in mass would mean that the particles would become out of step with the applied alternating potential.

FIGURE 16.12 Top view of a cyclotron. Note that for positively charged particles to travel clockwise as shown, the magnetic field must be directed out of the page.



Synchrotrons

The main accelerators today are synchrotrons. Synchrotrons keep the particles in a path of constant radius. As the particles gain energy, the magnetic field is increased to maintain the same path. Many powerful magnets are required around this path. The particles move through a small-diameter evacuated tube that forms a large-diameter ring.

On each circuit around the ring, the particles pass through regions where an applied radio frequency provides an electric field in a direction such that it produces the 'kick' that increases the energy of the particles. The radio frequency increases as the particles increase in energy and take shorter and shorter periods to complete their orbits. A disadvantage of a synchrotron is that a 'batch' of particles must complete their journey through the accelerator before another batch can enter. However, the advantages of the synchrotron, in terms of energy that can be achieved, far outweigh this disadvantage.

FIGURE 16.13 Part of the 6.28 km (circumference) Fermilab Tevatron accelerator. The yellow and red sections on the floor are parts of the new accelerator, which can accelerate protons to nearly 1000 GeV.

The dimensions and statistics of the large accelerators are impressive. The main accelerator, the Tevatron at Fermilab, has a circumference of 6.28 km. The original accelerator was able to accelerate protons to about 200 GeV. When higher energies were required, another accelerator was built below the first. It uses superconducting magnets to steer the particles around the ring, and because these do not heat up, they can be left on for very long periods. Another accelerator in a separate ring now accelerates protons to 150 GeV, at which point they are transferred to the new accelerator and accelerated further to about 1000 GeV(1 TeV).

The Large Electron Positron collider (LEP) at CERN occupied a tunnel of 27 km circumference straddling the Swiss–French border. In 2000 it was shut down, and construction of its replacement, the Large Hadron Collider (LHC) began. The LHC was scheduled to commence operation in 2005 but that was delayed until 2007. Then in 2007 a problem with one of the magnet support structures caused a further delay until 2008. The LHC has 1232 superconducting magnets, each 15 m long, around 85% of its circumference. The magnets were supplied by Fermilab. These magnets are powered by superconducting cables carrying currents of 12 000 amps and are cooled by liquid helium to −271°C. The LHC is able to accelerate protons to 7 TeV and collide these protons with other protons travelling in the opposite direction, also with an energy of 7 TeV. It is also able to collide heavy ions, such as lead, with a total energy of 1250 TeV, about thirty times that of the RHIC at Brookhaven.

16.3 Exercise 1

1 In 1930, Ernest Lawrence constructed the first cyclotron. It was approximately 12.5 cm in diameter and accelerated hydrogen ions in a magnetic field of 1.27 T between the poles of a magnet 10 cm in diameter. The accelerating voltage applied to the dees was 2000 V and Lawrence determined that the hydrogen ions had been accelerated to an energy of 80 000 eV

(a) How many times had the hydrogen ions experienced the 2000 V accelerating potential and how many orbits of the cyclotron did they complete?

(b) Determine the velocity of an 80 000 eV hydrogen ion (proton).

(c) A charged particle moving in a circular path in a magnetic field has a centripetal force of magnitude $\boldsymbol{F}_c = \frac{m\boldsymbol{v}^2}{r}$ provided by the magnetic force $\boldsymbol{F}_B = q\boldsymbol{vB}$. Determine the radius of an 80 000 eV proton in a magnetic field of 1.22 T.

2 The Stanford Linear Accelerator (SLAC) was used to accelerate electrons to very high velocities. Suggest a reason why a linear accelerator was preferable to a cyclic accelerator, such as a cyclotron, as a device to produce very high energy electrons. (Hint: you may wish to think back to the Rutherford model of the atom and the electrons in orbit around the nucleus.)
3 In 1983, particle physicists in the United States proposed that a new accelerator be constructed. This new accelerator was called the Superconducting Supercollider (SSC) and was to be approximately 86 kilometres in circumference. It was planned to be approximately 60 metres below the ground surrounding the city of Waxahachie, Texas. The total cost was estimated to be eight billion dollars. Construction began in 1990, but the project was cancelled in 1994 when it was about 20% complete. (A search on the internet will yield information about the project and reasons for and against stopping the construction.) Prepare material that could be used in a debate, either to support or to oppose the expenditure of such a large amount of money on such a project.
4 Describe the advantages and disadvantages of a synchrotron.

16.4 The quark model

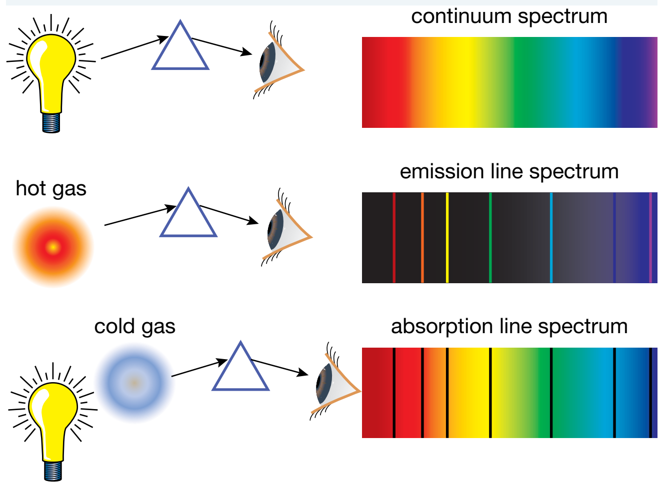
16.4.1 On the trail of the quark

In the late nineteenth century, when visible light was shone through a gas of atoms of a particular element, a spectrum of black lines was observed. Each element produced a unique pattern of these lines, called an absorption pattern.

The lines in the atomic absorption patterns suggested that there was some complexity or structure inside the atom. This structure was discovered early in the twentieth century.

More information about the internal structure of the nucleus can be determined by the energy of alpha particles emitted through radioactive decay. This energy is specific to the nucleus undergoing decay. If a system is showing evidence that it can have only certain energy values, then it must have a structure, that is, be made up of smaller particles.

FIGURE 16.14 A continuous spectrum and two different ways of producing an element's fingerprint.



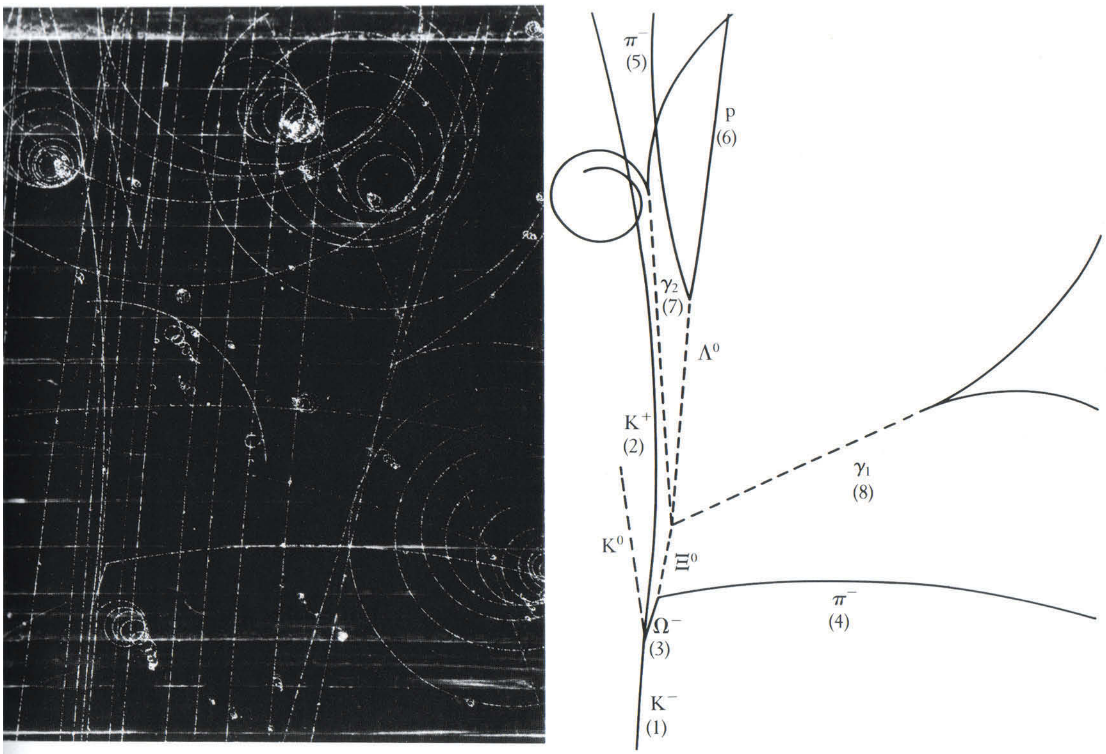
During the 1960s it was discovered that when protons and neutrons were hit by a beam of particles, a type of spectra was evident, much like molecules, atoms and nuclei. This meant that protons and neutrons are made up of even smaller particles. This was the beginning of the quark model.

In 1961, Murray Gell-Mann (1929–) in the USA and Yuval Ne'eman (1925–2006), an Israeli theorist in England, independently discovered a method of organising particles on the basis of their symmetry properties. Gell-Mann called the method, perhaps a little irreverently, by the Buddhist term 'The Eightfold Way'. The theory suggested that there was a missing particle, which they referred to as the Ω^- (omega minus).

In 1963, a search for this particle was started using the bubble chamber at Brookhaven. The bubble chamber was about two metres in diameter and contained liquid hydrogen. Every few seconds, a burst of kaons (K-mesons) collided with protons (the nuclei of the atoms of liquid hydrogen) in the bubble chamber. This produced a spray of particles that, it was hoped, would include the Ω^-. Eventually, in photograph number 50 321, an event indicating the existence of the Ω^- was discovered.

FIGURE 16.15 (a) The bubble chamber photograph taken at Brookhaven National Laboratory that shows the path of the Ω^- particle. (b) The diagram at the right identifies the particles responsible for the various trails. Dotted lines show the paths of particles not visible in the photograph. The incident K^- particle (1) collided with a positron at (3). The short tail at (3) was produced by the Ω^- before it decayed to the π^- and ultimately a number of other particles of which some left trails in the photograph.

(a) (b)



While the finding of the Ω^- particle confirmed the Eightfold Way's organisation of particles, the underlying reason for this organisation remained a mystery until 1964. At this time, Gell-Mann and George Zweig (1937–) independently proposed that there were three fundamental particles that were the constituents of hadrons. Gell-Mann named these particles 'quarks'.

16.4.2 The quark model

Gell-Mann and Zweig first proposed that hadrons were composed of only three quarks. It was predicted that there were three different types of quarks and these were called 'up', 'down' and 'strange'. Later it was necessary to add more quarks and these became 'charm' (discovered in 1974), 'bottom' (1977) and 'top' (1995). In the strange language of particle physics, these types of quarks became known as 'flavours'.

The six quarks have different masses and possess charges that are either $+\frac{2}{3}$ or $-\frac{1}{3}$ of the charge on an electron. Each quark has its own antiparticle.

The characteristics of the known quarks are shown in Table 16.3.

The charges on baryons (made of three quarks) and mesons (made of two quarks) are the result of the charges carried by the quarks of which they are composed. Mesons are composed of one quark and one antiquark. A positive pion (π^+) is made of one up quark and one down antiquark to give a charge of +1 while its antiparticle, π^-, is made of one up antiquark and one down quark to give a total charge of −1. Baryons have three quarks. A proton is composed of two up quarks, giving the proton a charge of +1.

TABLE 16.3 Quark characteristics.

Quark	Symbol	Charge	Multiple of proton mass	First observed
Up	u	$+\frac{2}{3}$	0.003	1968
Down	d	$-\frac{1}{3}$	0.006	1968
Charm	c	$+\frac{2}{3}$	1.3	1974
Strange	s	$-\frac{1}{3}$	0.1	1968
Top	t	$+\frac{2}{3}$	184	1995
Bottom	b	$-\frac{1}{3}$	4.5	1977

A neutron is also composed of up and down quarks, but one up and two down quarks are required to produce a neutral particle with a charge of 0.

AS A MATTER OF FACT

Where did the name 'quark' come from?

Murray Gell-Mann was seeking a name for the particle model he was proposing. He was reading *Finnegans Wake* by James Joyce and came across the invented word 'quark' in three lines of a poem:

Three quarks for Muster Mark!
Sure he has not got much of a bark
And sure any he has it's all beside the mark.

Where do the names for the various quarks come from?

- Up and down refer to a type of spin that characterises all subatomic particles.
- Strange quarks are components of particular baryons that were found in cosmic ray showers. They had surprisingly long half-lives and so were called 'strange particles'.
- The charm quark was so-called by its discoverers because they were 'fascinated and pleased by the symmetry its discovery brought to the subnuclear world'.
- Top and bottom quarks were named as 'logical partners for the up and down quarks'.

How do you pronounce 'quark'?

It seems there are two possibilities: one sounding like 'mark' and the other sounding like 'quart'. The pronunciation rhyming with 'mark' is the more common.

16.4 Exercise 1

1 (a) The composition of the baryon called the 'charmed double bottom' is 'cbb'. What is its charge?
(b) What would you call a 'ccb' and what is its charge?
2 How many baryons, in theory, could be 'strangely charming'?
3 A neutron is described as a 'udd' and a proton as a 'uud'. What do these descriptions mean?
4 (a) The lambda baryon (Λ) was discovered by researchers at the University of Melbourne in 1950. Its quark composition is uds. What is its charge?
(b) The lambda baryon decays into a proton (uud) and a pion (ud). The quarks differ in mass. Suggest a possible mechanism for the decay.
5 Use the periodic table to determine which element is most comparable in mass to the 'top' quark.

16.5 The Standard Model

16.5.1 Particles of the Standard Model

The Standard Model (SM) of physics is a theory that seeks to explain the fundamental composition of matter in terms of elementary particles, as well as the four basic universal forces — gravity, electromagnetism, the weak nuclear force, and the strong nuclear force. The elementary particles in the Standard Model can be described as being either fermions or bosons.

Fermions are those particles that combine to make up matter. They have half-integer spin and obey Pauli's Exclusion Principle, which dictates that no two particles can have the same quantum state. Today it is accepted that there are six flavours of quarks and six flavours of leptons. Quarks and leptons can be divided neatly into groups called 'generations'. All the visible matter in the universe is composed of first-generation quarks and leptons (the up and down quarks and electrons).

TABLE 16.4 Fermions.

	LEPTONS				QUARKS			
GENERATION	NAME	SYMBOL	REST MASS (MeV)	ELECTRIC CHARGE	NAME	SYMBOL	REST MASS (MeV)	ELECTRIC CHARGE
I	Electron neutrino	ν_e	$\simeq 0$	0	Up	u	$\simeq 5$	$+\frac{2}{3}$
	Electron	e^-	0.511	-1	Down	d	$\simeq 7$	$-\frac{1}{3}$
II	Muon neutrino	ν_μ	$\simeq 0$	0	Charm	c	1500	$+\frac{2}{3}$
	Muon	μ^-	105.7	-1	Strange	s	$\simeq 150$	$-\frac{1}{3}$
III	Tau neutrino	ν_τ	< 35	0	Top	t	170 000	$+\frac{2}{3}$
	Tau	τ^-	1784	-1	Bottom	b	$\simeq 5000$	$-\frac{1}{3}$

Unlike fermions, bosons do not obey the Pauli exclusion principle and so more than one of them is able to occupy the same space at the same time. **Gauge bosons** carry forces and come in three main forms: photons, which carry the electromagnetic force, gluons, which essentially hold quarks together and are therefore considered to carry the strong nuclear force, and the W and Z bosons, which are responsible for the weak nuclear force. As yet, no boson has been found that carries the gravitational force and so gravity is the one force that the Standard Model does not explain.

The Higgs boson (referred to in 1993 by physicist Leon Lederman as the 'God particle') is the only known member of the second type of boson called a scalar boson. It was theorised in the 1960s by the English physicist Peter Higgs that the interaction of this particle with the accompanying Higgs field gives mass to all the other particles. In March 2013, CERN announced that they had identified a particle that they tentatively identified as the elusive Higgs boson. The implications of this discovery are still being explored.

16.5.2 The Standard Model today

As it presently stands, the Standard Model of particle physics is made up of 17 elementary particles (18 including the Higgs boson), which combine to form larger particles of matter or are responsible for at least three of the four universal forces. The Standard Model is a great achievement and a large number of experiments have confirmed, sometimes to incredible precision, the predictions of the Standard Model. However, it is acknowledged that it is not a perfect model, leaving many questions unanswered and appearing to contradict other notable theories. For example:

- It is incompatible with Einstein's general theory of relativity. Therefore, unification of the forces cannot involve the force of gravity. The Standard Model is a quantum–mechanical model while general relativity is not.
- The Standard Model provides no reason for the numbers of particles. Why are there six quarks and six leptons? Is it coincidence or is there an underlying reason?
- Is there some underlying truly fundamental particle such as a 'leptoquark'? This might explain why three leptons have electrical charges of one unit and quarks have electrical charges of $+\frac{2}{3}$ or $-\frac{1}{3}$ of this unit of electrical charge.

16.5.3 Future research directions

In the United States, the National Academy of Sciences has set up a special committee to assess key questions about the nature of the universe. They have identified eleven key questions, which they hope will either be answered in the next decade or that we should be thinking of answering in the following decades.

Their eleven questions are:

1. What is dark matter?
2. What are the masses of neutrinos, and how have they shaped the evolution of the universe?
3. What is the nature of dark energy?
4. Are protons unstable?
5. How did the universe begin?
6. Are there new states of matter at exceedingly high densities and temperatures?
7. Is a new theory of matter and light needed at the highest energies?
8. How were the elements from iron to uranium made?
9. Are there additional space–time dimensions?
10. Did Einstein have the last word on gravity?
11. How do cosmic accelerators work and what are they accelerating?

Most of these questions involve the close linking of particle physics and cosmology. Perhaps students in high school today may choose a career in particle physics and will one day be able to shed light on some of these questions.

16.5 Exercise 1

1 Which fermions can be described as being 'second generation'?
2 How do (a) gauge bosons differ from scalar bosons? (b) fermions differ from bosons?
3 Give two examples of phenomena that are not explained by the Standard Model.

16.6 Review

16.6.1 Summary

- There are two types of fermions, the fundamental particles of matter: quarks and leptons.
- There are six leptons. They are the electron, muon and tau particle, and the neutrinos associated with each of these.
- There are six quarks. They all have an electric charge, which is a fraction of the charge size of the electron. Three have a charge of $+\frac{2}{3}$, and three have a charge of $-\frac{1}{3}$. Their masses vary significantly.
- The quarks combine to form particles called hadrons, of which there are two types: mesons and baryons.
- Mesons are composed of one quark and one antiquark. There are many mesons. They can be positively charged, neutral or negatively charged, and have short half-lives.
- Baryons are composed of three quarks. There are a large number of different types of baryons, with a range of masses and charges ranging from +2 to −2.
- Gauge bosons are particles that carry forces. The Higgs boson interacting with a Higgs field is believed to produce mass.
- Some of the subatomic particles were predicted well before they were detected.
- Electric charges can be accelerated in a number of ways to produce electromagnetic radiation. While linear accelerators and cyclotrons were once very important accelerators, the main accelerators today are synchrotrons in which particles are accelerated around a constant radius path.

16.6.2 Questions

1. (a) A proton is composed of two up quarks and one down quark, and a neutron is composed of one up quark and two down quarks. Show that a proton will have a charge of plus one unit and a neutron will be neutral.
   (b) An antiproton would be expected to have a charge opposite to that of a proton. Identify the quarks (or antiquarks) in an antiproton and show that it will have a charge of minus one unit.
   (c) A neutron has no charge but still has an antiparticle. Identify the quarks in an antineutron and show that an antineutron is neutral.
2. Why do you think it has taken so long and been so difficult to find neutral particles such as the neutron, neutrino and the Higgs boson?
3. What do a particle and its antiparticle have in common? How do they differ?
4. How do mesons and baryons differ?
5. Some mesons are their own antiparticle. Explain with an example.
6. How many different mesons are possible, taking into account that some mesons are their own antiparticle. How many different baryons are possible? Ignore antiparticle baryons.
7. The bottom antiquark forms a family of mesons called B mesons with each of the four lighter quarks. Determine the charge and name of each meson in this family.
8. Design baryons with:
   (a) a charge of +2
   (b) a charge of −2
   (c) a charge of zero.
9. What is the charge of:
   (a) the triple bottom baryon
   (b) the baryon that is full of charm?
10. Why do you think baryons with a top quark would be hard to detect?
11. A neutron (udd) can decay into a proton (uud), an electron and an antineutrino. What do you think has happened inside the neutron?

12. The heavier mesons decay into lighter mesons, which then decay into leptons. Mesons consist of two quarks. Leptons are not quarks. What does this suggest about what can happen to quarks?
13. The tau particle decays into two pions. In light of the quark model, what is unusual about this decay?
14. It is very convincing if a scientific explanation or theory can predict a future event or discovery. What are the examples of this given in this topic and how significant was the discovery for the status of the explanation in each case?

eBook*plus* RESOURCES

Try out this interactivity: Electrons and positrons
Searchlight ID: int-6394

Complete this digital doc: Investigation: Cloud chambers
Searchlight ID: doc-26324

PRACTICAL INVESTIGATIONS

Investigation 16.1: Cloud chambers

Aim

To observe the tracks produced by charged particles passing through a cloud chamber.

Safety

Tongs or forceps should be used when handling radioactive materials. These materials should not be handled with bare hands. The sources encased in perspex are probably not suitable for use in cloud chambers.

Some cloud chambers have sources already mounted. Special care should be taken with others that rely on a radioactive salt such as thorium oxide. Gloves should be worn if it is necessary to open a bottle containing thorium oxide powder.

You should wash your hands after handling any radioactive sources.

If using a diffusion cloud chamber, gloves should be worn when handling dry ice and special care taken when breaking and or crushing the dry ice.

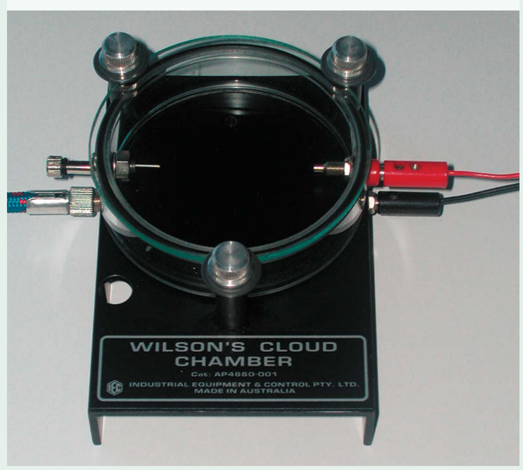
Introduction and theory

As there are two different types of cloud chamber available for use in schools, we will describe how each can be used. The principle of operation is the same for both types as they rely on the production of a supersaturated vapour. As ionising radiation passes through the supersaturated vapour, droplets condense on the ions formed and reveal the path of the radiation.

Diffusion cloud chamber

The diffusion cloud chamber was developed in 1939 by Dr Alexander Langsdorf, Jnr at the University of California at Berkeley. A diffusion cloud chamber can operate continuously, apparently for hours at a time instead of the few seconds of an expansion type chamber, but the small devices used in schools are unable to maintain the correct conditions for much longer than fifteen to twenty minutes. In these cloud chambers, alcohol from the top of the chamber evaporates and then diffuses downwards towards the base of the chamber. The base of the chamber is cooled with dry ice (beneath the base) and as the alcohol vapour diffuses downwards, the vapour becomes supersaturated. There should be a region in the chamber that maintains a supersaturated vapour and the trails of charged particles through this region can be observed.

FIGURE 16.16 A Wilson cloud chamber.



Expansion cloud chamber

This was the original type of cloud chamber and was invented by C.T.R. Wilson in 1911 to detect radiation. In such a cloud chamber, a piston below the chamber was lowered to reduce the pressure and cool the gas in the chamber. This produced a supersaturated vapour. In the expansion cloud chamber made in Australia by IEC, the expansion is performed by a bicycle pump (with the washer reversed to extract air from the chamber).

Reduction of the pressure causes the vapour to become supersaturated and, hopefully, trails appear. A voltage applied to a metal ring inside the chamber sweeps the ions from the chamber before another expansion or pressure reduction can be performed. A disadvantage of such a chamber is that it is then necessary to wait perhaps a minute for the alcohol to evaporate into the chamber before repeating the process. The small school version can usually be used without this wait. Once the correct conditions have been attained, withdrawing the handle of the bicycle pump can usually be done every few seconds.

Part A: Observing tracks in a diffusion cloud chamber

Apparatus

cloud chamber (which probably has a built-in radioactive source)
woollen cloth
alcohol (Propan-2-ol also known as iso-propyl alcohol recommended but ethyl alcohol should work)
dry ice
light source (possibly a microscope lamp)

Method

We will assume that the cloud chamber has a built-in source.

Remove the perspex top and moisten the felt ring with a few drops of alcohol. Turn the chamber upside-down and unscrew the base and remove the foam pad. Place some small pieces of dry ice over the black metal plate. These will be held against the metal by the foam when the base is screwed back on.

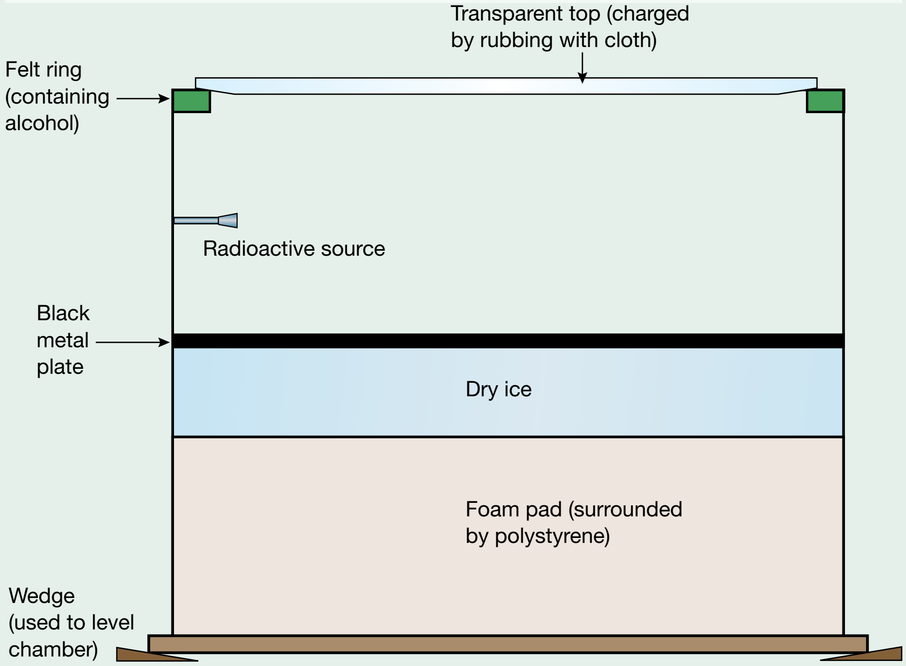
Place the chamber right side up on small wedges (probably provided) and level it as carefully as possible. If the metal plate is not horizontal, convection currents can hinder the production of tracks. Replace the perspex top and rub it gently with the woollen cloth to charge it. (Charging the top of the chamber produces an electric field that will remove dust and ions and this should assist in the production of tracks in the chamber.)

After a few minutes, tracks should become visible.

If a light is shone horizontally through the side of the chamber, the visibility of the tracks may be enhanced.

The chamber should continue to operate while the temperature difference between the top and bottom of the chamber region is maintained. If the trails become hard to see, recharging the perspex top may help.

FIGURE 16.17 A diffusion cloud chamber.



It is sufficient to set up a cloud chamber and observe the tracks of alpha particles through the chamber.

If the cloud chamber works particularly well, you may be able observe other tracks that do not come from the source. (Cloud chambers played an important role in the discovery of cosmic rays.)

Part B: Observing tracks in an expansion cloud chamber

The IEC Wilson's Expansion Cloud Chamber is described below. Usually expansion cloud chambers have the disadvantage that they provide a brief view of the tracks when expansion occurs and then there is a delay before the expansion can be repeated. However, with this particular cloud chamber, expansions can be performed every few seconds.

Apparatus

Wilson's Expansion Cloud Chamber including modified bicycle pump, radioactive source
alcohol (propan-2-ol, also known as isopropyl alcohol, is recommended but ethyl alcohol should work.)
light source (possibly a microscope lamp)
high-voltage DC power supply (at least 300 V)

Method

Preparing the radioactive source

The cloud chamber is supplied with a small quantity of thorium oxide. The thorium oxide can be used to prepare a point radioactive source (which can be screwed into the side of the chamber) or radon 220 gas, produced from the decay of thorium, can be used as the source. (If you search for information on the decay of thorium, you will find that radon 220 is produced part of the way along the decay series.)

Special care must be taken if it is necessary to transfer thorium oxide into the squeeze bottle that is used to inject the radon into the cloud chamber. Your teacher will probably do this for you. (The apparatus works well with about 10 grams of thorium oxide in the squeeze bottle.)

Setting up the expansion cloud chamber

Make sure that the base of the cloud chamber is level.

Remove and clean the glass top. Pour a few millilitres of propan-2-ol onto the black metal disc at the bottom of the chamber (2 to 3 millilitres works well).

Replace the top and gently tighten the screws to produce a seal.

Attach the hose from the squeeze bottle containing the thorium oxide to the fitting on the side of the chamber.

Connect the terminals on the side and base of the chamber to a high-voltage power supply. (The instructions say up to 600 volts can be used but excellent tracks were observed with a voltage of 200 volts. This high voltage is essential for the production of tracks.)

Set up the microscope lamp so that it shines horizontally into the chamber. (It should not be too close to the chamber as it is essential that the chamber is not heated.)

Release the Mohr clip on the hose connecting the squeeze bottle to the chamber, squeeze the bottle gently once and then replace the clip. (As the purpose of this is to inject some radon 220 gas into the chamber, it is easier to do this if the bicycle pump is disconnected, which means that the chamber is not sealed.)

Reconnect the bicycle pump, and quickly and smoothly withdraw the handle of the pump.

Tracks should be visible throughout the chamber. The radon gas should have spread throughout the chamber and, as it decays by alpha particle emission, the short tracks produced by alpha particles should be visible throughout the chamber.

The trails disappear quickly but repeating within a few seconds should produce another set of trails.

Occasionally much longer trails, which are probably thinner than those of the alpha particles, will be observed. What particles are likely to produce these trails and what might be their origin?

The number of tracks produced will gradually decrease. (Check the half-life of radon 220 and you will see why.) Eventually it will be necessary to squeeze another puff of gas into the chamber.

Further investigations

It is an achievement to set up a cloud chamber and observe the trails. If you find that you are able to get your cloud chamber working very well without any difficulty you might like to try to investigate further.

You could try different sources and compare the tracks produced by different types of radiation. (This might be impractical with a diffusion cloud chamber but might well be possible with the expansion cloud chamber.)

You could try to produce a magnetic field in the chamber and see if it is possible to deflect the particles in the magnetic field.

GLOSSARY

absorption spectrum: a series of dark lines on a coloured background that is produced when white light is passed through a cool gas and viewed through a spectroscope. *p. 369*

accrete: the accumulation of particles attracted by gravitational force. *p. 325*

air resistance or **drag:** the force applied to an object, opposite to its direction of motion, by the air through which it is moving. *p. 2*

angular momentum, L: of a point mass, m, which is in circular motion of radius, r, with velocity, v, is given by: $L = m\,v\,r$. Angular momentum is the rotational equivalent to linear momentum and is an important quantity in rotational motion. (It follows a similar conservation principle to linear momentum.) *p. 370*

anode/cathode: cathode rays are now known to be streams of electrons emitted within an evacuated tube from a cathode (negative electrode) to an anode (positive electrode). They were first observed in discharge tubes. *p. 341*

armature: a frame around which a coil of wire is wound, which rotates in a motor's magnetic field. *p. 158*

average angular velocity ω: the rate of change in angular displacement with respect to time. *p. 46*

average binding energy per nucleon: the total binding energy of a nucleus divided by the number of nucleons in the nucleus. It is a measure of the stability of the nucleus. *p. 405*

back emf: an electromagnetic force that opposes the main current flow in a circuit. When the coil of a motor rotates, a back emf is induced in the coil due to its motion in the external magnetic field. *p. 166*

banking: the angling of a surface following a curved path to allow the reaction force of a body travelling on the banked surface to contribute to the net force acting. *p. 39*

baryons: hadrons with three quarks. *p. 429*

binding energy: the energy equivalent of the mass defect of the nucleus. It is the energy that would have to be provided and converted to mass to enable all the nucleons in a nucleus to be separated from each other. *p. 405*

black dwarf: a white dwarf that has cooled sufficiently so that it no longer emits heat or light. *p. 331*

black hole: the crushed remnant of the core (greater than 5 solar masses) of a very massive star. Theoretically, it is a point of zero volume and infinite density. *p. 326*

brushes: conductors that make electrical contact with the moving split metal ring of a commutator. *p. 160*

cathode ray tube or **discharge tube:** a sealed glass tube from which most of the air is removed by vacuum pump. A beam of electrons travels from the cathode to the anode and can be de.ected by electrical and/or magnetic fields. *p. 341*

centripetal force: the force that acts to maintain circular motion and is directed towards the centre of the circle. *p. 33*

coherent: when there is a constant phase difference between light waves; that is, the peaks line up and the troughs line up. *p. 217*

commutator: a device for reversing the direction of a current flowing through an electric circuit, for example, the coil of a motor. *p. 158*

constructive interference: the disturbance caused when two waves reach a position at the same time and give rise to an amplitude that is greater than that due to each of the waves alone. *p. 215*

cosmology: the study of how the universe began, has evolved and will end. *p. 305*

de Broglie wavelength: the wavelength associated with a particle or discrete piece of matter. *p. 378*

decay chain: also known as a decay series; the sequence of stages a radioisotope passes through to become more stable. At each stage, a more stable isotope forms. The chain ends when a stable isotope forms. *p. 402*

decay equation: a representation of a decay reaction. It shows the changes occurring in nuclei and lists the products of the decay reaction. *p. 408*

destructive interference: the disturbance that occurs when the sum of two superimposed waves is zero. *p. 215*

eccentricity: the degree that an ellipse differs from a perfect circle. *p. 61*

eddy current: a circular or whirling current induced in a conductor that is stationary in a changing magnetic field, or that is moving through a magnetic field. They resemble the eddies or swirls left in the water after a boat has gone by. *p. 134*

electromagnetic induction: the generation of an emf and/or electric current through the use of a magnetic field. *p. 126*

emission spectrum: a series of brightly coloured lines on a dark background that is produced when light from an excited gas is viewed through a spectroscope. *p. 369*

empirical equation: one that has no theoretical basis but can be used to calculate correct values. Kepler's Third Law, $T^2 \propto R^3$ is another example of an empirical equation. *p. 367*

escape velocity: the initial velocity required by a projectile to rise vertically and just escape the gravitational field of a planet. *p. 82*

excited state: when an electron exists in a stationary state in which it has more energy. *p. 376*

fermions: particles that have half-integer spins. They obey the Pauli exclusion principle. *p. 438*

fission fragments: the products from a nucleus that undergoes fission. The fission fragments are smaller than the original nucleus. *p. 413*

fluorescence: the emission of light from a material when it is exposed to streams of particles or external radiation. *p. 341*

flux: from the Latin word *fluo* meaning 'flow'. Flux is a state of flowing or movement. In physics, flux is the rate of flow of a fluid, radiation or particles. *p. 129*

fusion: a thermonuclear reaction in which nuclei of light atoms join to form nuclei of heavier atoms. *p. 328*

galvanometer: an instrument for detecting small electrical currents. *p. 127*

gauge bosons: force carriers such as photos, gluons and the W and Z bosons. *p. 438*

geostationary orbit: an altitude at which the period of the orbit precisely matches that of the Earth. This corresponds to an altitude of approximately 35 800 km. *p. 72*

globular cluster: a very old, densely packed cluster of stars in the shape of a sphere. *p. 304*

ground state: the state an electron is in when it has the lowest possible amount of energy. *p. 376*

half-life: the time taken for half the radioactive nuclei in a sample to decay. If we exclude the activity of daughter nuclei, it is the time taken for the activity of a particular sample to drop to half its initial value. *p. 399*

Hohmann transfer orbit: the most efficient intermediate orbit to transfer from one circular orbit to another. *p. 73*

Hubble's constant: the constant of proportionality relating the speed that galaxies are receding from Earth and their distance from Earth. *p. 308*

Hubble's Law: states that the speed of recession of galaxies is proportional to their distance from Earth. *p. 308*

induction: a process where one object with magnetic or electrical properties can produce the same properties in another object without making physical contact. *p. 129*

induction motor: an AC machine in which torque is produced by the interaction of a rotating magnetic field produced by the stator and currents induced in the rotor. *p. 176*

inertial reference frames: reference frames that are not accelerating. *p. 265*

invariant: describes a quantity that has the same value in all reference frames. *p. 266*

isotope: a nuclide that has the same number of protons but different numbers of neutrons compared to another nuclide of the same element. *p. 398*

leptons: particles that do not experience the strong nuclear force. They are all fermions with half-integer spin. An electron is a lepton. *p. 428*

low Earth orbit: an orbit higher than 250 km and lower than 1000 km. *p. 72*

magnetic flux, ϕ_B: the amount of magnetic field passing through a given area. In the SI system, ϕ_B is measured in weber (Wb). *p. 129*
magnetic flux density: the strength of a magnetic field, B. In the SI system, B is measured in tesla (T) or weber per square metre (Wb m^{-2}). *p. 129*
magnetic permeability: the ability of a material to support the formation of a magnetic field. *p. 117*
main sequence star: characterised by the fusion of hydrogen to helium in its core. *p. 322*
mass defect: the difference between the mass of the constituent nucleons of a nucleus and the mass of the nucleus. *p. 404*
mesons: hadrons with two quarks. *p. 429*
motor effect: the action of a force experienced by a current-carrying conductor in an external magnetic field. *p. 113*
muon: an unstable subatomic particle of the same charge as an electron but with a mass approximately two hundred times greater. *p. 427*
nuclide: refers to a particular nucleus with certain values of Z (atomic number) and A (mass number). *p. 398*
path difference: the difference between the lengths of the paths from each of two sources of waves to a point. *p. 218*
period: (of a cycle or series of events) is the amount of time one cycle or one event takes to occur. *p. 28*
phosphorescent: a substance that absorbs radiation of one wavelength and then emits radiation of a different wavelength over a period of time. The hands of some analogue watches are coated with a phosphorescent substance to enable them to be seen in the dark. *p. 396*
photoelectric effect: the release of electrons from a metal surface as a result of exposure to electromagnetic radiation. *p. 242*
photon: a quantum (or discrete packet) of electromagnetic radiation. It can be thought of as an elementary particle with zero rest mass and charge, travelling at the speed of light. *pp. 240, 366*
planetary nebula: a shell-shaped cloud of gas that is the blown-away outer layers of a star. *p. 331*
polarisation: the blocking of transverse waves except for those travelling in a single plane. *p. 223*
positron: a positively charged particle with the same mass as an electron. *pp. 408, 427*
principal quantum number: the value of n for each stationary state or orbit of the Bohr atom. *p. 372*
proper length: the length of an object measured in its rest frame. *p. 281*
proper time: the time measured in a frame of reference where the events occur at the same point in space. The proper time of a clock is the time the clock measures in its own reference frame. *p. 276*
quantum (plural: **quanta**)**:** can be considered to be the smallest amount of energy possible in a given situation. Planck's atomic oscillators could oscillate only with certain precise amounts of energy. *p. 366*
radians: One radian is the angle subtended at the centre of a circle by an arc that is equal in length to the radius of the circle. It is equal to approximately 57.3 degrees. *p. 46*
range: horizontal displacement of a projectile from its starting point when it lands. *p. 11*
red giant: a star characterised by a helium-burning core surrounded by a hydrogen-burning shell. *p. 330*
red-shift: the increase in wavelength that results from a light source moving away from the observer. *p. 306*
relative: describes a quantity that has different values for different observers. *p. 260*
rest mass: the mass of an object measured at rest. *p. 289*
right-hand grip rule: used to find the direction of a magnetic field around a straight current-carrying conductor. *p. 134*
right-hand push rule: (also called the right-hand palm rule) used to find the direction of the force acting on a moving charged particle or current-carrying conductor in an external magnetic field. *p. 113*
rotor: the rotating part of an electrical rotating machine. *p. 168*
scalar field: a varying mathematical function that assigns a value to every point in a space. *p. 76*
slip speed: the difference between the speed of the rotating magnetic field and the speed of the rotor. *p. 179*
solenoid: a coil of wire wound into a cylindrical shape. *p. 110*

split metal ring: the two-piece conducting metal surface of a commutator. Each part is connected to the coil. *p. 159*

squirrel-cage rotor: an assembly of parallel conductors and short-circuiting end rings in the shape of a cylindrical squirrel cage. *p. 177*

stationary state: the state an electron is in when it orbits the nucleus without emitting any electromagnetic radiation. *p. 370*

stator: the non-rotating magnetic part of a motor. *pp. 158, 168*

step-down transformer: provides an output voltage that is less than the input voltage. *p. 137*

step-up transformer: provides an output voltage that is greater than the input voltage. *p. 137*

stroboscope: a light that produces quick flashes at regular (usually small) time periods. *p. 7*

strong nuclear force: the force that holds nucleons together in a nucleus of an atom. It acts over only very short distances. *p. 402*

supernova: a violent explosion of uncontrolled nuclear reactions that completely blows away the various layers of a massive star (original mass greater than five solar masses). *p. 325*

tension: the force that acts to resist the stretching of a material; also refers to process of pulling the particles of a material further apart. *p. 34*

terminal: the free end of a cell or battery to which a connection is made to the rest of a circuit. *p. 170*

terminal velocity: the constant velocity reached by a falling object when the upwards air resistance becomes equal to the downward force of gravity. *p. 6*

thought experiments: also known as gedanken experiments; imaginary scenarios designed to explore what the laws of physics predict would happen. *p. 274*

time dilation: the slowing of time by clocks moving relative to the observer. *p. 276*

torque: also referred to as moment; the turning effect of a force about a pivot or reference point. *pp. 50, 164*

trajectory: the path that a projectile follows during its flight. *p. 7*

transfer orbit: an orbit used to manoeuvre a satellite from one orbit to another. *p. 72*

transformer: a device in which two multi-turn coils are wound around an iron core. One coil acts as an input while the other acts as an output. The purpose of the transformer is to produce an output AC voltage that is different from the input AC voltage. *p. 136*

transmutation: when a radioactive atom emits an alpha particle or a beta particle and an atom of a new element is produced. A new daughter element is formed from a parent element. *p. 398*

triple alpha process: the process of helium fusion in the core of a red giant. *p. 330*

uniform circular motion: circular motion with a uniform orbital speed. *p. 28*

universal motor: a series-wound motor that may be operated on either AC or DC electricity. *p. 175*

vector field: a space in which each point is assigned a vector that has a magnitude and a direction. *p. 76*

wavefront: either the crest or trough of a wave. The wavefront is perpendicular to the direction of the velocity of the wave. *p. 214*

wave–particle duality: describes light as having characteristics of both waves and particles. This duality means that neither the wave model nor the particle model adequately explains the properties of light on its own. *p. 385*

weight: the force applied to an object due to gravitational attraction. *p. 2*

white dwarf: a dense star made of degenerate matter. It is the end point of small- to medium-sized stars. *p. 331*

work function: the minimum energy required to release an electron from the surface of a material. *p. 246*

APPENDIX 1
Data and formulae sheet

DATA

Charge on electron, q_e	-1.602×10^{-19} C
Mass of electron, m_e	9.109×10^{-31} kg
Mass of neutron, m_n	1.675×10^{-27} kg
Mass of proton, m_p	1.673×10^{-27} kg
Speed of sound in air	$340\ \mathrm{m\,s^{-1}}$
Earth's gravitational acceleration, g	$9.8\ \mathrm{m\,s^{-2}}$
Speed of light, c	$3.00 \times 10^{8}\ \mathrm{m\,s^{-1}}$
Electric permittivity constant, ε_0	$8.854 \times 10^{-12}\ \mathrm{A^2 s^4 kg^{-1} m^{-3}}$
Magnetic permeability constant, μ_0	$4\pi \times 10^{-7}\ \mathrm{NA^{-2}}$
University gravitational constant, G	$6.67 \times 10^{-11}\ \mathrm{N\,m^2 kg^{-2}}$
Mass of Earth, M_E	6.0×10^{24} kg
Radius of Earth, r_E	6.371×10^{6} m
Planck constant, h	6.626×10^{-34} J s
Rydberg constant, R (hydrogen)	$1.097 \times 10^{7}\ \mathrm{m^{-1}}$
Atomic mass unit, u	1.661×10^{-27} kg
	$931.5\ \mathrm{MeV/c^2}$
1 eV	1.602×10^{-19} J
Density of water, ρ	$1.00 \times 10^{3}\ \mathrm{kg\,m^{-3}}$
Specific heat capacity of water	$4.18 \times 10^{3}\ \mathrm{J\,kg^{-1}\,K^{-1}}$
Wien's displacement constant, b	2.898×10^{-3} m K

FORMULAE

Motion, forces and gravity

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

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

$$\Delta U = mg\Delta h$$

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

$$\sum \frac{1}{2}mv^2_{\text{before}} = \sum \frac{1}{2}mv^2_{\text{after}}$$

$$\Delta p = F_{net}\Delta t$$

$$\omega = \frac{\Delta\theta}{t}$$

$$\tau = rF_{\perp} = rF\sin\theta$$

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

$$U = -\frac{GMm}{r}$$

$v = u + at$

$F_{net} = ma$

$W = F_{\parallel} s = Fs \cos\theta$

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

$P = F_{\parallel} v = Fv \cos\theta$

$\sum m v_{\text{before}} = \sum m v_{\text{after}}$

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

$Fc = \frac{mv^2}{r}$

$F = \frac{GMm}{r^2}$

$\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$

Waves and thermodynamics

$v = f\lambda$

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

$d \sin\theta = m\lambda$

$n_x = \frac{c}{v_x}$

$I = I_{\max} \cos^2\theta$

$Q = mc\Delta T$

$f_{\text{beat}} = |f_2 - f_1|$

$f' = f\frac{(v_{\text{wave}} + v_{\text{observer}})}{(v_{\text{wave}} - v_{\text{source}})}$

$n_1 \sin\theta_1 = n_2 \sin\theta_2$

$\sin\theta_c = \frac{n_2}{n_1}$

$I_1 r_1^2 = I_2 r_2^2$

$\frac{Q}{t} = \frac{kA\Delta T}{d}$

Electricity and magnetism

$E = \frac{V}{d}$

$V = \frac{\Delta U}{q}$

$W = qV$

$W = qEd$

$B = \frac{\mu_0 I}{2\pi r}$

$B = \frac{\mu_0 NI}{L}$

$\phi = B_{\parallel} A = BA \cos\theta$

$\varepsilon = -N\frac{\Delta\phi}{\Delta t}$

$\frac{V_p}{V_s} = \frac{N_p}{N_s}$

$F = qE$

$F = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r^2}$

$I = \frac{q}{t}$

$V = IR$

$P = VI$

$F = qv_{\perp}B = qvB \sin\theta$

$F = lI_{\perp}B = lIB \sin\theta$

$\frac{F}{l} = \frac{\mu_0}{2\pi} \frac{I_1 I_2}{r}$

$\tau = nIA_{\perp}B = nIAB \sin\theta$

$V_p I_p = V_s I_s$

Quantum mechanics, special relativity and nuclear physics

$$\lambda = \frac{h}{mv}$$

$$E_{k,\max} = hf - W$$

$$\lambda_{\max} = \frac{b}{T}$$

$$E = mc^2$$

$$E = hf$$

$$\frac{1}{\lambda} = R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

$$t = \frac{t_0}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$$

$$l = l_0\sqrt{\left(1 - \frac{v^2}{c^2}\right)}$$

$$p_v = \frac{m_0 v}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$$

$$N_t = N_0 e^{-\lambda t}$$

$$\lambda = \frac{\text{In}2}{t_{\frac{1}{2}}}$$

APPENDIX 2
Periodic table

Key: 4 — Atomic number; Be — Symbol; 9.012 — Atomic weight; Beryllium — Name

Period \ Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H 1.008 Hydrogen																	2 He 4.003 Helium
2	3 Li 6.941 Lithium	4 Be 9.012 Beryllium											5 B 10.81 Boron	6 C 12.01 Carbon	7 N 14.01 Nitrogen	8 O 16.00 Oxygen	9 F 19.00 Fluorine	10 Ne 20.18 Neon
3	11 Na 22.99 Sodium	12 Mg 24.31 Magnesium											13 Al 26.98 Aluminium	14 Si 28.09 Silicon	15 P 30.97 Phosphorus	16 S 32.07 Sulfur	17 Cl 35.45 Chlorine	18 Ar 39.95 Argon
4	19 K 39.10 Potassium	20 Ca 40.08 Calcium	21 Sc 44.96 Scandium	22 Ti 47.87 Titanium	23 V 50.94 Vanadium	24 Cr 52.00 Chromium	25 Mn 54.94 Manganese	26 Fe 55.85 Iron	27 Co 58.93 Cobalt	28 Ni 58.69 Nickel	29 Cu 63.55 Copper	30 Zn 65.38 Zinc	31 Ga 69.72 Gallium	32 Ge 72.64 Germanium	33 As 74.92 Arsenic	34 Se 78.96 Selenium	35 Br 79.90 Bromine	36 Kr 83.80 Krypton
5	37 Rb 85.47 Rubidium	38 Sr 87.61 Strontium	39 Y 88.91 Yttrium	40 Zr 91.22 Zirconium	41 Nb 92.91 Niobium	42 Mo 95.96 Molybdenum	43 Tc [98] Technetium	44 Ru 101.1 Ruthenium	45 Rh 102.9 Rhodium	46 Pd 106.4 Palladium	47 Ag 107.9 Silver	48 Cd 112.4 Cadmium	49 In 114.8 Indium	50 Sn 118.7 Tin	51 Sb 121.8 Antimony	52 Te 127.6 Tellurium	53 I 126.9 Iodine	54 Xe 131.3 Xenon
6	55 Cs 132.9 Caesium	56 Ba 137.3 Barium	57–71 * Lanthanides	72 Hf 178.5 Hafnium	73 Ta 180.9 Tantalum	74 W 183.8 Tungsten	75 Re 186.2 Rhenium	76 Os 190.2 Osmium	77 Ir 192.2 Iridium	78 Pt 195.1 Platinum	79 Au 197.0 Gold	80 Hg 200.6 Mercury	81 Tl 204.4 Thallium	82 Pb 207.2 Lead	83 Bi 209.0 Bismuth	84 Po [209] Polonium	85 At [210] Astatine	86 Rn [222] Radon
7	87 Fr [223] Francium	88 Ra [226] Radium	89–103 ** Actinides	104 Rf [261] Rutherfordium	105 Db [262] Dubnium	106 Sg [266] Seaborgium	107 Bh [264] Bohrium	108 Hs [265] Hassium	109 Mt [268] Meitnerium	110 Ds [271] Darmstadtium	111 Rg [280] Roentgenium	112 Cn Copernicium	113 Nh Nihonium	114 Fl Flerovium	115 Mc Moscovium	116 Lv Livermorium	117 Ts Tennessine	118 Og Oganesson

*Lanthanide series

57 La 138.9 Lanthanum	58 Ce 140.1 Cerium	59 Pr 140.9 Praseodymium	60 Nd 144.2 Neodymium	61 Pm [145] Promethium	62 Sm 150.4 Samarium	63 Eu 152.0 Europium	64 Gd 157.3 Gadolinium	65 Tb 158.9 Terbium	66 Dy 162.5 Dysprosium	67 Ho 164.9 Holmium	68 Er 167.3 Erbium	69 Tm 168.9 Thulium	70 Yb 173.1 Ytterbium	71 Lu 175.0 Lutetium

**Actinide series

89 Ac [227] Actinium	90 Th 232.0 Thorium	91 Pa 231.0 Protactinium	92 U 238.0 Uranium	93 Np [237] Neptunium	94 Pu [244] Plutonium	95 Am [243] Americium	96 Cm [247] Curium	97 Bk [247] Berkelium	98 Cf [251] Californium	99 Es [252] Einsteinium	100 Fm [257] Fermium	101 Md [258] Mendelevium	102 No [259] Nobelium	103 Lr [262] Lawrencium

- For elements with no stable nuclides, the mass of the longest living isotope is given in square brackets.
- The atomic weights of Np and Tc are given for the isotopes ^{237}Np and ^{99}Tc.

ANSWERS

Topic 1 Projectile motion

1.2 Exercise 1

5. $-9.9\ \text{m s}^{-1}$

6. 2.0 m

7. $-10\ \text{m s}^{-1}$

8. (a) 44 m; (b) $29.4\ \text{m s}^{-1}$, rounded to $29\ \text{m s}^{-1}$

1.3 Exercise 1

2. (a) 8.0 m; (b) 2.6 s; (c) $7.6\ \text{m s}^{-1}$; (d) $2.2\ \text{m s}^{-1}$

3. (a) $3.13\ \text{m s}^{-1}$, rounded to $3.1\ \text{m s}^{-1}$; (b) $0.638 = 0.64$ s

4. $u_x = 900\ \text{m s}^{-1}$, $u_y = -7.84\ \text{m s}^{-1}$

5. $u_x = 5\ \text{m s}^{-1}$, $u_y = -4.9\ \text{m s}^{-1}$

6. $13.7\ \text{m s}^{-1}$ at 59.3° below the horizontal

7. (a) $230\ \text{m s}^{-1}$; (b) 1.715 m; (c) 120 m

8. 0.4 s

9. 12.4 m

10. 3.9 m

11. (a) 76.4 m, rounded to 76 m; (b) 160 m;
(c) $39.2\ \text{m s}^{-1}$, rounded to $39\ \text{m s}^{-1}$; (d) 44.4°, rounded to 44°

1.4 Exercise 1

1. 21.8 m

2. (a) 2.24 s; (b) 42.6 m; (c) 4.8 m

3. 13.7 m

4. $59\ \text{m s}^{-1}$

5. (a) 2.1 m; (b) 32 m

1.5.2 Questions

9. (a) Vertical equals $15\ \text{m s}^{-1}$, horizontal equals $13\ \text{m s}^{-1}$;
(b) $10\ \text{m s}^{-1}$, $4.3\ \text{m s}^{-1}$;
(c) $5\ \text{m s}^{-1}$, zero;
(d) Zero, $10\ \text{km h}^{-1}$;
(e) $17\ \text{m s}^{-1}$, $29\ \text{m s}^{-1}$

10. (a) 1.1 s; (b) $-10.8\ \text{m s}^{-1}$; (c) 5.95 m

11. $4.4 \times 10^2\ \text{m s}^{-1}$

12. 3.14 m

13. (a) 2.39 m; (b) 6.14 m

14. (a) 6.2 s; (b) 108.5 m

15. Yes

16. 3390 m

17. (a) 5.5 s; (b) $55\ \text{m s}^{-1}$

18. (a) 3.6 s; (b) 16 m

19. No, as the range is only 4.9 m

20. $5.5\ \text{m s}^{-1}$

21. 19°

22. 1.22 metres above the target

23. 44.3 m

Topic 2 Circular motion

2.2 Exercise 1

1. 26 m s^{-1}

2. 2.7 rev s^{-1}

3. 160 s

4. (a) 3.1 m; (b) 1.3 m s^{-1}; (c) 0.828 m s^{-1}, rounded to 0.83 m s^{-1}; (d) 0 m s^{-1}

5. 12 m s^{-2}

6. 230 m s^{-2}

7. 230 N

8. (a) 7.85 m s^{-1}; (b) 1.2 s; (c) 1.7 revs

9. (a) 9.0 m s^{-2} towards the centre of the roundabout;
(b) 1.1×10^4 N towards the centre of the roundabout

10. (a) 22 m s^{-2};
(b) 1326 N, rounded to 1300 N, towards the centre of the Gravitron

2.3 Exercise 1

3. 60°

4. 46.2°

5. 17.7m s^{-1}

6. 11 m s^{-1}

7. 13.3 m s^{-1}

8. 18°

9. 53°

10. (a) 61°; (b) 1011 N, rounded to 1000 N

2.4 Exercise 1

2. (a) 1127 N, rounded to 1100 N;
(b) 1715 N, rounded to 1700 N;
(c) 2.9, almost three times the weight force

3. 2507 N, rounded to 2500 N

4. 4.849 m s^{-1}, rounded to 4.8 m s^{-1}

5. (a) 8.3 m s^{-1}; (b) 13.5 N;
(c) At the bottom; (d) 21 m s^{-1}

2.5 Exercise 1

1. (a) 0.52 rad; (b) 3.14 rad; (c) 6.1 rad; (d) 8.0 rad

2. (a) South;
(b) Downwards towards the ground under the wheel

3. (a) 150 rad s^{-1}; (b) 31.4 m s^{-1}; (c) 10 030 rpm

4. 500 N

5. 21.7°

6. 1 m

2.6.2 Questions

1. $\boldsymbol{F} = 31$ N, $\boldsymbol{a} = 78$ m s^{-2}

2. 19 400 N

3. (a) 5 s; (b) 2.5 m s^{-1}; (c) 3 m s^{-2}; (d) 1.26 rad s^{-1}

4. (a) 0.06 s; (b) 104 rad s^{-1}; (c) 42 m s^{-1}

5. 1176 N m

7. Upwards

8. (a) 0.024 m s^{-2} towards the centre of the circle;
(b) 1.6 N towards the centre of the circle

9. Lucy

10. (a) $0.050\ \text{m s}^{-2}$ towards the centre of the circle;
(b) 1.7 N towards the centre of the circle;
(c) 75 N towards the centre of the circle

11. (a) $0.95\ \text{m s}^{-2}$ towards the centre of the circle;
(b) 0.11 N towards the centre of the circle

14. (a) $11.7\ \text{m s}^{-1}$;
(b) $92.0\ \text{m s}^{-2}$ towards the centre of the circle;
(c) 4.60 N towards the centre of the circle;
(d) 4.60 N towards the centre of the circle;

15. (a) 12 N (11.8 N); (b) $11\ \text{m s}^{-1}$ $(10.9\ \text{m s}^{-1})$; (c) 1.2 s

16. (a) 350 N towards the centre of the circle;
(b) 350 N towards the centre of the circle;
(c) It will increase to 5.0 m.

17. 78.9 kg

18. 82° (i.e. banking alone is not the solution)

19. (b) (i) 8000 N downwards; (ii) $6.3\ \text{m s}^{-1}$

20. (a) $4.5\ \text{m s}^{-1}$; (b) 3.3×10^2 N upwards

21. 25 N

22. 8.7 N

23. (a) 816 m; (b) 4.8×10^3 N

24. (a) 5.2 s; (b) $3.6\ \text{m s}^{-1}$

25. 2.64 m

Topic 3 Motion and gravitational fields

3.2 Exercise 1

1. A

2. 0.72 AU

3. (a) (i) $3.4 \times 10^{18}\ \text{m}^3\text{s}^{-2}$; (ii) $3.4 \times 10^{18}\ \text{m}^3\text{s}^{-2}$

4. 1.1×10^9 m

3.3 Exercise 1

1. 686 N towards the centre of the Earth

2. (a) 1.98×10^{20} N; (b) 1.98×10^{20} N

3. 4.2×10^{-12} N

4. 1.64×10^{22} N

5. 0.36 N

6. 2.38×10^{20} N towards the Sun

7. 23.75 cm from the 6 kg mass

8. ~ 124 days

3.4 Exercise 1

3. Low Earth: 360.0, 1.53, 7686; Geostationary: 35 800, 23.93, 3070

5. $7749\ \text{m s}^{-1}$

6. $1629\ \text{m s}^{-1}$

7. $3885\ \text{m s}^{-1}$, 11 hr 51 min 4 s

8. 6900 N

3.5 Exercise 1

3. (a) 0; (b) $4.9\ \text{N kg}^{-1}$; (c) $9.71\ \text{N kg}^{-1}$

4. one-ninth less

5. 2.08 s longer

3.6 Exercise 1

1. 37 300 km h^{-1}

2. (a) -7.4×10^{30} J; (b) -3.24×10^{35} J

3. Escape velocity would double

4. (a) 1.8×10^{32} J; (b) 9×10^{31} J; (c) 9×10^{31} J

5. (a) 4.2×10^{7} m;
(b) 2.12×10^{26} kg;
(c) (i) 1.68×10^{10} J, (ii) -3.36×10^{10} J

3.7.2 Questions

7. (a) 28 400 km h^{-1}; (b) 85 min; (c) 2.44 N

8. (a) 28 050 km h^{-1}; (b) 88.1 min;
(c) 9.25 m s^{-2} towards Earth's centre; (d) 1020 000 N

13. 1.48×10^{-10} N

14. (a) 710 N; (b) 650 N

15. (a) Satellite:
orbital velocity, 7721 m s^{-1}
centripetal force, 1.21×10^{4} N
gravitational force, 1.21×10^{4} N
(b) Venus:
orbital velocity, 3.52×10^{4} m s^{-1}
centripetal force, 5.6×10^{22} N
gravitational force, 5.5×10^{22} N
(c) Callisto:
orbital velocity, 8186 m s^{-1}
centripetal force, 3.9×10^{21} N
gravitational force, 3.9×10^{21} N

16. (a) 0.124; (b) 0.515; (c) 0.904; (d) 0.466

17. Mercury: 4250 m s^{-1}, Venus: 10 400 m s^{-1}, Io: 2550 m s^{-1}, Callisto: 2470 m s^{-1}

18. (a) 97 N downwards; (b) 97 N; (c) 6.41×10^{6} m

19. 1.6×10^{22} kg

22. (a) 8.73 m s^{-2}; (b) 8.78 N kg^{-1};
(d) 1.1×10^{7} N; (e) 2.2×10^{-4} N

23. 0.034 m s^{-2}

29. 4.2×10^{7} m

31. (a) 1.4×10^{9} J;
(b) 5.8×10^{3} s or 97 minutes;
(c) $r^3 T^{-2} =$ constant for any satellite of Earth;
(d) (a) would be halved and (b) would remain the same.

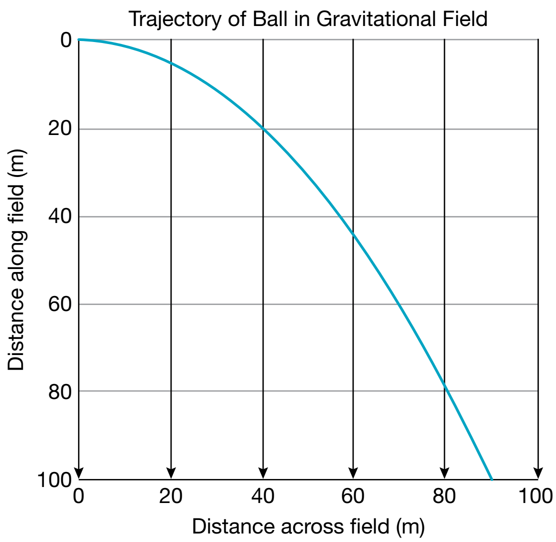
Topic 4 Electric and magnetic fields

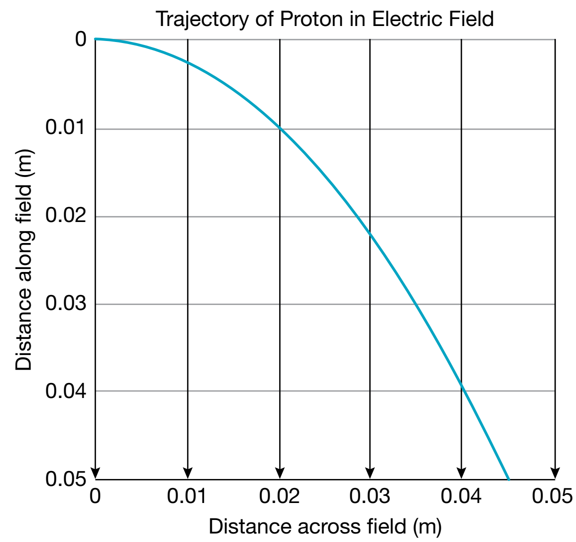
4.2 Exercise 1

1. Scenario 1:
(a) 4.52 s;
(b) 44.3 m s^{-1};
(c) 20 m s^{-1};
(d) 48.6 m s^{-1} at 24.3° from vertical

2. Scenario 2:
(a) 51.2 N C^{-1};
(b) 8.20×10^{-18} N down, 4.90×10^{9} m s^{-2} down;
(c) 4.52 μs;
(d) 2.21×10^{4} m s^{-1};
(e) 1.0×10^{4} m s^{-1};
(f) 2.43×10^{4} m s^{-1} at 24.3° from vertical

3. A graphing challenge:





4.3 Exercise 1

1. (a) 5.0×10^{-16} N; (b) 2.6×10^{-3} m

2. 4.4×10^{6} m s^{-1}

4.4.2 Questions

1. (a) 4.0×10^{-3} N; (b) 2.61×10^{-3} N; (c) 90 V

2. (a) The 100 V battery; (b) 1.6×10^{-17} J;
(c) The answer does not change; (d) The answer does not change;
(e) The electrons would not be accelerated, so would not gain any energy; (f) The field strength is doubled.

3. (a) 1.75×10^{17} m s^{-2}; (b) 1.7×10^{-11} s;
(c) 2.57×10^{-3} m; (d) 2.57×10^{3} V

4. (a) $d = \dfrac{El^2q}{2mV^2}$; (b) 8×10^{-13} C

5. (a) 4.6×10^{-14} N;

(c) 5.1×10^{16} m s^{-2}

6. (a) 2.0×10^{-13} N; (b) 2.2×10^{17} m s^{-2}; (c) 1.2×10^{14} m s^{-2}

10. (a) 0.043 mm; (b) 78 mm; (c) 16 mm

11. 1.7 mT

Topic 5 The motor effect

5.2 Exercise 1

1. (a) Out of page;
(b) Upwards in plane of page

2. Upwards

5.3 Exercise 1

1. 1.25 N

2. 2.33 T

5.6.2 Questions

11. (a) 6.8×10^{-3} N, down the page;
(b) 1.5×10^{-4} N, out of the page

12. (a) 12.5 T

13. (a) 6.0×10^{-2} N

14. 1.8×10^{-2} N

15. (a) 1.2×10^{-5} N

16. (a) 1.3×10^{-5} N; (b) 3.2×10^{-6} N; (c) 5.7×10^{-7} N

Topic 6 Electromagnetic induction

6.3 Exercise 1

40 μWb

6.4 Exercise 1

(a) 0.018 V;

(b) Anticlockwise

6.7 Exercise 1

1. 9.9, approximately 10 turns

2. 0.05 or 1 : 20 — a step-up transformer

6.9.2 Questions

3. (a) 3.0 Wb; (b) 2.3×10^{-2} Wb;
(c) 6.0×10^{-6} Wb; (d) 0

11. (a) 1.4×10^{-3} Wb

12. (a) 6.3×10^{-3} Wb; (b) Would be 25 times greater

15. (a) 48 A; (b) 0.6 A

16. (a) 24 A; (b) 220 V

22. (b) 64

23. 16 V

24. (a) 2.0 V: 4; 6.0 V: 12; (b) 2.5 A

25. (a) 400 V; (b) 200 W; (c) 200 W; (d) 10 A

26. (b) 3 (increase)

27. (a) 26.7; (b) 26.7 A

28. 0.020

29. (a) 1.3 V;
(c) 0.8 A

31. (a) 0.24 A; (b) 0.096 V; (c) 500 kV; (d) 2.3×10^{-2} W

32. (a) 7.0×10^{-2}; (b) 220 MW; (c) 6.7×10^{2} A

33. (a) 5.0 A; (b) 400 W; (c) 3.9×10^{3} V

Topic 7 Applications of the motor effect

7.3 Exercise 1

1. 100 N

2. 1 m

7.8.2 Questions

3. (b) 0.24 N;
(c) 2.7×10^{-2} m^2;
(d) 3.6×10^{-2}N m

4. (a) 0.98 N, upwards

Topic 8 Exploring the electromagnetic spectrum

8.2 Exercise 1

6. 2.998×10^{8} m s^{-1}

8.4 Exercise 1

2. 15

8.4 Exercise 2

1. 9.46×10^{15} m

8.8.2 Questions

16. (a) 149.9 million kilometres; (b) 389.7 million kilometres; (c) 6.206 billion kilometres; (d) 11.5 hours plus typing time

18. (b) 299 792 458 m

23. (a) A, B, D

Topic 9 The wave model of light

9.3 Exercise 2

1. 0.0135 m

2. (a) 3.33×10^{-6} m; (b) 9.15°; (c) 0.20 m

9.5 Exercise 1

1. (a) 0.5; (b) 71.6°; (c) 77.1°

9.6.2 Questions

4. Constructive: 1.06 μm, 2.12 μm, 3.18 μm …
Destructive: 0.53 μm, 1.59 μm, 2.65 μm …

5. (b) (i) 0; (ii) 950 nm; (iii) 1266 nm

6. 2.58×10^{-3} m

7. 450 nm

9. (b) 0.177 m for violet light;
(c) 0.290 m for red light;
(d) Separation of second order violet to red = 0.226 m

10. (a) $\frac{1}{2}$ of the original intensity;
(b) 50.7°

11. 69.3°

12. $\frac{1}{8}$ of the original intensity

Topic 10 The quantum model of light

10.2 Exercise 1

1. (a) order 10^{9} Hz; (b) order 10^{-8} m

2. an X-ray

10.2 Exercise 2

1. (a) 4.6×10^{14} Hz; (b) 3.1×10^{-19} J

2. (a) 4.8×10^{15} Hz; (b) 6.2×10^{-8} m; (c) ultraviolet

10.2 Exercise 3

1. 5.0×10^{-7} m

2. 0.85

10.3 Exercise 1

(a) Row 1: 389 nm, 7.70×10^{14} Hz, 3.19 eV, 2.36 V, 3.78×10^{-19} J
Row 2: 524 nm, 5.73×10^{14} Hz, 2.38 eV, 1.54 V, 2.46×10^{-19} J

(c) (i) From this data, $h = 8.3 \times 10^{-34}$ J s $= 5.2 \times 10^{-15}$ eV s; (ii) approximately 8.3×10^{13} Hz;
(iii) approximately 6.9×10^{-20} J

(e) 2.92×10^{-19} J = 1.83 eV

(f) 19 μA

(g) 1.59 V

10.4.2 Questions

1. wavelength decreases, brightness increases

4. (a) 3.96×10^{-7} m; (b) blue

5. (a) 6.5×10^{-7} m;
(b) 4460 K;
(c) 325 nm or 3.25×10^{-7}

9. (a) 5.63 V; (b) 3.74×10^{-18} J; (c) 5.64×10^{15} Hz

10. (a) positively charged;
(c) 4.2×10^{14} Hz;
(d) 6.6×10^{-34} J s
(e) 2.8×10^{-19} J

13. 3.07×10^{-19} J

14. 1.11 V

16. (a) 0.85 μA; (b) 1.0 μA;
(c) 1.0 μA; (d) only light intensity affects current;
(e) 1.7 V

17. (a) 2.57 eV = 4.11×10^{-19} J;
(b) 2.12 V;
(c) 2.45×10^{-7} m

18. 55 photons

19. $\frac{n_{red}}{n_{blue}} = 1.3320$

20. (a) 2.56×10^{-19} J; (b) 2.5 V; (c) 6.9×10^{-19} J

22. (a) 2.86 m; (b) 6.96×10^{-26} J

23. long-wave radio, microwave radio waves, infra-red light, red light, blue light, X-rays, gamma rays

25. 7.96×10^{-15} J

Topic 11 Light and special relativity

11.2 Exercise 1

1. B

2. (a) $0\,\text{m}\,\text{s}^{-1}$; (b) $4\,\text{m}\,\text{s}^{-1}$

3. (a) $5\,\text{km}\,\text{h}^{-1}$; (b) 19 s

11.3 Exercise 3

2. 4.59×10^{13} m

11.5 Exercise 1

4.329 min

11.5 Exercise 2

1 min 48 s

11.5 Exercise 3

0.866c

11.6 Exercise 1

2. 14 100 light years

11.6 Exercise 2

(a) 1934 m;

(b) 209 m

11.7 Exercise 1

5.64×10^{-22} kg m s^{-1}

11.8 Exercise 1

3×10^{16} kg

11.8 Exercise 2

1. 8.89×10^{-26} kg

2. 3.079×10^{-26} kg

11.9 Exercise 1

1. 1.6×10^{-13} J

2. 3.56×10^{-27} kg

11.10.2 Questions

6. 100 $km\,h^{-1}$ towards the front of each car

9. (a) 10 $km\,h^{-1}s^{-1}$ (b) 10 $km\,h^{-1}s^{-1}$

15. c

25. (a) 1.048 28;
(b) 95.4 beats per minute (it beats more slowly)

26. (a) Unchanged;
(b) It would be contracted to 60% of its proper length;
(c) c

27. (a) 2.82 s; (b) 0.7053 m

28. (a) 5 min; (b) 0.78c

29. (a) 3.0 light-years; (b) 4.3 years; (c) 6 years

30. (a) 57 m; (b) 190 m;
(c) 6.3×10^{-7} s; (d) 2.67×10^{-7} s

31. 43.6 m

33. 5.97 km

34. 7.839 microseconds

37. (a) 4.5×10^{17} J; (b) 1.4×10^{19} J;
(c) 6.0×10^{19} J; (d) 1.2×10^{20} J

40. C

42. (a) 117 kg (b) 4.2×10^{18} J

43. 5.4×10^{41} J

44. 2.99×10^{18} kg

45. 2.3×10^{20} J

46. 1.0×10^{4} $m\,s^{-1}$

47. 2.25×10^{16} J

51. 8.18×10^{-31} kg

Topic 12 Elemental origins

12.4 Exercise 2

2. 2.2 MeV

12.8 Exercise 1

1.7×10^{39}

12.9.2 Questions

20. (a) 1.39×10^{30} J; (b) 1.54×10^{14} kg

21. 2.4×10^{45} eV

22. (a) 1.93×10^{22} W; (b) 6.77×10^{12} kg

23. (a) 9.02×10^{37}; (b) 1×10^{9} years

24. (b) 2.83×10^{-30} kg

Topic 13 The structure of the atom

13.2 Exercise 1

(a) 200 $V\,m^{-1}$;

(b) 25 V;

(c) 0.040 N towards the negative plate;

(d) the force would diminish in size

13.2 Exercise 2

1. (a) The top plate would be positive;
(b) 1.04×10^{-10}kg

2. (a) magnetic force is always perpendicular to velocity;
(b) 5.0×10^{3} $m\,s^{-1}$

13.5.2 Questions

4. 2.0×10^{-5} N
7. (a) 4.0×10^{3} N C^{-1} to the left;
(b) 6.4×10^{-16} N to the right;
(c) 6.4×10^{-16} N to the left;
(d) The forces act in opposite directions;
(e) 3.2×10^{-17} J each

8. 4.8×10^{-19} C

9. (a) 200 V m^{-1}; (b) 0.40 N

10. 2.4×10^{-13} N

11. 8.64×10^{-16} N

12. 1.73×10^{-13}N

13. 1.9×10^{19} m s^{-2}

14. 0.49 N
20. (a) 3.0×10^{4}V m^{-1}; (b) 3.0×10^{6} m s^{-1}; (c) 4.8×10^{-15} N
23. ${}^{9}_{4}\text{Be} + {}^{4}_{2}\alpha \rightarrow {}^{12}_{6}\text{C} + {}^{1}_{0}n$

Topic 14 The atom and quantum mechanics

14.2 Exercise 1

1. 9.7×10^{-8}m, 2.05×10^{-18}J

2. It is in the UV region.

14.2 Exercise 2

2.86 eV, 4.57×10^{-19} J

14.3 Exercise 1

1. The electron has the greater wavelength since it has the least momentum: $\lambda = 3.6 \times 10^{-9}$ m

2. The wavelength is reduced.

3. 7.29×10^{7} m s^{-1}

14.3 Exercise 2

1. (a) 6.63×10^{-24}N s; they both have the same momentum;
(b) $E_{\text{photon}} = 1.99 \times 10^{-15}$ J, $E_{\text{electron}} = 2.42 \times 10^{-17}$J; the photon has much more energy

2. (a) $p_{\text{photon}} = 6.7 \times 10^{-25}$N s, $p_{\text{electron}} = 1.9 \times 10^{-23}$N s; the electron has the greater momentum;
(b) $\lambda_{\text{photon}} = 9.9 \times 10^{-10}$ m, $\lambda_{\text{electron}} = 3.5 \times 10^{-11}$ m; the photon has the greater wavelength

3. (a) 28 eV, 4.48×10^{-18} J;
(b) 2.86×10^{-24}N s;
(c) 2.32×10^{-10} m;
(d) it decreases

14.4.2 Questions

2. Ground state energy = −10.4 eV
first excited state energy = −5.5 eV
second excited state energy = −3.7 eV
third excited state energy = −1.6 eV

3. (c) 12.8 eV, 2.6 eV, 0.70 eV, 12.1 eV, 1.9 eV, 10.2 eV;
(d) least energy: $\lambda = 1.77 \times 10^{-6}$ m; greatest energy: $\lambda = 9.7 \times 10^{-8}$ m

4. (c) 5.89×10^{-7} (third excited state to ground state), 3.31×10^{-7}m (second excited state to ground state), 2.85×10^{-7}m (first excited state to ground state) and 2.42×10^{-7}m from the ionisation level.

5. (a) 9.5×10^{-8} m; (b) 4.34×10^{-7} m; (c) 1.28×10^{-6} m

6. (a) 3.88×10^{-7} m, 3.79×10^{-7} m, 3.75×10^{-7} m;
(b) series limits towards 3.64×10^{-7} m

7. (a) 2.12×10^{-10} m; (b) 4.77×10^{-10} m; (c) 8.48×10^{-10} m

8. (a) blue; (b) red; (c) blue

9. 10 different lines

10. (a) 1.21×10^{-7} m, 1.02×10^{-7} m;
(b) 6.57×10^{-7} m, 4.87×10^{-7} m;
(c) 1.88×10^{-6} m, 1.28×10^{-6} m

11. (a) 7.77×10^{-6}m; (b) 9.4×10^{-8}m

12. (a) $n = \infty$;
(b) Lyman: 9.1×10^{-8} m, Balmer: 3.65×10^{-7}m, Paschen: 8.23×10^{-7} m;
(c) 13.6 eV

13. (a) (i) 1.8 eV;
(ii) 4.34×10^{14} Hz, 6.91×10^{-7} m;
(b) 1.41×10^{-7} m

15. Re-emission is in all directions, not just in the original beam direction.

17. Emission spectra consist of discrete lines.

21. The electron has the longer de Broglie wavelength.

22. The slower electron has the greater de Broglie wavelength.

23. (a) $\frac{1}{2}$;
(b) the same factor increase in momentum results in the same factor decrease in wavelength, $\frac{1}{2}$

24. (a) 2.43×10^{-11} m; (b) 2.73×10^{-22} m s^{-1}

25. (a) 2.87×10^{-14} m; (b) 2.03×10^{-13} m

Topic 15 Properties of the nucleus

15.2 Exercise 1

(a) $^{241}_{95}\text{Am} \rightarrow {}^{237}_{93}\text{Np} + {}^{4}_{2}\text{He}$

(b) $^{197}_{78}\text{Pt} \rightarrow {}^{197}_{79}\text{Au} + {}^{0}_{-1}\text{e}$

15.3 Exercise 1

1. (a) 2.5 hours; (b) 5.56 mg

2. (a) 10.0 minutes;
(b) 14.15 minutes;
(c) 20 minutes

3. 2.2 hours, decay constant = 0.693/2.2 = 0.315 per h

15.4 Exercise 1

1.

ISOTOPE	IDENTITY	STABLE?	DECAY
$^{28}_{14}\text{X}$	Si-28	Yes	Does Not decay
$^{42}_{14}\text{X}$	Si-42	No	β−
$^{56}_{28}\text{X}$	Ni-56	No	β+
$^{78}_{28}\text{X}$	Ni-78	No	β−
$^{100}_{50}\text{X}$	Sn-100	No	β+
$^{222}_{90}\text{X}$	Th-222	No	α

2. $^{210}_{82}\text{Pb} \rightarrow {}^{210}_{83}\text{Bi} + {}^{0}_{-1}\text{e} + \bar{\nu}$

$^{210}_{83}\text{Bi} \rightarrow {}^{210}_{84}\text{Po} + {}^{0}_{-1}\text{e} + \bar{\nu}$

$^{210}_{84}\text{Po} \rightarrow {}^{206}_{82}\text{Pb} + {}^{4}_{2}\text{He}$

15.5 Exercise 1

1. 92.16 MeV

2. (a) 0.079 487 amu; (b) 74.04 MeV; (c) 6.170 MeV

3. (a) $^{64}_{29}\text{Cu}$; (b) 559.3 MeV

15.6 Exercise 1

1. $^{45}_{21}\text{Sc}$

2. positron, 3.717 MeV

15.7 Exercise 1

1. ${}^{18}_{9}F$

2. ${}^{238}_{92}U + {}^{1}_{0}n \rightarrow {}^{239}_{92}U$

${}^{238}_{92}U \rightarrow {}^{239}_{93}Np + {}^{0}_{-1}e + \bar{v}$

${}^{239}_{93}Np \rightarrow {}^{239}_{94}Pu + {}^{0}_{-1}e + \bar{v}$

15.7 Exercise 2

1. (a) mass lost = 0.198 26 amu, energy released = 184.7 MeV;
(b) mass lost = 0.193 18 amu, energy released = 179.9 MeV
The first reaction releases slightly more energy (4.8 MeV).

15.7 Exercise 3

1. (a) $6.688\,98 \times 10^{-27}$ kg;
(b) $6.683\,16 \times 10^{-27}$ kg;
(c) 5.82×10^{-30} kg;
(d) 5.2380×10^{-13} J;
(e) 7.8308×10^{-13} J;
(f) 100 J;
(g) 9.4673×10^{-13} J;
(h) The energy released from the kilogram of deuterium in (e) is less than that required to run the 100 W lamp for 30 000 years. The deuterium will only run the lamp for about 24.5 thousand years.

2. 25.73 MeV

15.8.2 Questions

1. (a) 30 protons, 36 neutrons, (b) 90 protons, 140 neutrons
(c) 20 protons, 25 neutrons (d) 14 protons, 17 neutrons

2. (a) ${}^{4}_{2}He$ (b) ${}^{13}_{7}N$ (c) ${}^{234}_{91}Pa$

3. (a) Gold: 79 protons, 118 neutrons
(b) Bismuth: 83 protons, 127 neutrons
(c) Lead: 82 protons, 128 neutrons

6. (a) β− (b) β− (c) α

7. (a) ${}^{226}_{86}Ra \rightarrow {}^{4}_{2}\alpha + {}^{222}_{86}Rn + \text{energy}$
(b) ${}^{214}_{84}Po \rightarrow {}^{4}_{2}\alpha + {}^{210}_{82}Pb + \text{energy}$
(c) ${}^{291}_{95}Am \rightarrow {}^{4}_{2}\alpha + {}^{237}_{93}Np + \text{energy}$

8. (a) ${}^{60}_{27}Co \rightarrow {}^{0}_{-1}e + {}^{60}_{28}Ni + \text{energy}$
(b) ${}^{90}_{38}Sr \rightarrow {}^{0}_{-1}e + {}^{90}_{39}Y + \text{energy}$
(c) ${}^{32}_{15}P \rightarrow {}^{0}_{-1}e + {}^{32}_{16}S + \text{energy}$

9. ${}^{24}_{12}Mg* \rightarrow \gamma + {}^{24}_{12}Mg$

10. (a) ${}^{Z}_{A}X \rightarrow {}^{4}_{2}\alpha + {}^{Z-4}_{A-2}D + \text{energy}$ (b) ${}^{Z}_{A}X \rightarrow {}^{0}_{-1}e + {}^{Z}_{A+1}E + \text{energy}$
(c) ${}^{Z}_{A}X^{*} \rightarrow {}^{Z}_{A}X + \gamma$ (d) ${}^{27}_{13}Al + {}^{2}_{1}H \rightarrow {}^{28}_{14}Si + {}^{1}_{0}n + \text{energy}$
(e) ${}^{22}_{11}Na + {}^{4}_{2}He \rightarrow {}^{25}_{12}Mg + {}^{1}_{1}H + \text{energy}$

13. 8α particles and 6β particles are emitted.

14. (a) 7α decays and 4β decays
(b) 6α decays and 4β decays

15. Astatine-218

16. No, the total changes in atomic number and mass number are identical.

17. 1.9 h

19. 17 190 years

20. 33 mg

21. (a) 27.0, 13.5, 6.75, 3.375, 1.69 kBq
(b) (i) 23 kBq
(ii) 25 kBq

22. 2.0 h

23. 1.6×10^{10}
27. 210 MeV
32. (a) 174 MeV
(b) 3.09×10^{-28} kg
(c) 2.78×10^{-11} J, 174 MeV

33. (a) 174 MeV
(b) 3.10×10^{-28} kg
(c) 2.79×10^{-11} J, 174 MeV

34. (a) (i) 17.6 MeV (ii) 3.14×10^{-29} kg (iii) 2.82×10^{-12} J, 17.6 MeV
(b) (i) 4.03 MeV (ii) 7.19×10^{-30} kg (iii) 6.46×10^{-13} J, 4.03 MeV
(c) (i) 22.4 MeV (ii) 3.99×10^{-29} kg (iii) 3.58×10^{-12} J, 22.4 MeV

*0.000 000 000 000 000 000 000 000 000 000 000 001 s and 0.000 000 000 000 000 000 000 000 000 000 01 s

Topic 16 Deep inside the atom

16.2 Exercise 1

4. X = 12, Y = 7, Z = N

16.3 Exercise 1

1. (a) 40 times, or twenty orbits;
(b) 3.9×10^6 m s^{-1};
(c) 3.2 cm

16.4 Exercise 1

1. (a) zero;
(b) double charmed bottom, +1

2. 6
4. (a) zero
5. gold

16.6.2 Questions

6. 30 mesons, 216 baryons

7. Strange B meson (sb), charge = zero; charm B meson (cb), charge = +1; up B meson (ub), charge = +1; down B meson (db), charge = zero
9. (a) −1; (b) +2

INDEX

M

N

O