CAMBRIDGE
UNIVERSITY PRESS

Physics

for Cambridge International AS & A Level

COURSEBOOK

David Sang, Graham Jones,
Gurinder Chadha & Richard Woodside

Third edition

Cambridge Elevate edition

Cambridge Assessment International Education

Endorsed for full syllabus coverage

CAMBRIDGE
UNIVERSITY PRESS

Physics

for Cambridge International AS & A Level

Coursebook

David Sang, Graham Jones, Gurinder Chadha & Richard Woodside

DEDICATED TEACHER AWARDS

Teachers play an important part in shaping futures. Our Dedicated Teacher Awards recognise the hard work that teachers put in every day.

Thank you to everyone who nominated this year, we have been inspired and moved by all of your stories. Well done to all of our nominees for your dedication to learning and for inspiring the next generation of thinkers, leaders and innovators.

Congratulations to our incredible winner and finalists

Ahmed Saya	Sharon Kong Foong	Abhinandan Bhattacharya	Anthony Chelliah	Candice Green	Jimrey Buntas Dapin
Cordoba School for A-Level, Pakistan	Sunway College, Malaysia	JBCN International School Oshiwara, India	Gateway College, Sri Lanka	St Augustine's College, Australia	University of San Jose-Recoletos, Philippines

For more information about our dedicated teachers and their stories, go to

dedicatedteacher.cambridge.org

〉Contents

〉Introduction

This book covers the entire syllabus of Cambridge International AS & A Level Physics (9702) for examination from 2022. This book is in three parts:

- Chapters 1–15 and P1: the AS Level content, covered in the first year of the course, including a chapter (P1) dedicated to the development of your practical skills
- Chapters 16–31 and P2: the A Level content, including a chapter (P2) dedicated to developing your ability to plan, analyse and evaluate practical investigations
- Appendices of useful formulae, a Glossary and an Index.

The main tasks of a textbook like this are to explain the various concepts of physics that you need to understand, and to provide you with questions that will help you to test your understanding and develop the key skills you need to succeed on this course. You will find a visual guide to the structure of each chapter and the features of this book on the next two pages.

In your study of physics, you will find that certain key concepts are repeated, and that these concepts form 'themes' that link the different areas of physics together. It will help you to progress and gain confidence in your understanding of physics if you take note of these themes. For this Coursebook, these key concepts include:

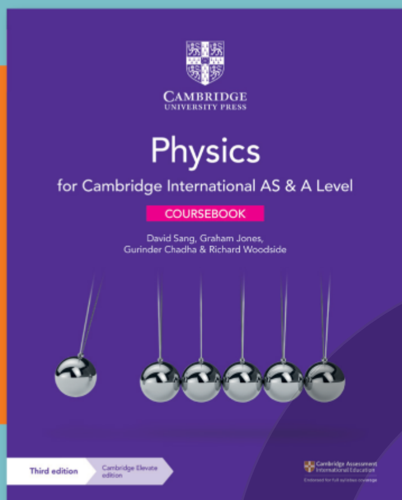
- models of physical systems
- testing predictions against evidence
- mathematics as a language and problem-solving tool
- matter, energy and waves
- forces and fields.

In this Coursebook, the mathematics has been kept to the minimum required by the Cambridge International AS & A Level Physics syllabus (9702). If you are also studying mathematics, you may find that more advanced techniques such as calculus will help you with many aspects of physics.

Studying physics is a stimulating and worthwhile experience. It is an international subject; no single country has a monopoly on the development of the ideas. It can be a rewarding exercise to discover how men and women from many countries have contributed to our knowledge and well-being, through their research into and application of the concepts of physics. We hope not only that this book will help you to succeed in your future studies and career, but also that it will stimulate your curiosity and fire your imagination. Today's students become the next generation of physicists and engineers, and we hope that you will learn from the past to take physics to ever-greater heights.

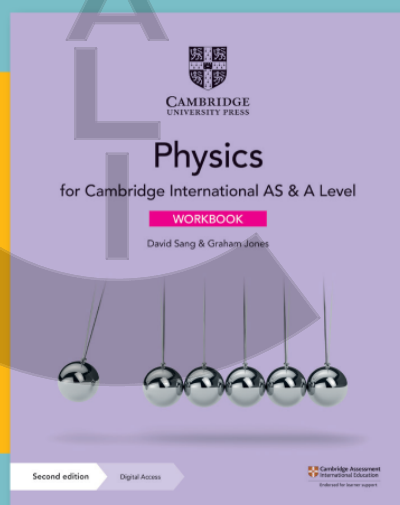
〉How to use this series

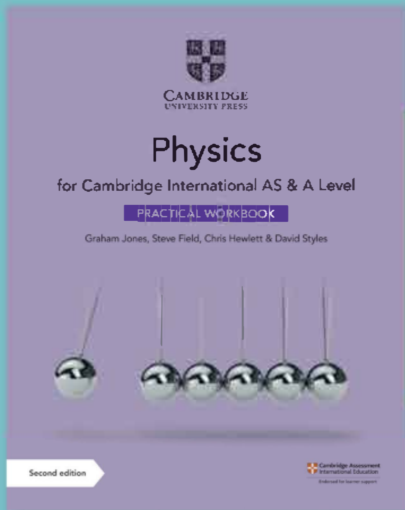
This suite of resources supports students and teachers following the Cambridge International AS & A Level Physics syllabus (9702). All of the books in the series work together to help students develop the necessary knowledge and scientific skills required for this subject.



The coursebook provides comprehensive support for the full Cambridge International AS & A Level Physics syllabus (9702). It clearly explains facts, concepts and practical techniques, and uses real-world examples of scientific principles. Two chapters provide full guidance to help students develop investigative skills. Questions within each chapter help them to develop their understanding, while exam-style questions provide essential practice.

The workbook contains over 100 exercises and exam-style questions, carefully constructed to help learners develop the skills that they need as they progress through their Physics course. The exercises also help students develop understanding of the meaning of various command words used in questions, and provide practice in responding appropriately to these.

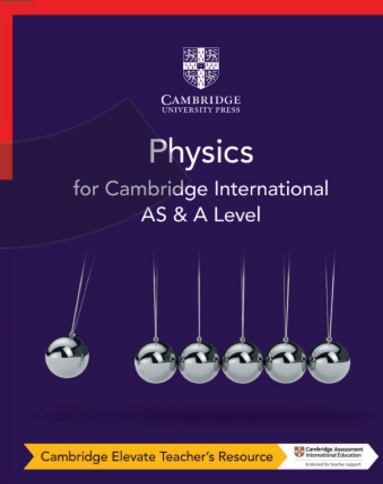


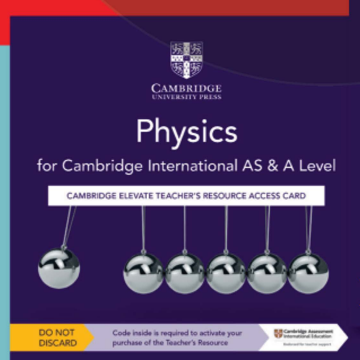


This write-in book provides students with a wealth of hands-on practical work, giving them full guidance and support that will help them to develop all of the essential investigative skills. These skills include planning investigations, selecting and handling apparatus, creating hypotheses, recording and displaying results, and analysing and evaluating data.

The teacher's resource supports and enhances the questions and practical activities in the coursebook. This resource includes detailed lesson ideas, as well as answers and exemplar data for all questions and activities in the coursebook and workbook. The practical teacher's guide, included with this resource, provides support for the practical activities and experiments in the practical workbook.

Teaching notes for each topic area include a suggested teaching plan, ideas for active learning and formative assessment, links to resources, ideas for lesson starters and plenaries, differentiation, lists of common misconceptions and suggestions for homework activities. Answers are included for every question and exercise in the coursebook, workbook and practical workbook.
Detailed support is provided for preparing and carrying out for all the investigations in the practical workbook, including tips for getting things to work well, and a set of sample results that can be used if students cannot do the experiment, or fail to collect results.





〉How to use this book

Throughout this book, you will notice lots of different features that will help your learning. These are explained below.

LEARNING INTENTIONS

These set the scene for each chapter, help with navigation through the Coursebook and indicate the important concepts in each topic.

BEFORE YOU START

This contains questions and activities on subject knowledge you will need before starting this chapter.

SCIENCE IN CONTEXT

This feature presents real-world examples and applications of the content in a chapter, encouraging you to look further into topics. There are discussion questions at the end that look at some of the benefits and problems of these applications.

PRACTICAL ACTIVITIES

This book does not contain detailed instructions for doing particular experiments, but you will find background information about the practical work you need to do in these boxes. There are also two chapters, P1 and P2, which provide detailed information about the practical skills you need to develop during the course.

Questions

Appearing throughout the text, questions give you a chance to check that you have understood the topic you have just read about. You can find the answers to these questions in the digital Coursebook.

KEY EQUATIONS

Key equations are highlighted in the text when an equation is first introduced. Definitions for the equation and further information are given in the margin.

KEY WORDS

Key vocabulary is highlighted in the text when it is first introduced. If you hover your cursor over the word, the definition will appear.

COMMAND WORDS

Command words that appear in the syllabus and might be used in exams are highlighted in the exam-style questions when they are first introduced. If you hover your cursor over the word, the Cambridge International definition will appear.

WORKED EXAMPLES

Wherever you need to know how to use a formula to carry out a calculation, there are worked examples boxes to show you how to do this.

KEY IDEAS

Important scientific concepts, facts and tips are given in these boxes.

REFLECTION

These activities ask you to look back on the topics covered in the chapter and test how well you understand these topics and encourage you to reflect on your learning.

SUMMARY

There is a summary of key points at the end of each chapter.

EXAM-STYLE QUESTIONS

Questions at the end of each chapter provide more demanding exam-style questions, some of which may require use of knowledge from previous chapters. Answers to these questions can be found in the digital Coursebook.

SELF-EVALUATION CHECKLIST

The summary checklists are followed by 'I can' statements that match the Learning intentions at the beginning of the chapter. You might find it helpful to rate how confident you are for each of these statements when you are revising. You should revisit any topics that you rated 'Needs more work' or 'Almost there'.

I can	See topic...	Needs more work	Almost there	Ready to move on

Resource index

The resource index is a convenient place for you to download all answer files for this resource.

〉Chapter 1

Kinematics: describing motion

LEARNING INTENTIONS

In this chapter you will learn how to:

- define and use displacement, speed and velocity
- draw and interpret displacement–time graphs
- describe laboratory methods for determining speed
- understand the differences between scalar and vector quantities and give examples of each
- use vector addition to add and subtract vectors that are in the same plane.

BEFORE YOU START

- Do you know how to rearrange an equation that involves fractions? Choose an equation that you know from your previous physics course, such as $P = \frac{V^2}{R}$ and rearrange it to make *R* or *V* the subject of the formula.
- Can you write down a direction using compass bearings, for example, as 014°, N14°E or 14° east of north?

DESCRIBING MOVEMENT

Our eyes are good at detecting movement. We notice even quite small movements out of the corners of our eyes. It's important for us to be able to judge movement – think about crossing the road, cycling or driving, or catching a ball.

Figure 1.1 shows a way in which movement can be recorded on a photograph. This is a stroboscopic photograph of a boy juggling three balls. As he juggles, a bright lamp flashes several times a second so that the camera records the positions of the balls at equal intervals of time.

How can the photograph be used to calculate the speed of the upper ball horizontally and vertically as it moves through the air? What other apparatus is needed? You can discuss this with someone else.

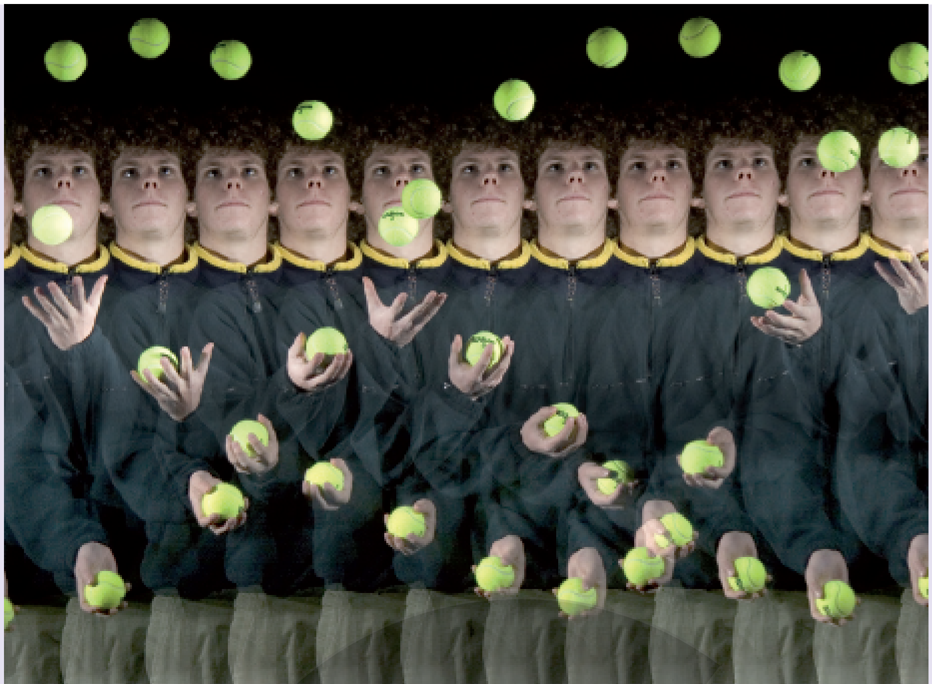


Figure 1.1: This boy is juggling three balls. A stroboscopic lamp flashes at regular intervals; the camera is moved to one side at a steady rate to show separate images of the boy.

1.1 Speed

We can calculate the average speed of something moving if we know the distance it moves and the time it takes:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}$$

In symbols, this is written as:

$$v = \frac{d}{t}$$

where *v* is the average speed and *d* is the distance travelled in time *t*.

If an object is moving at a constant speed, this equation will give us its speed during the time taken. If its speed is changing, then the equation gives us its **average speed**. Average speed is calculated over a period of time.

If you look at the speedometer in a car, it doesn't tell you the car's average speed; rather, it tells you its speed at the instant when you look at it. This is the car's **instantaneous speed**.

KEY EQUATION

$$\text{average speed} = \frac{\text{distance}}{\text{time}}$$

$$v = \frac{d}{t}$$

Question

1 Look at Figure 1.2. The runner has just run 10 000 m in a time of 27 minutes 5.17 s. Calculate his average speed during the race.

Figure 1.2: England's Mo Farah winning his second gold medal at the Rio Olympics in 2016.

Units

In the Système Internationale d'Unités (the SI system), distance is measured in metres (m) and time in seconds (s). Therefore, speed is in metres per second. This is written as $m\ s^{-1}$ (or as m/s). Here, s^{-1} is the same as 1/s, or 'per second'.

There are many other units used for speed. The choice of unit depends on the situation. You would probably give the speed of a snail in different units from the speed of a racing car. Table 1.1 includes some alternative units of speed.

Note that in many calculations it is necessary to work in SI units ($m\ s^{-1}$).

$m\ s^{-1}$	metres per second
$cm\ s^{-1}$	centimetres per second
$km\ s^{-1}$	kilometres per second
$km\ h^{-1}$ or km/h	kilometres per hour
mph	miles per hour

Table 1.1: Units of speed.

Questions

2 Here are some units of speed:

$m\ s^{-1}$ $mm\ s^{-1}$ $km\ s^{-1}$ $km\ h^{-1}$

Which of these units would be appropriate when stating the speed of each of the following?

- **a** a tortoise
- **b** a car on a long journey
- **c** light
- **d** a sprinter.

3 A snail crawls 12 cm in one minute. What is its average speed in $mm\ s^{-1}$?

Determining speed

You can find the speed of something moving by measuring the time it takes to travel between two fixed points. For example, some motorways have emergency telephones every 2000 m. Using a stopwatch you can time a car over this distance. Note that this can only tell you the car's average speed between the two points. You cannot tell whether it was increasing its speed, slowing down or moving at a constant speed.

PRACTICAL ACTIVITY 1.1

Laboratory measurements of speed

Here we describe four different ways to measure the speed of a trolley in the laboratory as it travels along a straight line. Each can be adapted to measure the speed of other moving objects, such as a glider on an air track or a falling mass.

Measuring speed using two light gates

The leading edge of the card in Figure 1.3 breaks the light beam as it passes the first light gate. This starts the timer. The timer stops when the front of the card breaks the second beam. The trolley's speed is calculated from the time interval and the distance between the light gates.

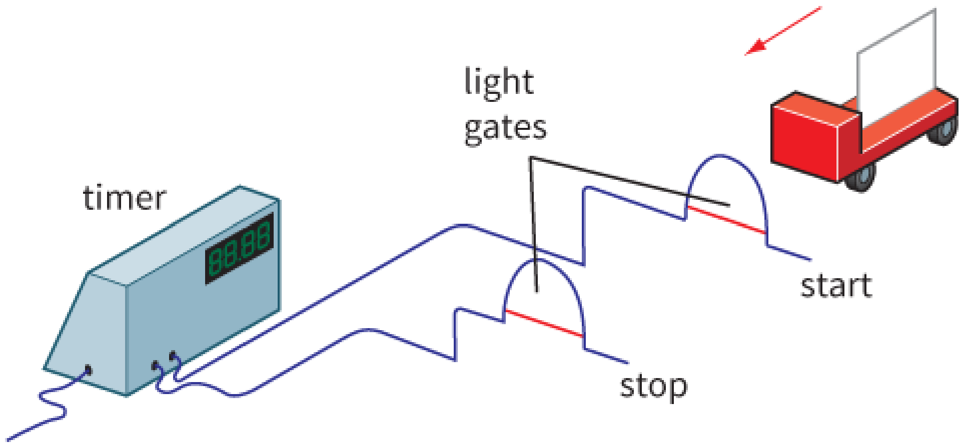


Figure 1.3: Using two light gates to find the average speed of a trolley.

Measuring speed using one light gate

The timer in Figure 1.4 starts when the leading edge of the card breaks the light beam. It stops when the trailing edge passes through. In this case, the time shown is the time taken for the trolley to travel a distance equal to the length of the card. The computer software can calculate the speed directly by dividing the distance by the time taken.

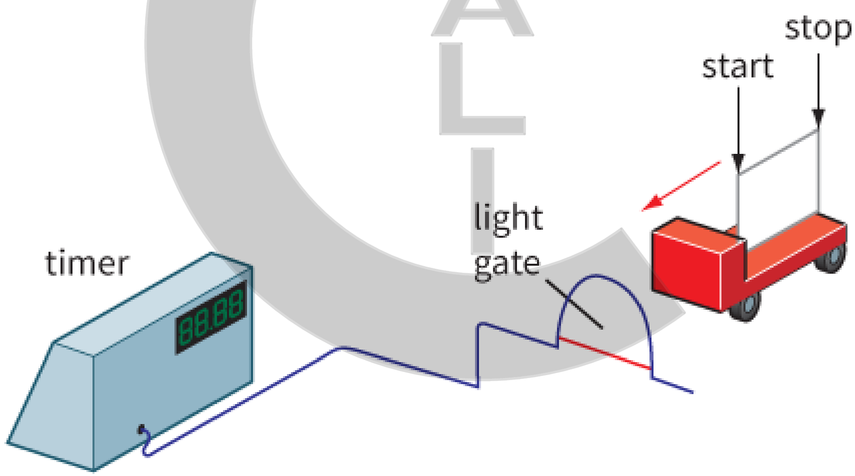


Figure 1.4: Using a single light gate to find the average speed of a trolley.

Measuring speed using a ticker-timer

The ticker-timer (Figure 1.5) marks dots on the tape at regular intervals, usually s (i.e. 0.02 s). (This is because it works with alternating current, and in most countries the frequency of the alternating mains is 50 Hz.) The pattern of dots acts as a record of the trolley’s movement.

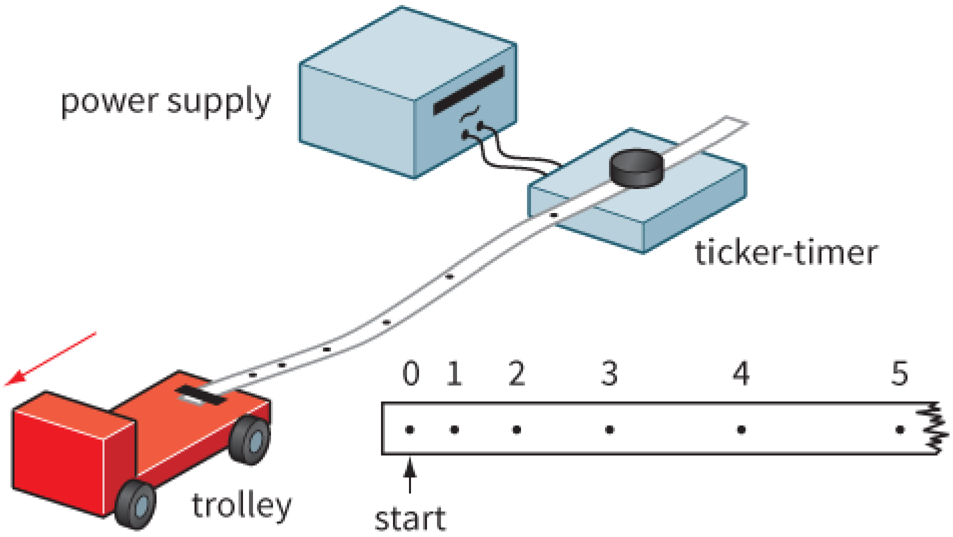


Figure 1.5: Using a ticker-timer to investigate the motion of a trolley.

Start by inspecting the tape. This will give you a description of the trolley's movement. Identify the start of the tape. Then, look at the spacing of the dots:

- even spacing – constant speed
- increasing spacing – increasing speed.

Now you can make some measurements. Measure the distance of every fifth dot from the start of the tape. This will give you the trolley's distance at intervals of 0.10 s. Put the measurements in a table and draw a distance–time graph.

Measuring speed using a motion sensor

The motion sensor (Figure 1.6) transmits regular pulses of ultrasound at the trolley. It detects the reflected waves and determines the time they took for the trip to the trolley and back. From this, the computer can deduce the distance to the trolley from the motion sensor. It can generate a distance–time graph. You can determine the speed of the trolley from this graph.

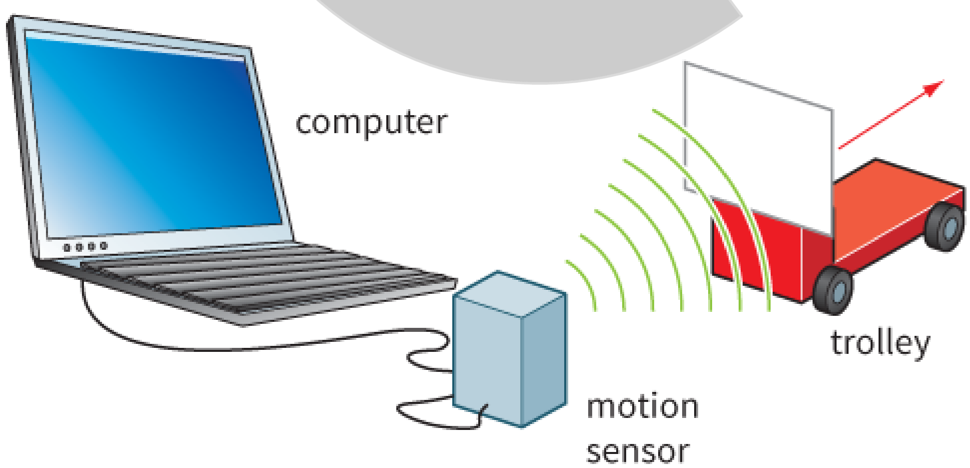


Figure 1.6: Using a motion sensor to investigate the motion of a trolley.

Choosing the best method

Each of these methods for finding the speed of a trolley has its merits. In choosing a method, you might think about the following points:

- Does the method give an average value of speed or can it be used to give the speed of the trolley at different points along its journey?
- How precisely does the method measure time–to the nearest millisecond?
- How simple and convenient is the method to set up in the laboratory?

Questions

4 A trolley with a 5.0 cm long card passed through a single light gate. The time recorded by a digital timer was 0.40 s. What was the average speed of the trolley in m s^{-1}?

5 Figure 1.7 shows two ticker-tapes. Describe the motion of the trolleys that produced them.

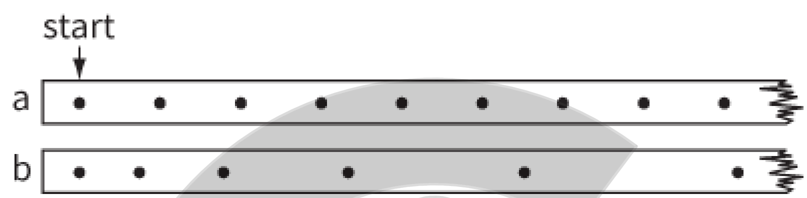


Figure 1.7: Two ticker-tapes. For Question 5.

6 Four methods for determining the speed of a moving trolley have been described. Each could be adapted to investigate the motion of a falling mass. Choose two methods that you think would be suitable, and write a paragraph for each to say how you would adapt it for this purpose.

1.2 Distance and displacement, scalar and vector

In physics, we are often concerned with the distance moved by an object in a particular direction. This is called its **displacement**.

Figure 1.8 illustrates the difference between distance and displacement. It shows the route followed by walkers as they went from town A to town C.

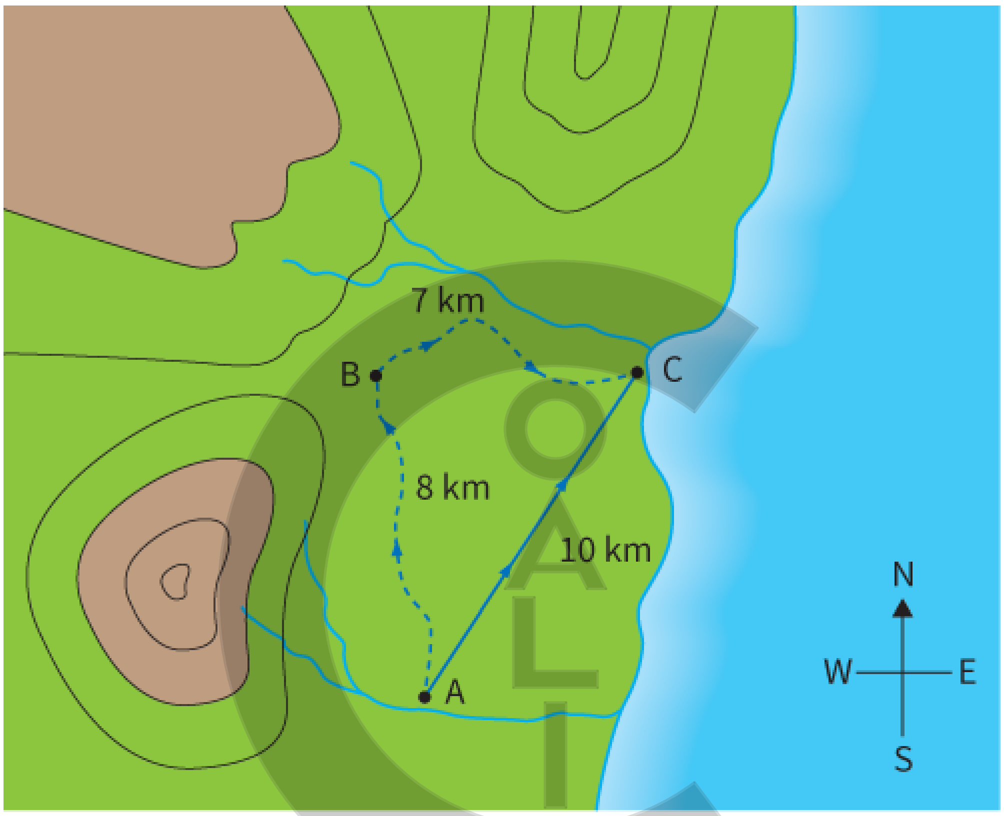


Figure 1.8: If you go on a long walk, the distance you travel will be greater than your displacement. In this example, the walkers travel a distance of 15 km, but their displacement is only 10 km, because this is the distance from the start to the finish of their walk.

Their winding route took them through town B, so that they covered a total distance of 15 km. However, their displacement was much less than this. Their finishing position was just 10 km from where they started. To give a complete statement of their displacement, we need to give both distance and direction:

$$\text{displacement} = 10\ \text{km at } 030^\circ \text{ or } 30^\circ \text{ E of N}$$

Displacement is an example of a **vector quantity**. A vector quantity has both magnitude (size) and direction. Distance, on the other hand, is a **scalar quantity**. Scalar quantities have magnitude only.

1.3 Speed and velocity

It is often important to know both the speed of an object and the direction in which it is moving.

Speed and direction are combined in another quantity, called **velocity**. The velocity of an object can be thought of as its speed in a particular direction. So, like displacement, velocity is a vector quantity. Speed is the corresponding scalar quantity, because it does not have a direction.

So, to give the velocity of something, we have to state the direction in which it is moving. For example, 'an aircraft flies with a velocity of 300 m s^{-1} due north'.

Since velocity is a vector quantity, it is defined in terms of displacement:

$$\text{velocity} = \frac{\text{change in displacement}}{\text{time taken}}$$

We can write the equation for velocity in symbols:

$$v = \frac{s}{t}$$

KEY EQUATION

$$\text{velocity} = \frac{\text{change in displacement}}{\text{time taken}}$$

Alternatively, we can say that velocity is the *rate of change* of an object's displacement:

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

where the symbol Δ (the Greek letter delta) means 'change in'. It does not represent a quantity (in the way that s and t do). Another way to write Δs would be $s_2 - s_1$, but this is more time-consuming and less clear.

From now on, you need to be clear about the distinction between velocity and speed, and between displacement and distance. Table 1.2 shows the standard symbols and units for these quantities.

Quantity	Symbol for quantity	Symbol for unit
distance	*d*	m
displacement	*s, x*	m
time	*t*	s
speed, velocity	*v*	m s^{-1}

Table 1.2: Standard symbols and units. (Take care not to confuse italic *s* for displacement with s for seconds. Notice also that *v* is used for both speed and velocity.)

Question

7 Do these statements describe speed, velocity, distance or displacement? (Look back at the definitions of these quantities.)

a The ship sailed south-west for 200 miles.

b I averaged 7 mph during the marathon.

c The snail crawled at 2 mm s^{-1} along the straight edge of a bench.
d The sales representative's round trip was 420 km.

Speed and velocity calculations

The equation for velocity, $v = \frac{\Delta s}{\Delta t}$, can be rearranged as follows, depending on which quantity we want to determine:

$$\text{change in displacement } \Delta s = v \times \Delta t$$
$$\text{change in time } \Delta t = \frac{\Delta s}{v}$$

Note that each of these equations is balanced in terms of units. For example, consider the equation for displacement. The units on the right-hand side are m s^{-1} × s, which simplifies to m, the correct unit for displacement.

We can also rearrange the equation to find distance *s* and time *t*:

$$\Delta s = v \times t$$
$$t = \frac{\Delta s}{v}$$

WORKED EXAMPLES

1 A car is travelling at 15 m s^{-1}. How far will it travel in 1 hour?

Step 1 It is helpful to start by writing down what you know and what you want to know:

v = 15 m s^{-1}

t = 1 h = 3600 s

s = ?

Step 2 Choose the appropriate version of the equation and substitute in the values. Remember to include the units:

$$\begin{aligned} s &= v \times t \\ &= 15 \times 3600 \\ &= 5.4 \times 10^4 \text{ m} \\ &= 54 \text{ km} \end{aligned}$$

The car will travel 54 km in 1 hour.

2 The Earth orbits the Sun at a distance of 150 000 000 km. How long does it take light from the Sun to reach the Earth? (Speed of light in space = 3.0 × 10^8 m s^{-1}.)

Step 1 Start by writing what you know. Take care with units; it is best to work in m and s. You need to be able to express numbers in scientific notation (using powers of 10) and to work with these on your calculator.

$$\begin{aligned} v &= 3.0 \times 10^8 \text{ m s}^{-1} \\ s &= 150\,000\,000 \text{ km} \\ &= 150\,000\,000 \text{ m} \\ &= 1.5 \times 10^{11} \text{ m} \end{aligned}$$

Step 2 Substitute the values in the equation for time:

$$t = \frac{s}{v}$$
$$= \frac{1.5\times10^{11}}{3.0\times10^{8}}$$
$$= 500\ \text{s}$$

Light takes 500 s (about 8.3 minutes) to travel from the Sun to the Earth.

Hint: *When using a calculator, to calculate the time t, you press the buttons in the following sequence:*

[1.5] [10^n] [11] [÷] [3] [10^n] [8]

Making the most of units

In Worked example 1 and Worked example 2, units have been omitted in intermediate steps in the calculations. However, at times it can be helpful to include units as this can be a way of checking that you have used the correct equation; for example, that you have not divided one quantity by another when you should have multiplied them. The units of an equation must be balanced, just as the numerical values on each side of the equation must be equal.

If you take care with units, you should be able to carry out calculations in non-SI units, such as kilometres per hour, without having to convert to metres and seconds.

For example, how far does a spacecraft travelling at 40 000 km h^{-1} travel in one day? Since there are 24 hours in one day, we have:

$$\text{distance travelled} = 40\,000\ \text{km h}^{-1} \times 24\ \text{h}$$
$$= 960\,000\ \text{km}$$

Questions

8 A submarine uses sonar to measure the depth of water below it. Reflected sound waves are detected 0.40 s after they are transmitted. How deep is the water? (Speed of sound in water = 1500 m s^{-1}.)

9 The Earth takes one year to orbit the Sun at a distance of 1.5×10^{11} m. Calculate its speed. Explain why this is its average speed and not its velocity.

1.4 Displacement–time graphs

We can represent the changing position of a moving object by drawing a displacement–time graph. The gradient (slope) of the graph is equal to its velocity (Figure 1.9). The steeper the slope, the greater the velocity. A graph like this can also tell us if an object is moving forwards or backwards. If the gradient is negative, the object's velocity is negative – it is moving backwards.

Deducing velocity from a displacement–time graph

A toy car moves along a straight track. Its displacement at different times is shown in Table 1.3. This data can be used to draw a displacement–time graph from which we can deduce the car's velocity.

Displacement s / m	1.0	3.0	5.0	7.0	7.0	7.0
Time t / s	0.0	1.0	2.0	3.0	4.0	5.0

Table 1.3: Displacement s and time t data for a toy car.

It is useful to look at the data first, to see the pattern of the car's movement. In this case, the displacement increases steadily at first, but after 3.0 s it becomes constant. In other words, initially the car is moving at a steady velocity, but then it stops.

The straight line shows that the object's velocity is constant.

The slope shows which object is moving faster. The steeper the slope, the greater the velocity.

The slope of this graph is 0. The displacement *s* is not changing. Hence the velocity *v* = 0. The object is stationary.

The slope of this graph suddenly becomes negative. The object is moving back the way it came. Its velocity *v* is negative after time *T*.

This displacement–time graph is curved. The slope is changing. This means that the object's velocity is changing – this is considered in Chapter 2.

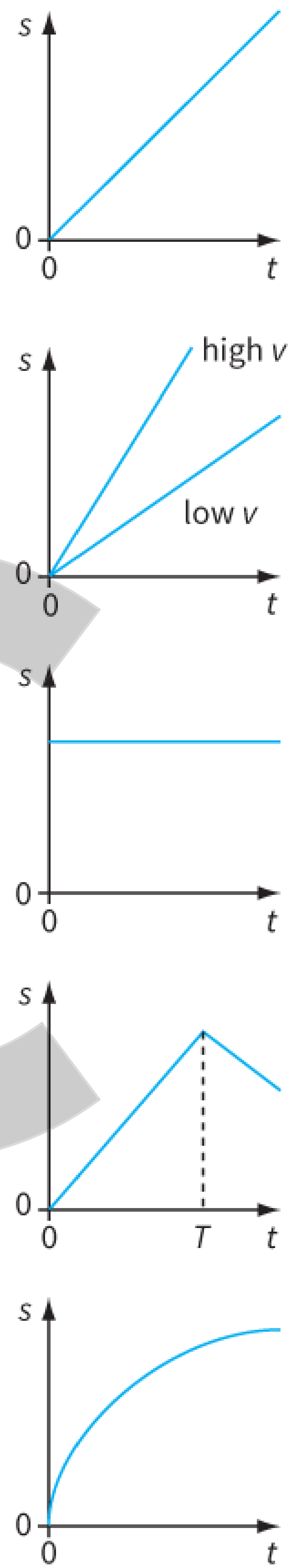


Figure 1.9: The slope of a displacement–time (*s*–*t*) graph tells us how fast an object is moving.

Now we can plot the displacement–time graph (Figure 1.10).

We want to work out the velocity of the car over the first 3.0 seconds. We can do this by working out the gradient of the graph, because:

velocity = gradient of displacement−time graph

We draw a right-angled triangle as shown. To find the car's velocity, we divide the change in displacement by the change in time. These are given by the two sides of the triangle labelled Δs and Δt.

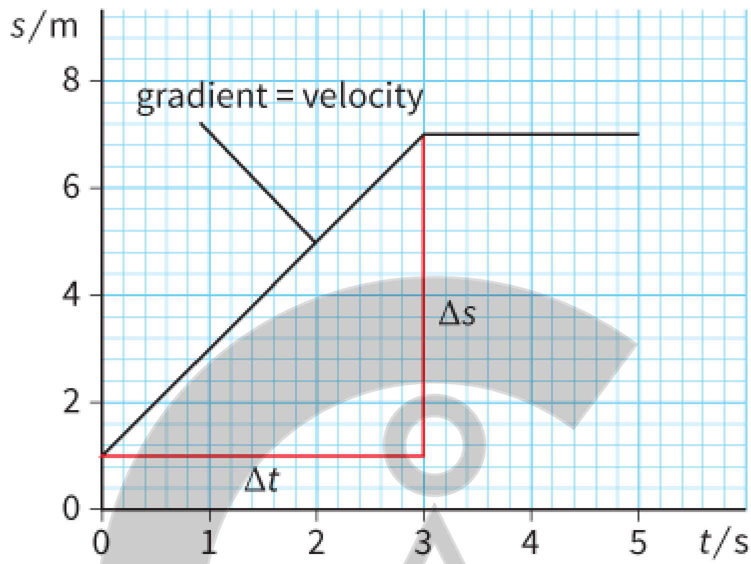


Figure 1.10: Displacement–time graph for a toy car; data as shown in Table 1.3.

$$\text{velocity} = \frac{\text{change in displacement}}{\text{time taken}}$$

$$= \frac{\Delta s}{\Delta t}$$

$$= \frac{(7.0-1.0)}{(3.0-0)}$$

$$= \frac{6.0}{3.0}$$

$$= 2.0\ \text{m s}^{-1}$$

If you are used to finding the gradient of a graph, you may be able to reduce the number of steps in this calculation.

Questions

10 The displacement–time sketch graph in Figure 1.11 represents the journey of a bus. What does the graph tell you about the journey?

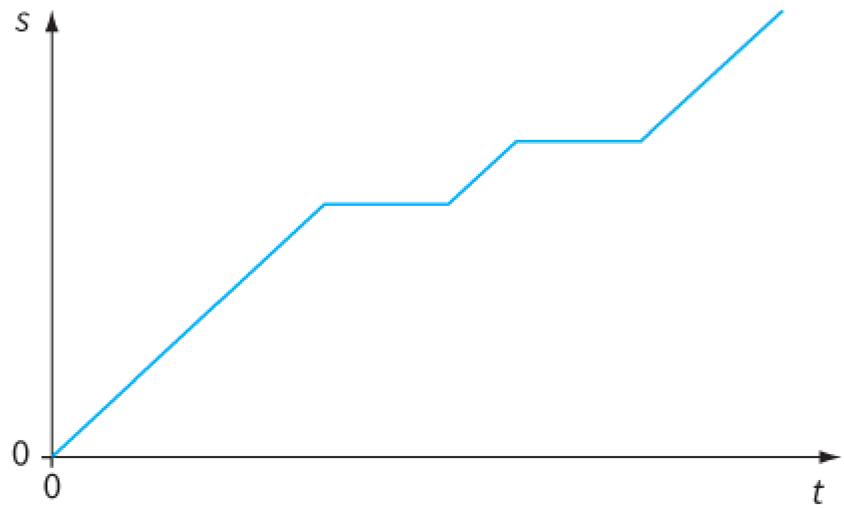


Figure 1.11: For Question 10.

11 Sketch a displacement–time graph to show your motion for the following event. You are walking at a constant speed across a field after jumping off a gate. Suddenly you see a horse and stop. Your friend says there's no danger, so you walk on at a reduced constant speed. The horse neighs, and you run back to the gate. Explain how each section of the walk relates to a section of your graph.

12 Table 1.4 shows the displacement of a racing car at different times as it travels along a straight track during a speed trial.

a Determine the car's velocity.

b Draw a displacement–time graph and use it to find the car's velocity.

Displacement / m	0	85	170	255	340
Time / s	0	1.0	2.0	3.0	4.0

Table 1.4: Displacement *s* and time *t* data for Question 12.

13 An old car travels due south. The distance it travels at hourly intervals is shown in Table 1.5.

a Draw a distance–time graph to represent the car's journey.

b From the graph, deduce the car's speed in km h^{-1} during the first three hours of the journey.

c What is the car's average speed in km h^{-1} during the whole journey?

Time / h	0	1	2	3	4
Distance / km	0	23	46	69	84

Table 1.5: Data for Question 13.

1.5 Combining displacements

The walkers shown in Figure 1.12 are crossing difficult ground. They navigate from one prominent point to the next, travelling in a series of straight lines. From the map, they can work out the distance that they travel and their displacement from their starting point:

distance travelled = 25 km

Figure 1.12: In rough terrain, walkers head straight for a prominent landmark.

(Lay thread along route on map; measure thread against map scale.)

displacement = 15 km in the direction 045°, N45° E or north-east

(Join starting and finishing points with straight line; measure line against scale.)

A map is a scale drawing. You can find your displacement by measuring the map. But how can you **calculate** your displacement? You need to use ideas from geometry and trigonometry. Worked examples 3 and 4 show how.

WORKED EXAMPLES

3 A spider runs along two sides of a table (Figure 1.13). Calculate its final displacement.

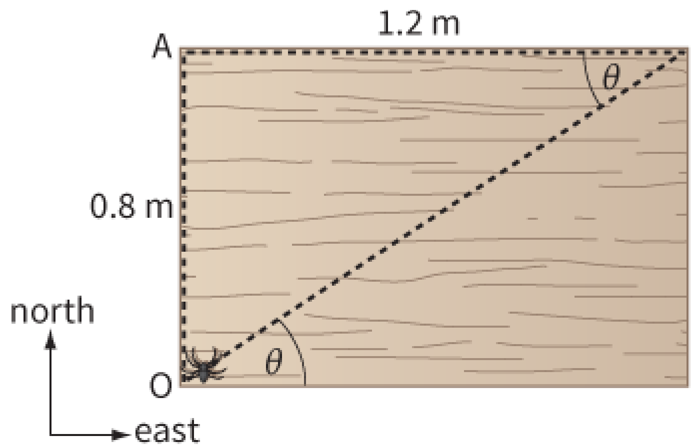


Figure 1.13: The spider runs a distance of 2.0 m. For Worked example 3.

Step 1 Because the two sections of the spider's run (OA and AB) are at right angles, we can **add** the two displacements using Pythagoras's theorem:

$$OB^2 = OA^2 + AB^2$$
$$= 0.8^2 + 1.2^2 = 2.08$$
$$OB = \sqrt{2.08} = 1.44 \text{ m} \approx 1.4 \text{ m}$$

Step 2 Displacement is a vector. We have found the **magnitude** of this vector, but now we have to find its direction. The angle θ is given by:

$$\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{0.8}{1.2}$$
$$= 0.667$$
$$\theta = \tan^{-1}(0.667)$$
$$= 33.7° \approx 34°$$

So the spider's displacement is 1.4 m at 056° or N56°E or at an angle of 34° north of east.

4 An aircraft flies 30 km due east and then 50 km north-east (Figure 1.14). Calculate the final displacement of the aircraft.

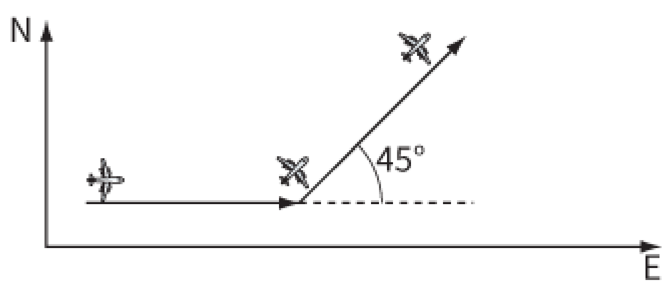


Figure 1.14: For Worked example 4.

Here, the two displacements are not at 90° to one another, so we can't use Pythagoras's theorem. We can solve this problem by making a scale drawing, and measuring the final displacement. (However, you could solve the same problem using trigonometry.)

Step 1 Choose a suitable scale. Your diagram should be reasonably large; in this case, a scale of 1 cm to represent 5 km is reasonable.

Step 2 Draw a line to represent the first vector. North is at the top of the page. The line is 6 cm long, towards the east (right).

Step 3 Draw a line to represent the second vector, starting at the end of the first vector. The line is 10 cm long, and at an angle of 45° (Figure 1.15).

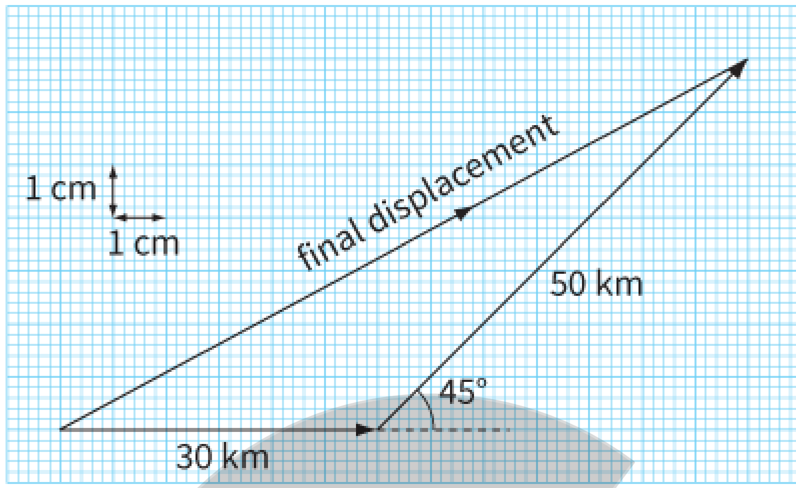


Figure 1.15: Scale drawing for Worked example 4. Using graph paper can help you to show the vectors in the correct directions.

Step 4 To find the final displacement, join the start to the finish. You have created a **vector triangle**. Measure this displacement vector, and use the scale to convert back to kilometres:

length of vector = 14.8 cm

final displacement = 14.8 × 5 = 74 km

Step 5 Measure the angle of the final displacement vector:

angle = 28° N of E

Therefore the aircraft's final displacement is 74 km at 28° north of east, 062° or N62°E.

Questions

14 You walk 3.0 km due north, and then 4.0 km due east.

- **a** Calculate the total distance in km you have travelled.
- **b** Make a scale drawing of your walk, and use it to find your final displacement. Remember to give both the magnitude and the direction.
- **c** Check your answer to part **b** by calculating your displacement.

15 A student walks 8.0 km south-east and then 12 km due west.

- **a** Draw a vector diagram showing the route. Use your diagram to find the total displacement.

  Remember to give the scale on your diagram and to give the direction as well as the magnitude of your answer.
- **b** Calculate the resultant displacement. Show your working clearly.

This process of adding two displacements together (or two or more of any type of vector) is known as vector addition. When two or more vectors are added together, their combined effect is known as the **resultant** of the vectors.

COALI

1.6 Combining velocities

Velocity is a vector quantity and so two velocities can be combined by vector addition in the same way that we have seen for two or more displacements.

Imagine that you are attempting to swim across a river. You want to swim directly across to the opposite bank, but the current moves you sideways at the same time as you are swimming forwards. The outcome is that you will end up on the opposite bank, but downstream of your intended landing point. In effect, you have two velocities:

- the velocity due to your swimming, which is directed straight across the river
- the velocity due to the current, which is directed downstream, at right angles to your swimming velocity.

These combine to give a resultant (or net) velocity, which will be diagonally downstream. In order to swim directly across the river, you would have to aim upstream. Then your resultant velocity could be directly across the river.

WORKED EXAMPLE

5 An aircraft is flying due north with a velocity of 200 m s^{-1}. A side wind of velocity 50 m s^{-1} is blowing due east. What is the aircraft's resultant velocity (give the magnitude and direction)?

Here, the two velocities are at 90°. A sketch diagram and Pythagoras's theorem are enough to solve the problem.

Step 1 Draw a sketch of the situation – this is shown in Figure 1.16a.

Step 2 Now sketch a vector triangle. Remember that the second vector starts where the first one ends. This is shown in Figure 1.16b.

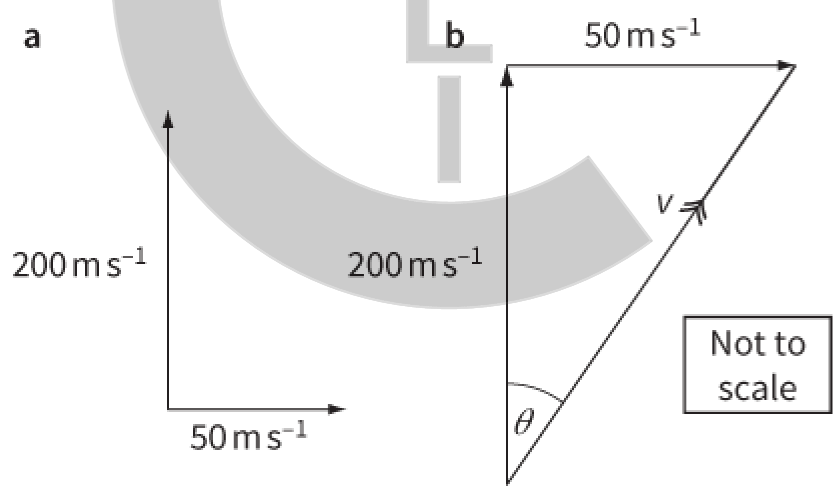


Figure 1.16: Finding the resultant of two velocities. For Worked example 5.

Step 3 Join the start and end points to complete the triangle.

Step 4 Calculate the magnitude of the resultant vector *v* (the hypotenuse of the right-angled triangle).

$$v^2 = 200^2 + 50^2 = 40\,000 + 2500 = 42\,500$$

$$v = \sqrt{42\,500} \approx 206\ \text{m s}^{-1}$$

Step 5 Calculate the angle θ:

$$\tan\theta = \frac{50}{200} = 0.25$$

$\theta = \tan^{-1}(0.25) \approx 14°$

So the aircraft's resultant velocity is 206 m s^{-1} at 14° east of north, 076° or N76°E.

Questions

16 A swimmer can swim at 2.0 m s^{-1} in still water. She aims to swim directly across a river that is flowing at 0.80 m s^{-1}. Calculate her resultant velocity. (You must give both the magnitude and the direction.)

17 A stone is thrown from a cliff and strikes the surface of the sea with a vertical velocity of 18 m s^{-1} and a horizontal velocity v. The resultant of these two velocities is 25 m s^{-1}.

a Draw a vector diagram showing the two velocities and the resultant.

b Use your diagram to find the value of v.

c Use your diagram to find the angle between the stone and the vertical as it strikes the water.

1.7 Subtracting vectors

Sometimes, vectors need to be subtracted rather than added. For example, if you are in a car moving at 2.0 m s^{-1} and another car on the same road is moving in the same direction at 5.0 m s^{-1}, then you approach the car at 5.0 – 2.0 = 3.0 m s^{-1}. You are subtracting two velocity vectors.

Subtraction of vectors can be done using the formula:

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$$

where **A** and **B** are vectors.

KEY IDEA

To subtract a vector, add on the vector to be subtracted in the opposite direction.

So, to subtract, just add the negative vector.

But first you have to understand what the negative of vector **B** means. The negative of vector **B** is another vector of the same size as **B** but in the opposite direction.

This is straightforward if the velocities are in the same direction. For example, to subtract a velocity of 4 m s^{-1} north from a velocity of 10 m s^{-1} north, you start by drawing a vector 10 m s^{-1} north and then add a vector of 4 m s^{-1} south. The answer is 6 m s^{-1} north.

It is less straightforward if the velocities are in the opposite direction. For example, to subtract a velocity of 4 m s^{-1} south from a velocity of 10 m s^{-1} north, you start by drawing a vector 10 m s^{-1} north and then add a vector of 4 m s^{-1} north. The answer is 14 m s^{-1} north.

The example in Figure 1.17 shows how to find **A − B** and **A + B** when the vectors are along different directions.

Question

18 A velocity of 5.0 m s^{-1} is due north. Subtract from this velocity another velocity that is:

- **a** 5.0 m s^{-1} due south
- **b** 5.0 m s^{-1} due north

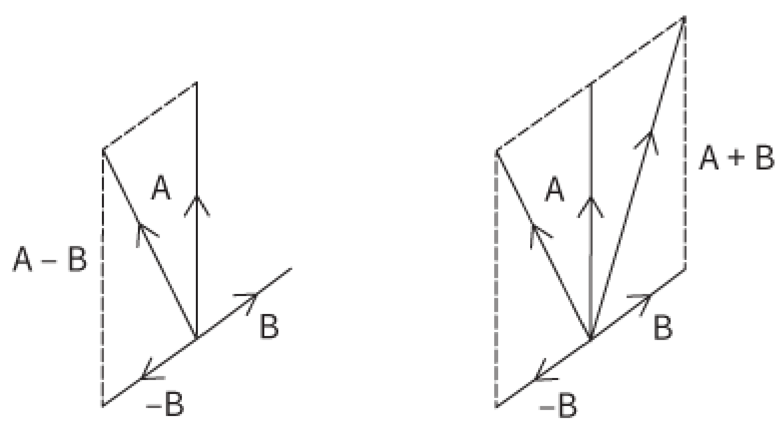


Figure 1.17: Subtracting and adding two vectors **A** and **B** in different directions.

c 5.0 m s^{-1} due west

d 5.0 m s^{-1} due east

(You can do a scale drawing or make a calculation but remember to give the direction of your answers as well as their size.)

1.8 Other examples of scalar and vector quantities

Direction matters when vectors are combined. You can use this to decide whether a quantity is a vector or a scalar. For example, if you walk for 3 minutes north and then 3 minutes in another direction, the total time taken is 6 minutes whatever direction you choose. A vector of 3 units added to another vector of 3 units can have any value between 0 and 6 but two scalars of 3 units added together always make six units. So, time is a scalar.

Mass and density are also both scalar quantities.

Force and acceleration, as you will see in later chapters, are both vector quantities. This is because, if an object is pushed with the same force in two opposite directions, the forces cancel out.

Work and pressure, which you will also study in later chapters, both involve force. However, work and pressure are both scalar quantities. For example, if you pull a heavy case along the floor north and then the same distance south, the total work done is clearly not zero. You just add scalar quantities even if they are in the opposite direction.

REFLECTION

- Write down anything that you found interesting or challenging in this chapter.
- Look at your notes later when you revise this topic.

SUMMARY

Displacement is the distance travelled in a particular direction.
Velocity is defined by the word equation: $$\text{velocity} = \frac{\text{change in displacement}}{\text{time taken}}$$
The gradient of a displacement–time graph is equal to velocity: $$v = \frac{\Delta s}{\Delta t}$$
Distance, speed, mass and time are scalar quantities. A scalar quantity has only magnitude.
Displacement and velocity are vector quantities. A vector quantity has both magnitude and direction.
Vector quantities may be combined by vector addition to find their resultant. The second vector can be subtracted from the first by adding the negative of the second vector, which acts in the opposite direction.

EXAM-STYLE QUESTIONS

1 Which of the following pairs contains one vector and one scalar quantity? **[1]**

A displacement : mass

B displacement : velocity

C distance : speed

D speed : time

2 A vector **P** of magnitude 3.0 N acts towards the right and a vector **Q** of magnitude 4.0 N acts upwards. **[1]**

What is the magnitude and direction of the vector (**P** − **Q**)?

A 1.0 N at an angle of 53° downwards to the direction of **P**

B 1.0 N at an angle of 53° upwards to the direction of **P**

C 5.0 N at an angle of 53° downwards to the direction of **P**

D 5.0 N at an angle of 53° upwards to the direction of **P**

3 A car travels one complete lap around a circular track at a constant speed of 120 km h^{-1}.

a If one lap takes 2.0 minutes, show that the length of the track is 4.0 km. **[2]**

b **Explain** why values for the average speed and average velocity are different. **[1]**

c **Determine** the magnitude of the displacement of the car in a time of 1.0 min. **[2]**

(The circumference of a circle = $2\pi R$, where R is the radius of the circle.) **[Total: 5]**

4 A boat leaves point A and travels in a straight line to point B. The journey takes 60 s.

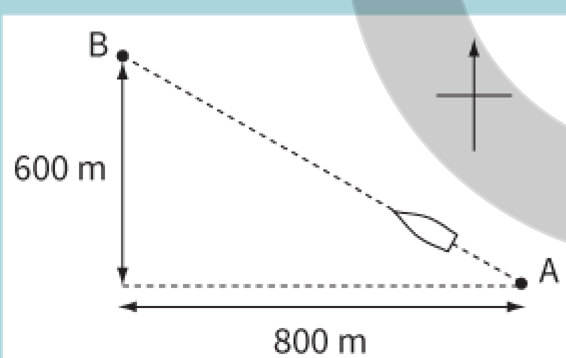


Figure 1.18

Calculate:

a the distance travelled by the boat **[2]**

b the total displacement of the boat **[2]**

c the average velocity of the boat. **[2]**

Remember that each vector quantity must be given a direction as well as a magnitude. **[Total: 6]**

5 A boat travels at 2.0 m s^{-1} east towards a port, 2.2 km away. When the boat reaches the port, the passengers travel in a car due north for 15 minutes at 60 km h^{-1}.

Calculate:

a the total distance travelled **[2]**

b the total displacement **[3]**

c the total time taken **[2]**

d the average speed in m s^{-1} **[2]**

e the magnitude of the average velocity. **[2]**

[Total: 11]

6 A river flows from west to east with a constant velocity of 1.0 m s^{-1}. A boat leaves the south bank heading due north at 2.4 m s^{-1}. Find the resultant velocity of the boat. **[3]**

7 a **Define** displacement. **[1]**

b Use the definition of displacement to explain how it is possible for an athlete to run round a track yet have no displacement. **[2]**

[Total: 6]

8 A girl is riding a bicycle at a constant velocity of 3.0 m s^{-1} along a straight road. At time t = 0, she passes her brother sitting on a stationary bicycle. At time t = 0, the boy sets off to catch up with his sister. His velocity increases from time t = 0 until t = 5.0 s, when he has covered a distance of 10 m. He then continues at a constant velocity of 4.0 m s^{-1}.

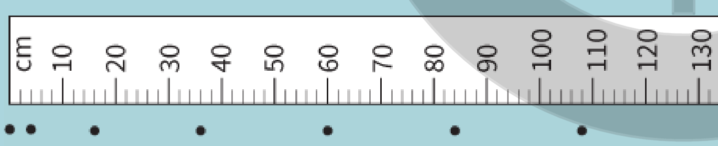
a Draw the displacement–time graph for the girl from t = 0 to t = 12 s. **[1]**

b On the same graph axes, draw the displacement–time graph for the boy. **[2]**

c Using your graph, determine the value of t when the boy catches up with his sister. **[1]**

[Total: 4]

9 A student drops a small black sphere alongside a vertical scale marked in centimetres. A number of flash photographs of the sphere are taken at 0.10 s intervals:



This diagram is shown sideways – the first black dot is at 0 cm and the next at 4 cm.

Figure 1.19

The first photograph is taken with the sphere at the top at time t = 0 s.

a Explain how Figure 1.19 shows that the sphere reaches a constant speed. **[2]**

b Determine the constant speed reached by the sphere. **[2]**

c Determine the distance that the sphere has fallen when t = 0.80 s. **[2]**

d In a real photograph, each image of the sphere appears slightly blurred because each flash is not instantaneous and takes a time of 0.0010 s.

Determine the absolute uncertainty that this gives in the position of each position of the black sphere when it is travelling at the final constant speed.

Suggest whether this should be observable on the diagram. **[2]**

[Total: 8]

10 a State **one** difference between a scalar quantity and a vector quantity and give an example of each. **[3]**

b A plane has an air speed of 500 km h^{-1} due north. A wind blows at 100 km h^{-1} from east to west.

Draw a vector diagram to calculate the resultant velocity of the plane. Give the direction of travel of the plane with respect to north. **[4]**

c The plane flies for 15 minutes. Calculate the displacement of the plane in this time. **[1]**

[Total: 8]

11 A small aircraft for one person is used on a short horizontal flight. On its journey from A to B, the resultant velocity of the aircraft is 15 m s^{-1} in a direction 60° east of north and the wind velocity is 7.5 m s^{-1} due north.

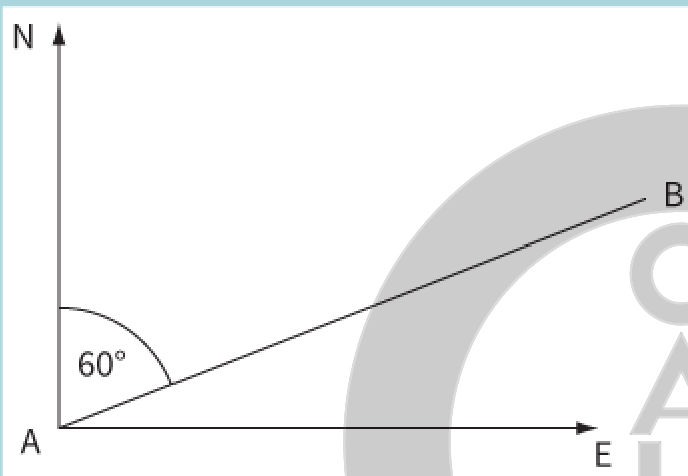


Figure 1.20

a Show that for the aircraft to travel from A to B it should be pointed due east. **[2]**

b After flying 5 km from A to B, the aircraft returns along the same path from B to A with a resultant velocity of 13.5 m s^{-1}. Assuming that the time spent at B is negligible, calculate the average speed for the complete journey from A to B and back to A. **[3]**

[Total: 5]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic...	Needs more work	Almost there	Ready to move on
define and use displacement, speed and velocity	1.1, 1.2, 1.3			
draw and interpret displacement–time graphs	1.4			
describe laboratory methods for determining speed	1.1			
understand the differences between scalar and vector quantities and give examples of each	1.2			
use vector addition to add and subtract vectors that are in the same plane.	1.6, 1.7			


80
100
120
140
160
180
200
km/h
220
240

〉Chapter 2

Accelerated motion

LEARNING INTENTIONS

In this chapter you will learn how to:

- define acceleration
- draw and interpret graphs of speed, velocity and acceleration
- calculate displacement from the area under a velocity–time graph
- calculate velocity and acceleration using gradients of a displacement–time graph and a velocity–time graph
- derive and use the equations of uniformly accelerated motion
- describe an experiment to measure the acceleration of free fall, *g*
- use perpendicular components to represent a vector
- explain projectile motion in terms of uniform velocity and uniform acceleration.

BEFORE YOU START

- Write down definitions of speed and velocity.
- Write a list of all the vectors that you know. Why are some quantities classed as vectors?

QUICK OFF THE MARK

The cheetah (Figure 2.1) has a maximum speed of more than 30 m s^{-1} (108 km/h). A cheetah can reach 20 m s^{-1} from a standing start in just three or four strides, taking only two seconds.

A car cannot increase its speed as rapidly but on a long straight road it can easily travel faster than a cheetah.

How do you think such measurements can be made? What apparatus is needed?

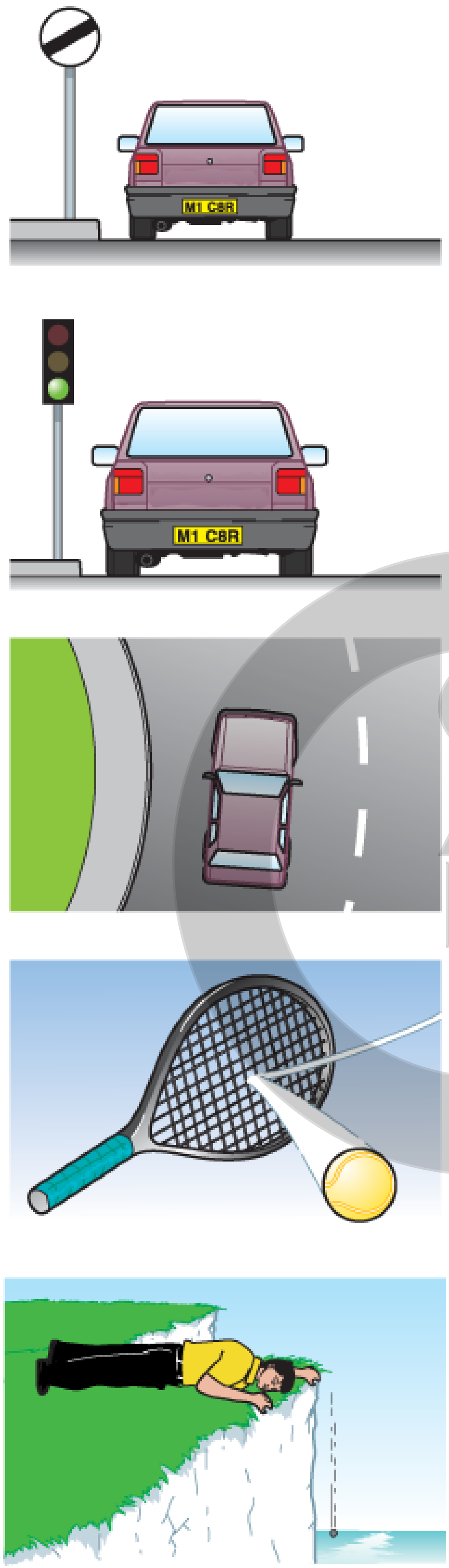
Figure 2.1: The cheetah is the world's fastest land animal. Its acceleration is impressive, too.

2.1 The meaning of acceleration

In everyday language, the term **accelerating** means 'speeding up'. Anything whose speed is increasing is accelerating. Anything whose speed is decreasing is decelerating.

To be more precise in our definition of acceleration, we should think of it as changing velocity. Any object whose speed is changing or which is changing its direction has **acceleration**. Because acceleration is linked to velocity in this way, it follows that it is a vector quantity.

Some examples of objects accelerating are shown in Figure 2.2.



A car speeding up as it leaves the town. The driver presses on the accelerator pedal to increase the car's velocity.

A car setting off from the traffic lights. There is an instant when the car is both stationary **and** accelerating. Otherwise it would not start moving.

A car travelling round a bend at a steady speed. The car's speed is constant, but its velocity is changing as it changes direction.

A ball being hit by a tennis racket. Both the ball's speed and direction are changing. The ball's velocity changes.

A stone dropped over a cliff. Gravity makes the stone go faster and faster. The stone accelerates as it falls.

Figure 2.2: Examples of objects accelerating.

COALI

2.2 Calculating acceleration

The acceleration of something indicates the rate at which its velocity is changing. Language can get awkward here. Looking at the sprinter in Figure 2.3, we might say, 'The sprinter accelerates faster than the car.' However, 'faster' really means 'greater speed'. It is better to say, 'The sprinter has a greater acceleration than the car.'

Acceleration is defined as follows:

$$\text{acceleration} = \text{rate of change of velocity}$$

$$\text{average acceleration} = \frac{\text{change in velocity}}{\text{time taken}}$$

So to calculate acceleration *a*, we need to know two quantities – the change in velocity Δv and the time taken Δt:

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

Sometimes this equation is written differently. We write *u* for the initial velocity and *v* for the final velocity (because *u* comes before *v* in the alphabet). The moving object accelerates from *u* to *v* in a time *t* (this is the same as the time represented by Δt in the equation). Then the acceleration is given by the equation:

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

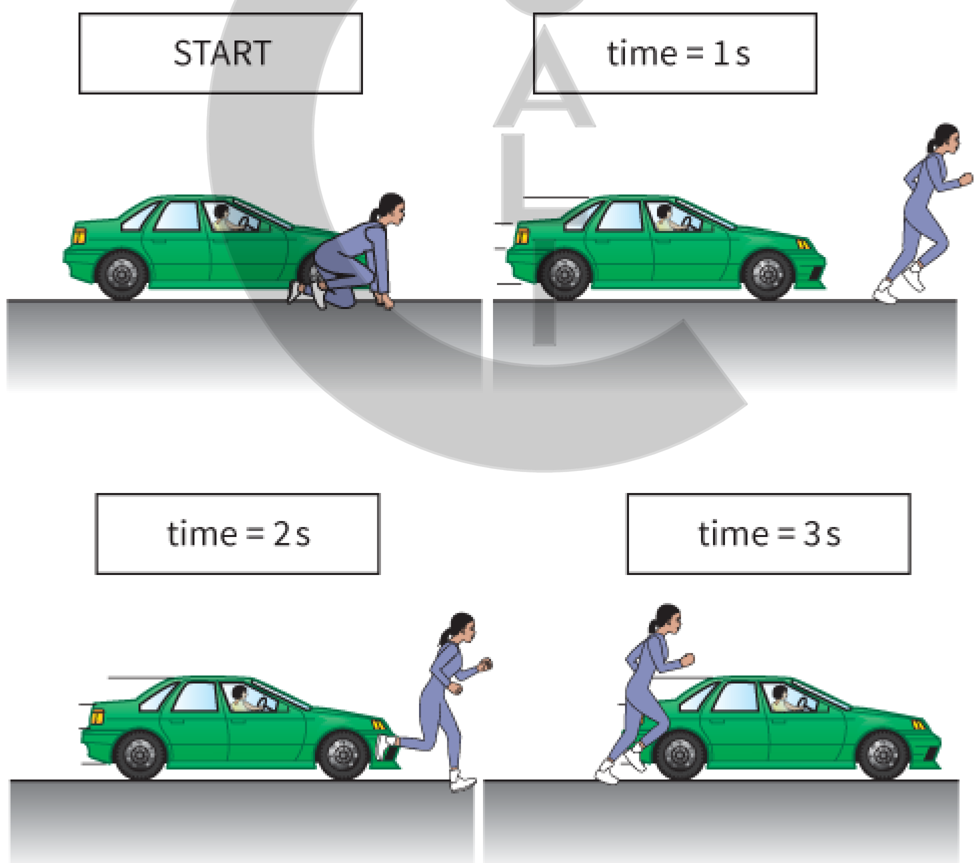


Figure 2.3: The sprinter has a greater acceleration than the car, but her top speed is less.

You must learn the definition of acceleration. It can be put in words or symbols. If you use symbols you must state what those symbols mean.

COALI

2.3 Units of acceleration

The unit of acceleration is m s^{-2} (metres per second squared). The sprinter might have an acceleration of 5 m s^{-2}; her velocity increases by 5 m s^{-1} every second. You could express acceleration in other units. For example, an advertisement might claim that a car accelerates from 0 to 60 miles per hour (mph) in 10 s. Its acceleration would then be 6 mph s^{-1} (6 miles per hour per second). However, mixing together hours and seconds is not a good idea, and so acceleration is almost always given in the standard SI unit of m s^{-2}.

WORKED EXAMPLES

1 Leaving a bus stop, a bus reaches a velocity of 8.0 m s^{-1} after 10 s. Calculate the acceleration of the bus.

Step 1 Note that the bus's initial velocity is 0 m s^{-1}.

Therefore:

$$\text{change in velocity } \Delta v = (8.0 - 0)\text{ m s}^{-1}$$
$$\text{time taken } \Delta t = 10\text{ s}$$

Step 2 Substitute these values in the equation for acceleration:

$$\text{acceleration} = \frac{\Delta v}{\Delta t} = \frac{8.0}{10} = 0.80\text{ m s}^{-2}$$

2 A sprinter starting from rest has an acceleration of 5.0 m s^{-2} during the first 2.0 s of a race. Calculate her velocity after 2.0 s.

Step 1 Rearranging the equation $a = \frac{v-u}{t}$ gives:

$v = u + at$

Step 2 Substituting the values and calculating gives:

$v = 0 + (5.0 \times 2.0) = 10$ m s^{-1}

3 A train slows down from 60 m s^{-1} to 20 m s^{-1} in 50 s. Calculate the magnitude of the deceleration of the train.

Step 1 Write what you know:

$u = 60$ m s^{-1} $v = 20$ m s^{-1} $t = 50$ s

Step 2 Take care! Here the train's final velocity is less than its initial velocity. To ensure that we arrive at the correct answer, we will use the alternative form of the equation to calculate *a*.

$$a = \frac{v-u}{t} = \frac{20-60}{50} = \frac{-40}{50} = -0.80\text{ m s}^{-2}$$

The minus sign (negative acceleration) indicates that the train is slowing down. It is decelerating. The magnitude of the deceleration is 0.80 m s^{-2}.

Questions

1 A car accelerates from a standing start and reaches a velocity of $18\ \mathrm{m\ s^{-1}}$ after 6.0 s. Calculate its acceleration.

2 A car driver brakes gently. Her car slows down from $23\ \mathrm{m\ s^{-1}}$ to $11\ \mathrm{m\ s^{-1}}$ in 20 s. Calculate the magnitude (size) of her deceleration. (Note that, because she is slowing down, her acceleration is negative.)

3 A stone is dropped from the top of a cliff. Its acceleration is $9.81\ \mathrm{m\ s^{-2}}$. How fast is it moving:

a after 1.0 s?

b after 3.0 s?

2.4 Deducing acceleration

The gradient of a velocity–time graph tells us whether the object's velocity has been changing at a high rate or a low rate, or not at all (Figure 2.4). We can deduce the value of the acceleration from the gradient of the graph:

acceleration = gradient of velocity–time graph

KEY IDEA

acceleration = gradient of velocity–time graph

The graph (Figure 2.5) shows how the velocity of a cyclist changed during the start of a sprint race. We can find his acceleration during the first section of the graph (where the line is straight) using the triangle as shown.

The change in velocity Δv is given by the vertical side of the triangle. The time taken Δt is given by the horizontal side.

$$\begin{aligned} \text{acceleration} &= \frac{\text{change in velocity}}{\text{time taken}} \\ &= \frac{25-0}{5} \\ &= 4.0\ \text{m s}^{-2} \end{aligned}$$

A more complex example where the velocity–time graph is curved is shown in Figure 2.18.

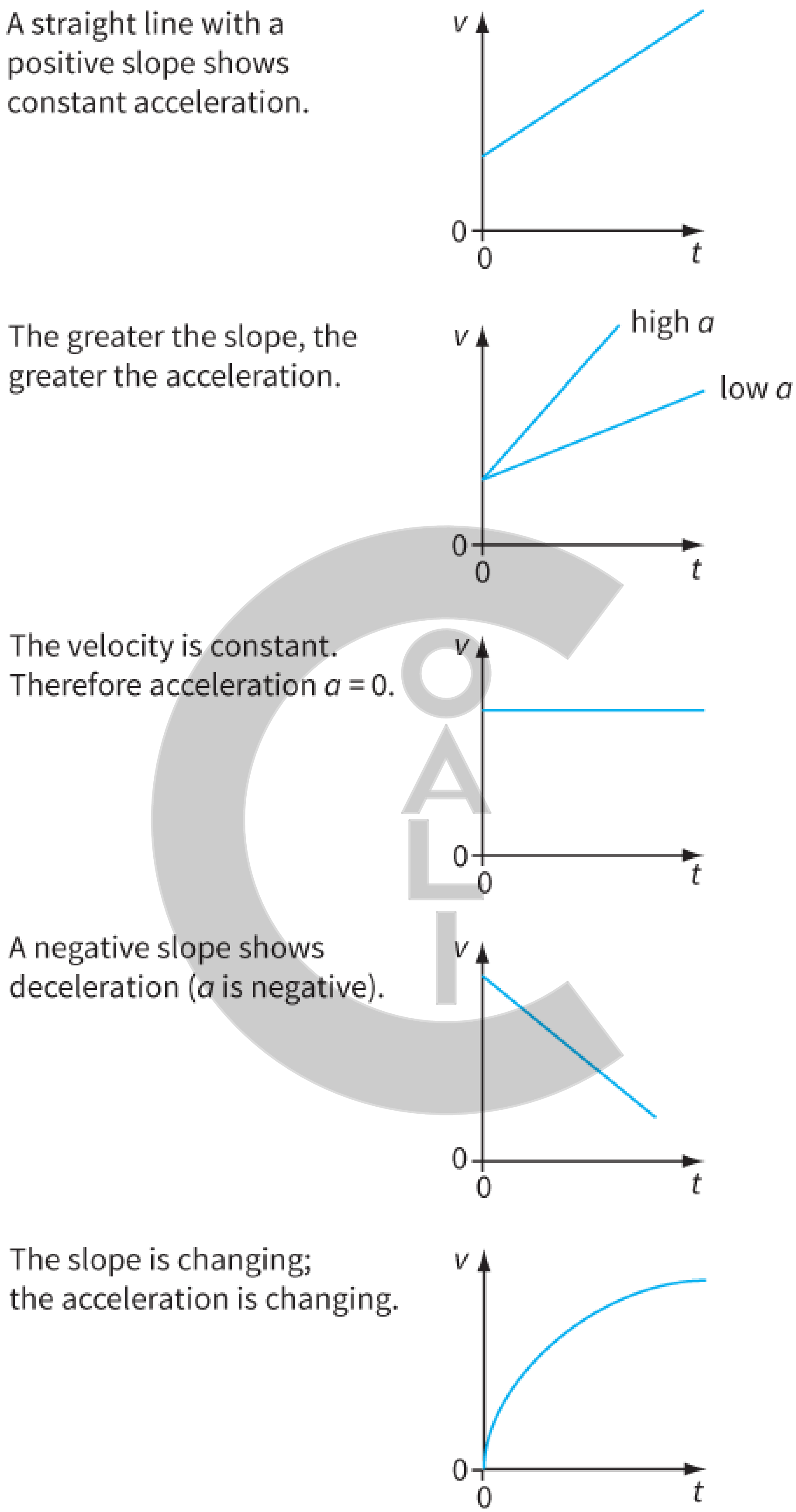


Figure 2.4: The gradient of a velocity–time graph is equal to acceleration.

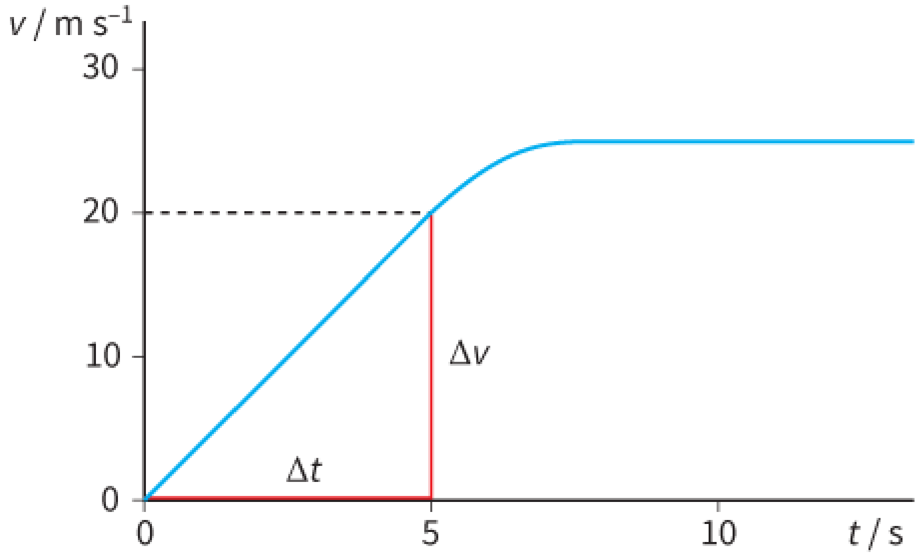


Figure 2.5: Deducing acceleration from a velocity–time graph.

2.5 Deducing displacement

We can also find the displacement of a moving object from its velocity–time graph. This is given by the area under the graph:

displacement = area under velocity–time graph

KEY IDEA

displacement = area **under** velocity–time graph

It is easy to see why this is the case for an object moving at a constant velocity. The displacement is simply velocity × time, which is the area of the shaded rectangle (Figure 2.6a).

For changing velocity, again the area under the graph gives displacement (Figure 2.6b).

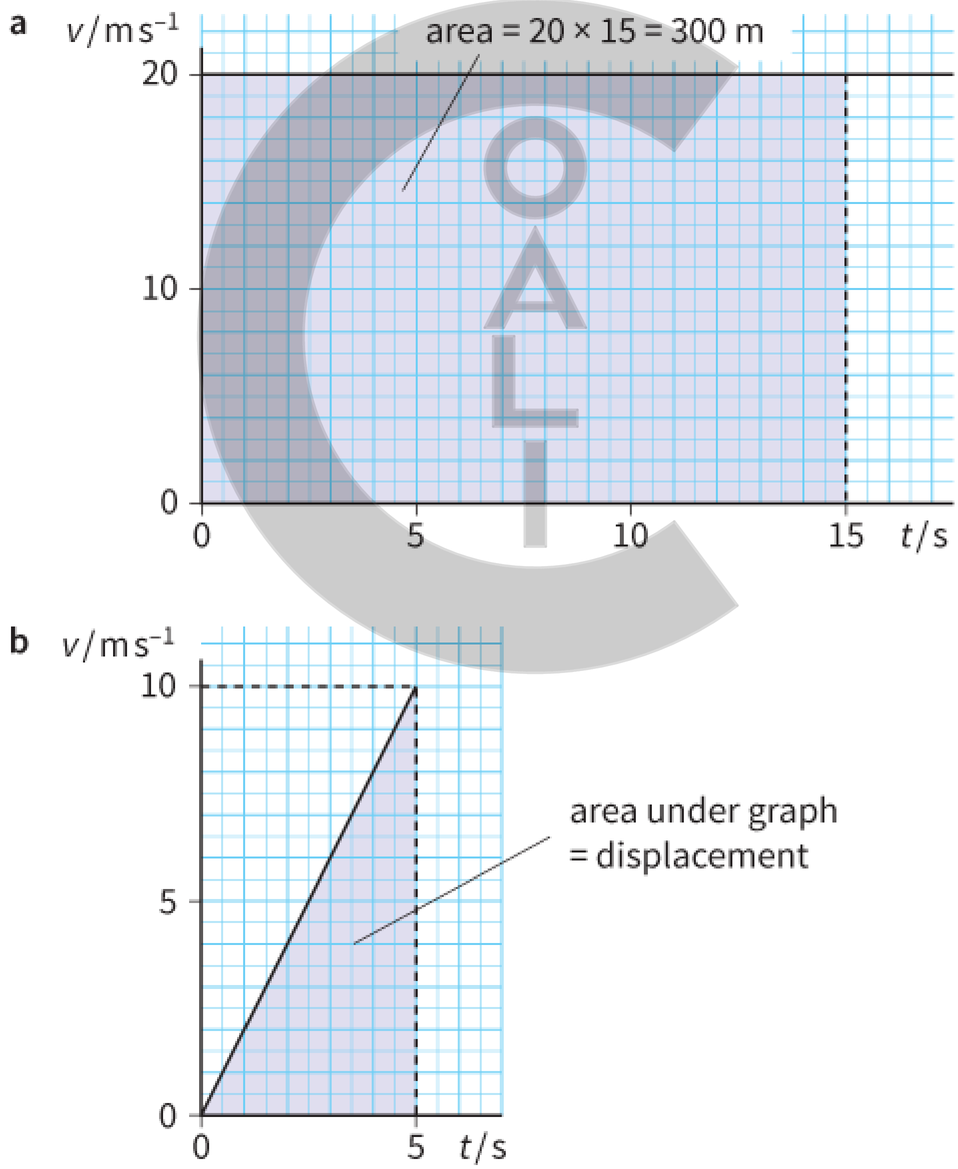


Figure 2.6: The area under the velocity–time graph is equal to the displacement of the object.

So, for this simple case in which the area is a triangle, we have:

$$\begin{aligned} \text{displacement} &= \tfrac{1}{2} \times \text{base} \times \text{height} \\ &= \tfrac{1}{2} \times 5.0 \times 10 \\ &= 25\ \text{m} \end{aligned}$$

It is easy to confuse displacement–time graphs and velocity–time graphs. Check by looking at the quantity marked on the vertical axis.

For more complex graphs, you may have to use other techniques such as counting squares to deduce the area, but this is still equal to the displacement.

(Take care when counting squares: it is easiest when the sides of the squares stand for one unit. Check the axes, as the sides may represent 2 units, 5 units or some other number.)

Questions

4 A lorry driver is travelling at the speed limit on a motorway. Ahead, he sees hazard lights and gradually slows down. He sees that an accident has occurred, and brakes suddenly to a halt. Sketch a velocity–time graph to represent the motion of this lorry.

5 Table 2.1 shows how the velocity of a motorcyclist changed during a speed trial along a straight road.

- **a** Draw a velocity–time graph for this motion.
- **b** From the table, deduce the motorcyclist's acceleration during the first 10 s.
- **c** Check your answer by finding the gradient of the graph during the first 10 s.
- **d** Determine the motorcyclist's acceleration during the last 15 s.
- **e** Use the graph to find the total distance travelled during the speed trial.

Velocity / m s^{-1}	0	15	30	30	20	10	0
Time / s	0	5	10	15	20	25	30

Table 2.1: Data for a motorcyclist.

2.6 Measuring velocity and acceleration

In a car crash, the occupants of the car may undergo a very rapid deceleration. This can cause them serious injury, but can be avoided if an air-bag is inflated within a fraction of a second. Figure 2.7 shows the tiny accelerometer at the heart of the system, which detects large accelerations and decelerations.

The acceleration sensor consists of two rows of interlocking teeth. In the event of a crash, these move relative to one another, and this generates a voltage that triggers the release of the air-bag.

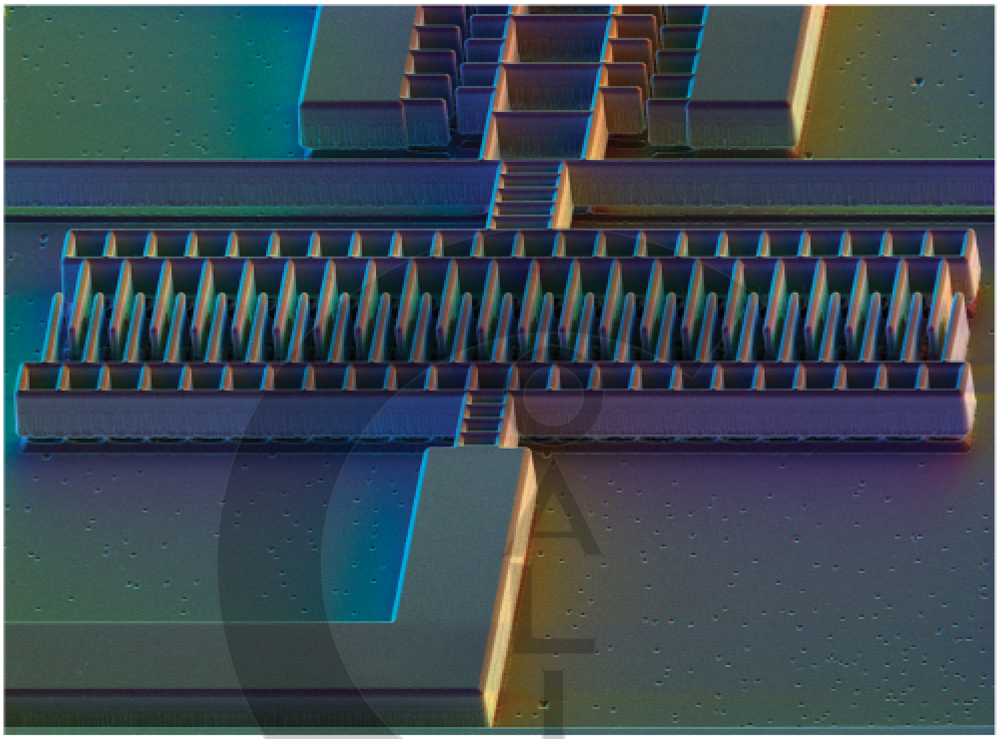


Figure 2.7: A micro-mechanical acceleration sensor is used to detect sudden accelerations and decelerations as a vehicle travels along the road. This electron microscope image shows the device magnified about 1000 times.

At the top of the photograph (Figure 2.7), you can see a second sensor that detects sideways accelerations. This is important in the case of a side impact.

These sensors can also be used to detect when a car swerves or skids, perhaps on an icy road. In this case, they activate the car's stability-control systems.

2.7 Determining velocity and acceleration in the laboratory

In Chapter 1, we looked at ways of finding the velocity of a trolley moving in a straight line. These involved measuring distance and time, and deducing velocity. Practical Activity 2.1 shows how these techniques can be extended to find the acceleration of a trolley.

PRACTICAL ACTIVITY 2.1: LABORATORY MEASUREMENTS OF ACCELERATION

Measurements using light gates

The computer records the time for the first 'interrupt' section of the card to pass through the light beam of the light gate (Figure 2.8). Given the length of the interrupt, it can work out the trolley's initial velocity *u*. This is repeated for the second interrupt to give final velocity *v*. The computer also records the time interval $t_3 - t_1$ between these two velocity measurements. Now it can calculate the acceleration *a* as shown:

$$u = \frac{l_1}{t_2 - t_1}$$

(l_1 = length of first section of the interrupt card)

and

$$v = \frac{l_2}{t_4 - t_3}$$

(l_2 = length of second section of the interrupt card)

Therefore:

$$a = \frac{\text{change in velocity}}{\text{time taken}} = \frac{v - u}{t_3 - t_1}$$

(Note that this calculation gives only an approximate value for *a*. This is because *u* and *v* are **average** speeds over a period of time; for an accurate answer we would need to know the speeds at times t_1 and t_3.)

Sometimes two light gates are used with a card of length *l*. The computer can still record the times as shown and calculate the acceleration in the same way, with $l_1 = l_2 = l$.

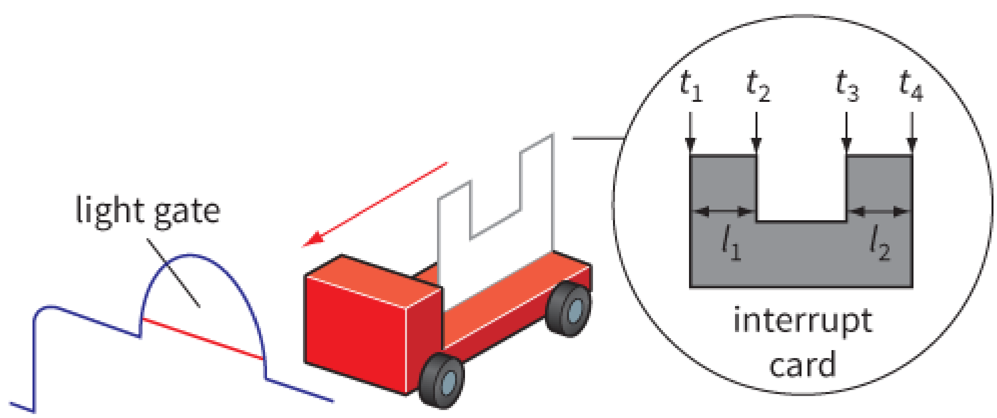


Figure 2.8: Determining acceleration using a single light gate.

Measurements using a ticker-timer

The practical arrangement is the same as for measuring velocity. Now we have to think about how to interpret the tape produced by an accelerating trolley (Figure 2.9).

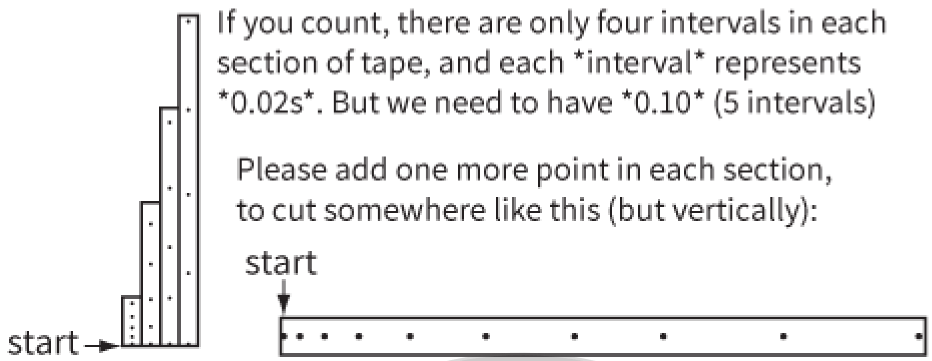


Figure 2.9: Ticker-tape for an accelerating trolley.

The tape is divided into sections, as before, every five dots. Remember that the time interval between adjacent dots is 0.02 s. Each section represents 0.10 s.

By placing the sections of tape side by side, you can picture the velocity–time graph.

The length of each section gives the trolley's displacement in 0.10 s, from which the average velocity during this time can be found. This can be repeated for each section of the tape, and a velocity–time graph drawn. The gradient of this graph is equal to the acceleration. Table 2.2 and Figure 2.10 show some typical results.

The acceleration is calculated to be:

$$a = \frac{\Delta v}{\Delta t} = \frac{0.93}{0.20} \approx 4.7\ \text{ms}^{-2}$$

Section of tape	Time at start / s	Time interval / s	Length of section / cm	Velocity / m s^{-1}
1	0.0	0.10	2.3	0.23
2	0.10	0.10	7.0	0.70
3	0.20	0.10	11.6	1.16

Table 2.2: Data for Figure 2.10.

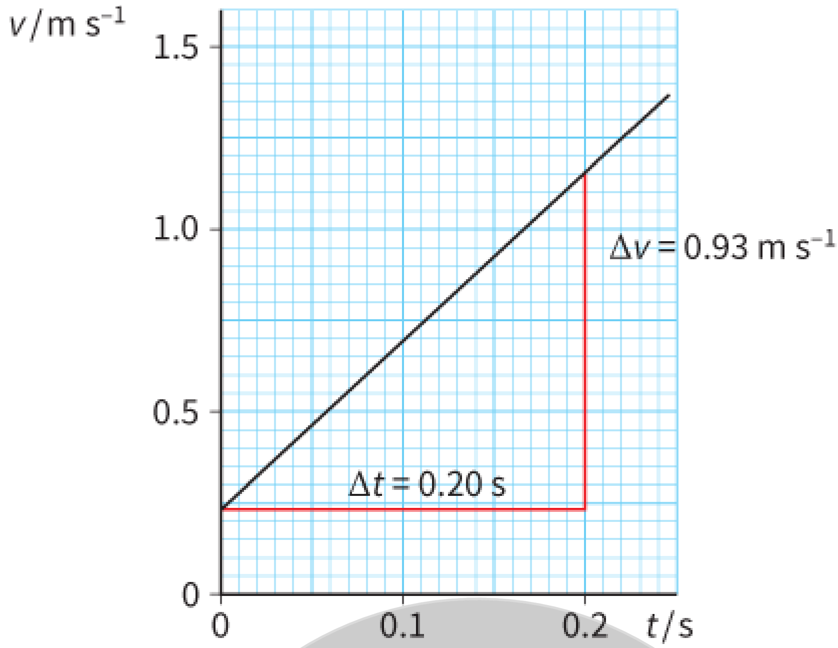


Figure 2.10: Deducing acceleration from measurements of a ticker-tape.

Measurements using a motion sensor

The computer software that handles the data provided by the motion sensor can calculate the acceleration of a trolley. However, because it deduces velocity from measurements of position, and then calculates acceleration from values of velocity, its precision is relatively poor.

Questions

6 Sketch a section of ticker-tape for a trolley that travels at a steady velocity and then decelerates.

7 Figure 2.11 shows the dimensions of an interrupt card, together with the times recorded as it passed through a light gate. Use these measurements to calculate the acceleration of the card. (Follow the steps outlined in Practical Activity 2.1.)

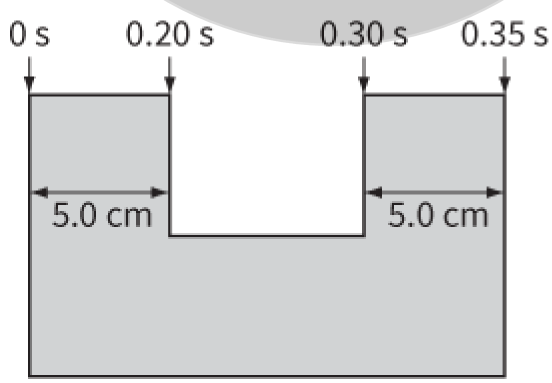


Figure 2.11: For Question 7.

8 Two adjacent five-dot sections of a ticker-tape measure 10 cm and 16 cm, respectively. The interval between dots is 0.02 s. Deduce the acceleration of the trolley that produced the tape.

COALI

2.8 The equations of motion

As a space rocket rises from the ground, its velocity steadily increases. It is accelerating (Figure 2.12).

Eventually, it will reach a speed of several kilometres per second. Any astronauts aboard find themselves pushed back into their seats while the rocket is accelerating.

Figure 2.12: A rocket accelerates as it lifts off from the ground.

The engineers who planned the mission must be able to calculate how fast the rocket will be travelling and where it will be at any point in its journey. They have sophisticated computers to do this, using more elaborate versions of the four equations of motion.

There is a set of equations that allows us to calculate the quantities involved when an object is moving with a **constant acceleration**.

The quantities we are concerned with are:

s displacement

u initial velocity

v final velocity

a acceleration

t time taken

The four **equations of motion** are shown above.

Take care using the equations of motion. They can only be used for:

- motion in a straight line

- an object with constant acceleration.

KEY EQUATIONS

The four equations of motion:

equation 1: $v = u + at$

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

equation 3: $s = ut + \frac{1}{2}at^2$

equation 4: $v^2 = u^2 + 2as$

To get a feel for how to use these equations, we will consider some worked examples. In each example, we will follow the same procedure:

Step 1 We write down the quantities that we know, and the quantity we want to find.

Step 2 Then we choose the equation that links these quantities, and substitute in the values.

Step 3 Finally, we calculate the unknown quantity.

We will look at where these equations come from in the next topic, ‘Deriving the equations of motion’.

WORKED EXAMPLES

4 The rocket shown in Figure 2.12 lifts off from rest with an acceleration of 20 m s^{-2}. Calculate its velocity after 50 s.

Step 1 What we know:

$u = 0$ m s^{-1}

$a = 20$ m s^{-2}

$t = 50$ s

and what we want to know: v = ?

Step 2 The equation linking u, a, t and v is equation 1:

$v = u + at$

Substituting gives:

$v = 0 + (20 \times 50)$

Step 3 Calculation then gives:

$v = 1000$ m s^{-1}

So the rocket will be travelling at 1000 m s^{-1} after 50 s. This makes sense, since its velocity increases by 20 m s^{-1} every second, for 50 s.

You could use the same equation to work out how long the rocket would take to reach a velocity of 2000 m s^{-1}, or the acceleration it must have to reach a speed of 1000 m s^{-1} in 40 s and so on.

5 The car shown in Figure 2.13 is travelling along a straight road at 8.0 m s^{-1}. It accelerates at 1.0 m s^{-2} for a distance of 18 m. How fast is it then travelling?

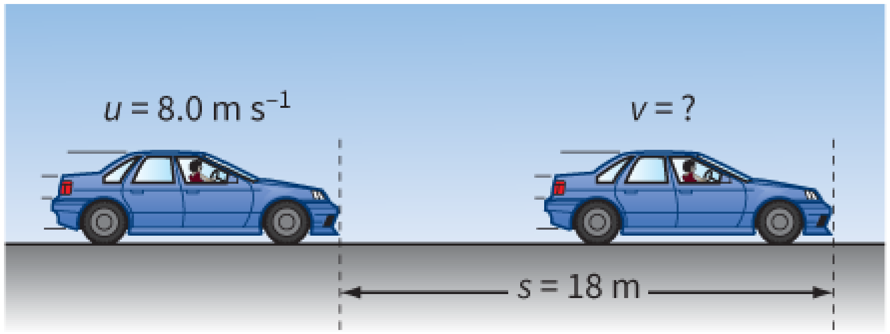


Figure 2.13: For Worked example 5. This car accelerates for a short distance as it travels along the road.

In this case, we will have to use a different equation, because we know the distance during which the car accelerates, not the time.

Step 1 What we know:

$u = 8.0\ \text{m s}^{-1}$

$a = 1.0\ \text{m s}^{-2}$

$s = 18\ \text{m}$

and what we want to know: v = ?

Step 2 The equation we need is equation 4:

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

Substituting gives:

$v^2 = 8.0^2 + (2 \times 1.0 \times 18)$

Step 3 Calculation then gives:

$v^2 = 64 + 36 = 100\ \text{m}^2\ \text{s}^{-2}$

$v = 10\ \text{m s}^{-1}$

So the car will be travelling at $10\ \text{m s}^{-1}$ when it stops accelerating.

(You may find it easier to carry out these calculations without including the units of quantities when you substitute in the equation. However, including the units can help to ensure that you end up with the correct units for the final answer.)

6 A train (Figure 2.14) travelling at $20\ \text{m s}^{-1}$ accelerates at $0.50\ \text{m s}^{-2}$ for 30 s. Calculate the distance travelled by the train in this time.

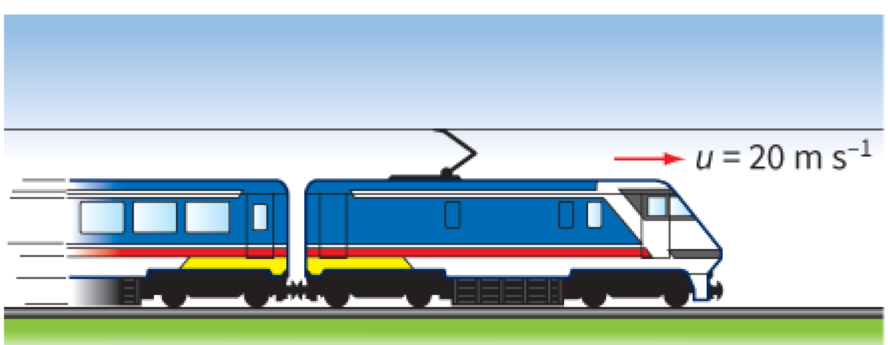


Figure 2.14: For Worked example 6. This train accelerates for 30 s.

Step 1 What we know:

$$u = 20\ \text{m s}^{-1}$$
$$t = 30\ \text{s}$$
$$a = 0.50\ \text{m s}^{-2}$$

and what we want to know: s = ?

Step 2 The equation we need is equation 3:

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

Substituting gives:

$$s = (20 \times 30) + \frac{1}{2} \times 0.5 \times (30)^2$$

Step 3 Calculation then gives:

$s = 600 + 225 = 825$ m

So the train will travel 825 m while it is accelerating.

7 The cyclist in Figure 2.15 is travelling at 15 m s^{-1}. She brakes so that she doesn't collide with the wall. Calculate the magnitude of her deceleration.

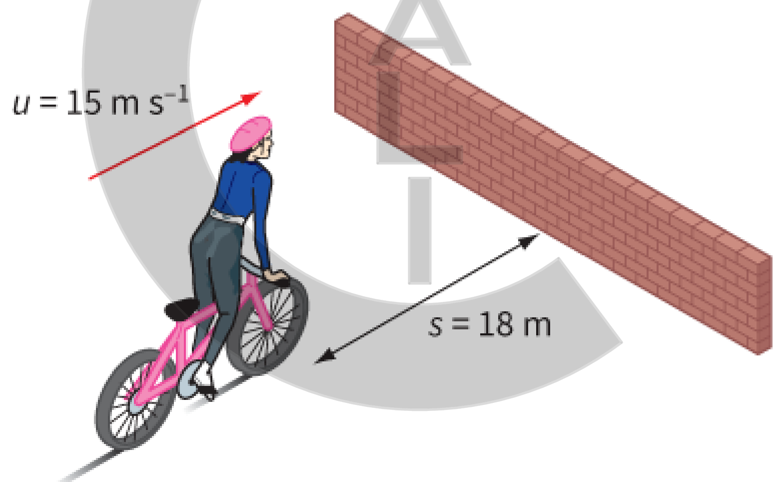


Figure 2.15: For Worked example 7. The cyclist brakes to stop herself colliding with the wall.

This example shows that it is sometimes necessary to rearrange an equation, to make the unknown quantity its subject. It is easiest to do this before substituting in the values.

Step 1 What we know:

$$u = 15\ \text{m s}^{-1}$$
$$v = 0\ \text{m s}^{-1}$$
$$s = 18\ \text{m}$$

and what we want to know: a = ?

Step 2 The equation we need is equation 4:

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

Rearranging gives:

$$a = \frac{v^2 - u^2}{2s}$$

$$a = \frac{0^2 - 15^2}{2 \times 18}$$

$$= \frac{-225}{36}$$

Step 3 Calculation then gives:

$a = -6.25 \text{ m s}^{-2} \approx -6.3 \text{ m s}^{-2}$

So the cyclist will have to brake hard to achieve a deceleration of magnitude 6.3 m s^{-2}. The minus sign shows that her acceleration is negative; in other words, a deceleration.

Questions

9 A car is initially stationary. It has a constant acceleration of 2.0 m s^{-2}.

a Calculate the velocity of the car after 10 s.

b Calculate the distance travelled by the car at the end of 10 s.

c Calculate the time taken by the car to reach a velocity of 24 m s^{-1}.

10 A train accelerates steadily from 4.0 m s^{-1} to 20 m s^{-1} in 100 s.

a Calculate the acceleration of the train.

b From its initial and final velocities, calculate the average velocity of the train.

c Calculate the distance travelled by the train in this time of 100 s.

11 A car is moving at 8.0 m s^{-1}. The driver makes it accelerate at 1.0 m s^{-2} for a distance of 18 m. What is the final velocity of the car?

2.9 Deriving the equations of motion

We have seen how to make use of the equations of motion. But where do these equations come from? They arise from the definitions of velocity and acceleration.

We can find the first two equations from the velocity–time graph shown in Figure 2.16. The graph represents the motion of an object. Its initial velocity is u. After time t, its final velocity is v.

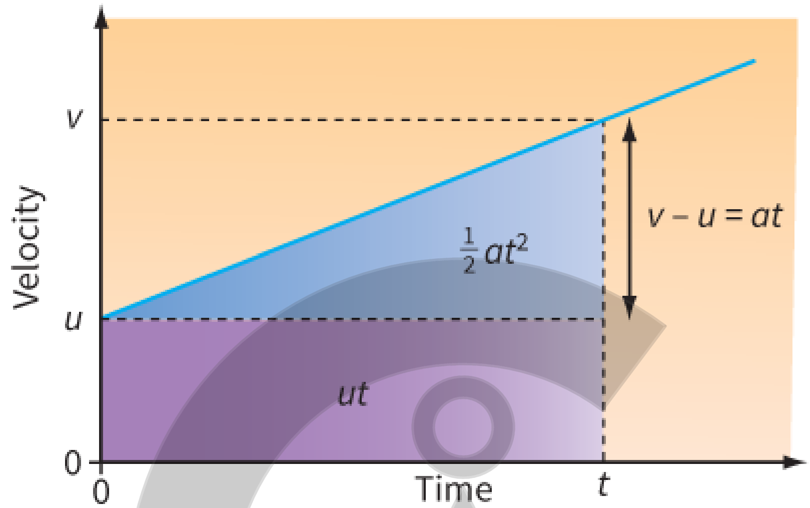


Figure 2.16: This graph shows the variation of velocity of an object with time. The object has constant acceleration.

Equation 1

The graph of Figure 2.16 is a straight line, therefore the object's acceleration a is constant. The gradient (slope) of the line is equal to acceleration.

The acceleration is defined as:

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

which is the gradient of the line. Rearranging this gives the first equation of motion:

$$v = u + at \qquad \text{(equation 1)}$$

Equation 2

Displacement is given by the area under the velocity–time graph. Figure 2.17 shows that the object's average velocity is half-way between u and v. So the object's average velocity, calculated by averaging its initial and final velocities, is given by:

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

The object's displacement is the shaded area in Figure 2.17. This is a rectangle, and so we have:

displacement = average velocity × time taken and hence:

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

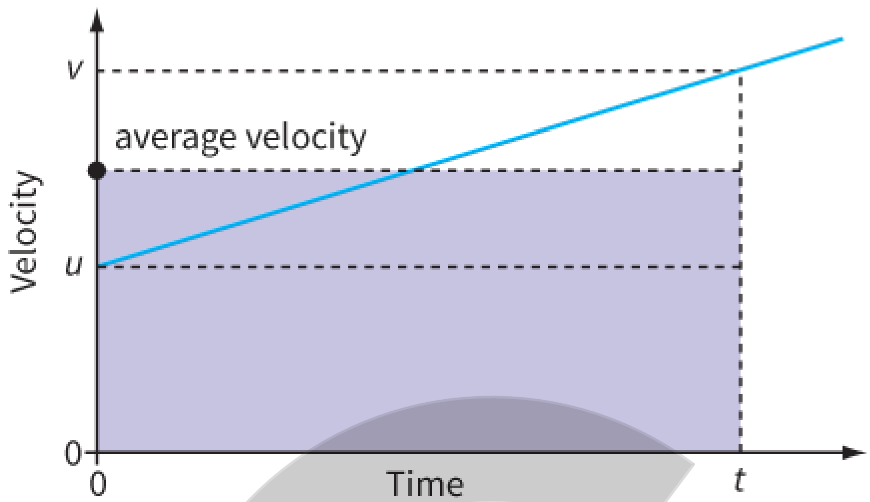


Figure 2.17: The average velocity is half-way between *u* and *v*.

Equation 3

From equations 1 and 2, we can derive equation 3:

$$v = u + at \qquad \text{(equation 1)}$$

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

Substituting *v* from equation 1 gives:

$$s = \frac{(u+u+at)}{2} \times t$$
$$= \frac{2ut}{2} + \frac{at^2}{2}$$

So

$$s = ut + \frac{1}{2}at^2 \qquad \text{(equation 3)}$$

Looking at Figure 2.16, you can see that the two terms on the right of the equation correspond to the areas of the rectangle and the triangle that make up the area under the graph. Of course, this is the same area as the rectangle in Figure 2.17.

Equation 4

Equation 4 is also derived from equations 1 and 2:

$$v = u + at \qquad \text{(equation 1)}$$

(equation 2)

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

Substituting for time *t* from equation 1 gives:

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

Rearranging this gives:

$$\begin{aligned} 2as &= (u+v)(v-u) \\ &= v^2 - u^2 \end{aligned}$$

or simply:

$$v^2 = u^2 + 2as \qquad \text{(equation 4)}$$

Investigating road traffic accidents

The police frequently have to investigate road traffic accidents. They make use of many aspects of physics, including the equations of motion. The next two questions will help you to apply what you have learned to situations where police investigators have used evidence from skid marks on the road.

Questions

12 Trials on the surface of a new road show that, when a car skids to a halt, its acceleration is $-7.0\ m\ s^{-2}$. Estimate the skid-to-stop distance of a car travelling at a speed limit of $30\ m\ s^{-1}$ (approximately $110\ km\ h^{-1}$ or 70 mph).

13 At the scene of an accident on a country road, police find skid marks stretching for 50 m. Tests on the road surface show that a skidding car decelerates at $6.5\ m\ s^{-2}$. Was the car that skidded exceeding the speed limit of $25\ m\ s^{-1}$ ($90\ km\ h^{-1}$) on this road?

2.10 Uniform and nonuniform acceleration

It is important to note that the equations of motion only apply to an object that is moving with a constant acceleration. If the acceleration *a* was changing, you wouldn't know what value to put in the equations. Constant acceleration is often referred to as **uniform acceleration**.

The velocity–time graph in Figure 2.18 shows **non-uniform acceleration**. It is not a straight line; its gradient is changing (in this case, decreasing).

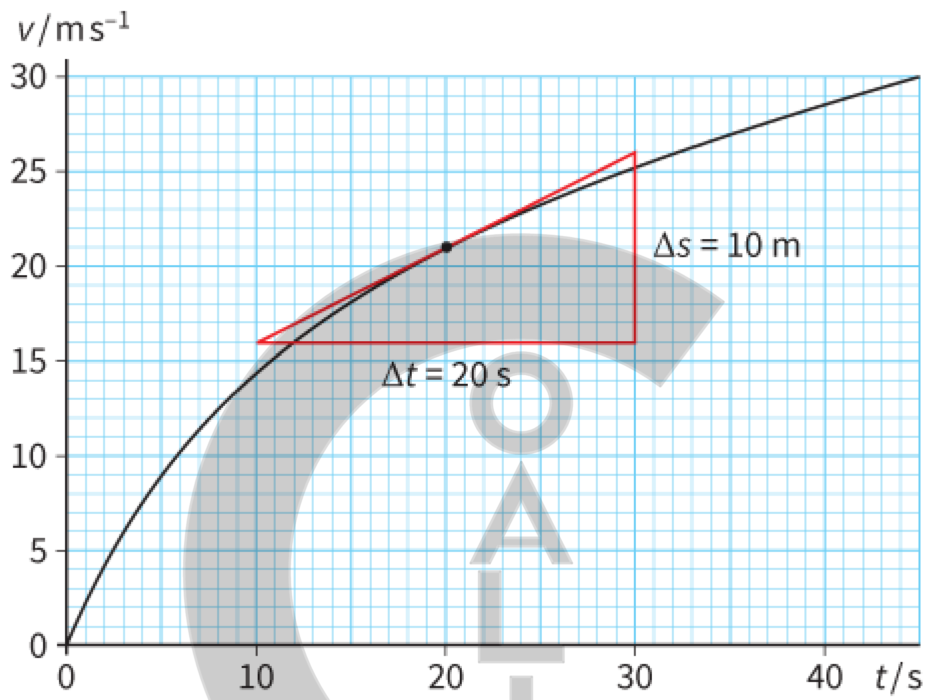


Figure 2.18: This curved velocity–time graph cannot be analysed using the equations of motion.

The acceleration at any instant in time is given by the gradient of the velocity–time graph. The triangle in Figure 2.18 shows how to find the acceleration at t = 20 seconds:

- At the time of interest, mark a point on the graph.
- Draw a **tangent** to the curve at that point.
- Make a large right-angled triangle, and use it to find the gradient.

You can find the change in displacement of the body as it accelerates by determining the area under the velocity–time graph.

To find the displacement of the object in Figure 2.18 between t = 0 and t = 20 s, the most straightforward, but lengthy, method is just to count the number of small squares.

In this case, up to t = 20 s, there are about 250 small squares. This is tedious to count but you can save yourself a lot of time by drawing a line from the origin to the point at 20 s. The area of the triangle is easy to find (200 small squares) and then you only have to count the number of small squares between the line you have drawn and the curve on the graph (about 50 squares)

In this case, each square is 1 m s^{-1} on the *y*-axis by 1 s on the *x*-axis, so the area of each square is 1 × 1 = 1 m and the displacement is 250 m. In other cases, note carefully the value of each side of the square you have chosen.

Questions

14 The graph in Figure 2.19 represents the motion of an object moving with varying acceleration. Lay your ruler on the diagram so that it is tangential to the graph at point P.

- **a** What are the values of time and velocity at this point?
- **b** Estimate the object's acceleration at this point.

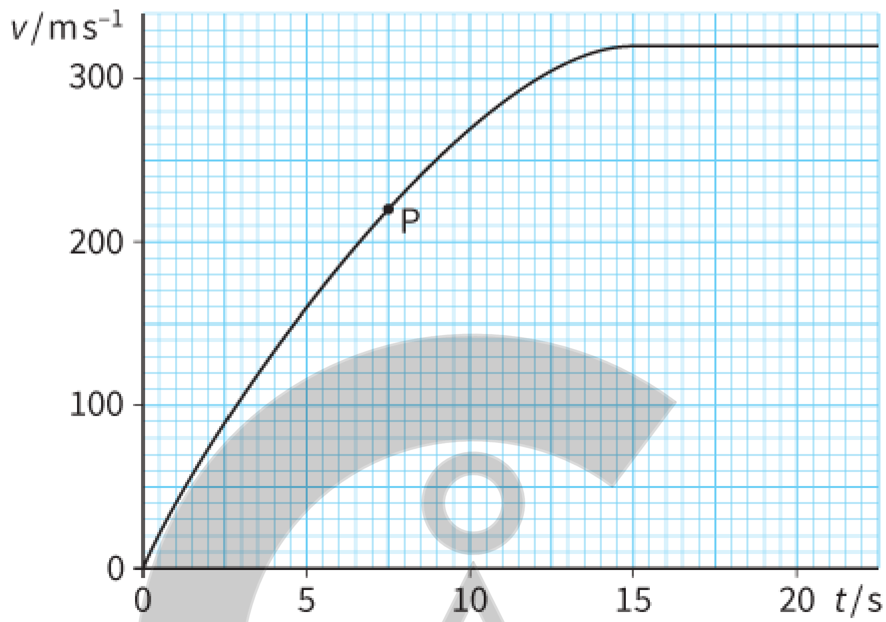


Figure 2.19: For Question 14.

15 The velocity–time graph (Figure 2.20) represents the motion of a car along a straight road for a period of 30 s.

- **a** Describe the motion of the car.
- **b** From the graph, determine the car's initial and final velocities over the time of 30 s.
- **c** Determine the acceleration of the car.
- **d** By calculating the area under the graph, determine the displacement of the car.
- **e** Check your answer to part **d** by calculating the car's displacement using $s = ut\frac{1}{2}at^2$

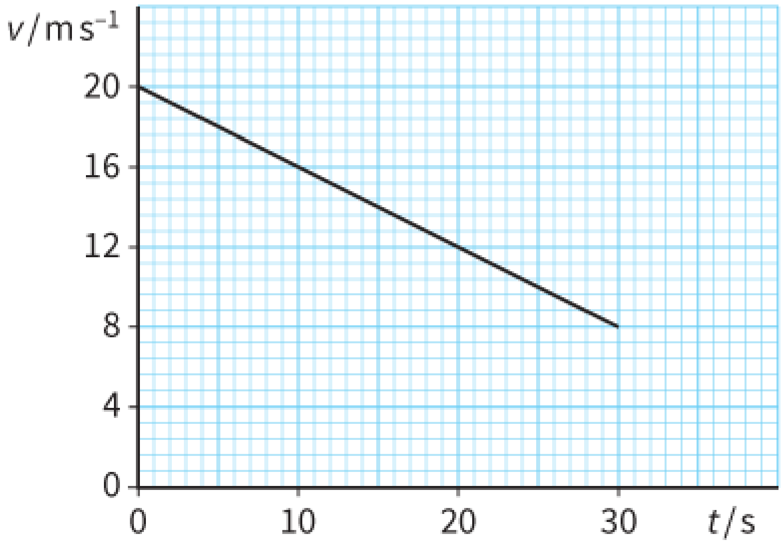


Figure 2.20: For Question 15.

2.11 Acceleration caused by gravity

If you drop a ball or stone, it falls to the ground. Figure 2.21, based on a multiflash photograph, shows the ball at equal intervals of time. You can see that the ball's velocity increases as it falls because the spaces between the images of the ball increase steadily. The ball is accelerating.

A multiflash photograph is useful to demonstrate that the ball accelerates as it falls. Usually, objects fall too quickly for our eyes to be able to observe them speeding up. It is easy to imagine that the ball moves quickly as soon as you let it go, and falls at a steady speed to the ground. Figure 2.21 shows that this is not the case.

If we measure the acceleration of a freely falling object on the surface of the Earth, we find a value of about 9.81 m s^{-2}. This is known as the acceleration of **free fall**, and is given the symbol g:

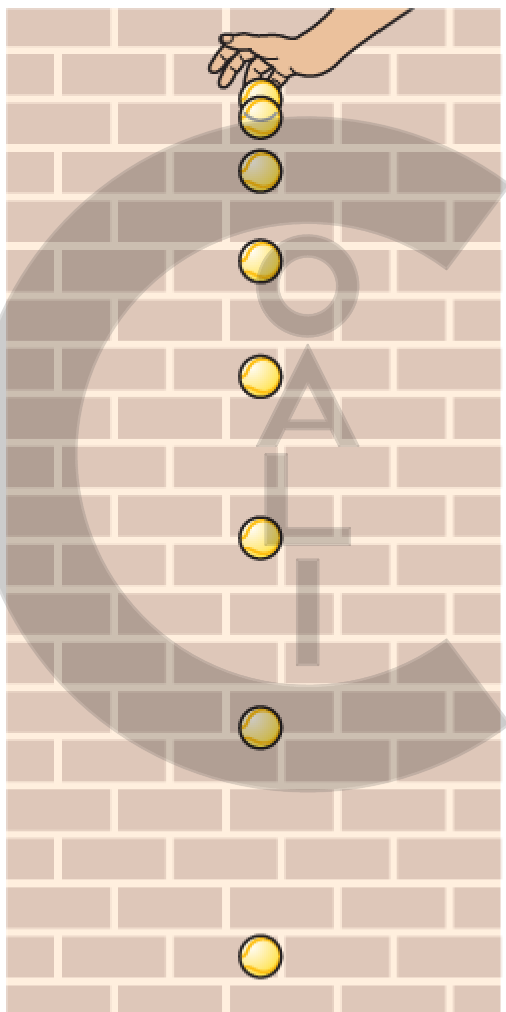


Figure 2.21: This diagram of a falling ball, based on a multiflash photo, clearly shows that the ball's velocity increases as it falls.

$$\text{acceleration of free fall } g = 9.81 \text{ m s}^{-2}$$

The value of g depends on where you are on the Earth's surface, but we usually take g = 9.81 m s^{-2}.

If we drop an object, its initial velocity u = 0. How far will it fall in time t? Substituting in $s = ut\frac{1}{2}at^2$ gives displacement s:

$$s = \tfrac{1}{2} \times 9.81 \times t^2$$
$$= 4.9 \times t^2$$

Hence, by timing a falling object, we can determine *g*.

Questions

16 If you drop a stone from the edge of a cliff, its initial velocity $u = 0$, and it falls with acceleration $g = 9.81$ m s^{-2}. You can calculate the distance *s* it falls in a given time *t* using an equation of motion.

a Copy and complete Table 2.3, which shows how *s* depends on *t*.

b Draw a graph of *s* against *t*.

c Use your graph to find the distance fallen by the stone in 2.5 s.

d Use your graph to find how long it will take the stone to fall to the bottom of a cliff 40 m high. Check your answer using the equations of motion.

Time / s	0	1.0	2.0	3.0	4.0
Displacement / m	0	4.9			

Table 2.3: Time t and displacement s data fo

17 An egg falls off a table. The floor is 0.8 m from the table-top.

a Calculate the time taken to reach the ground.

b Calculate the velocity of impact with the ground.

2.12 Determining *g*

One way to measure the acceleration of free fall *g* would be to try bungee-jumping (Figure 2.22). You would need to carry a stopwatch, and measure the time between jumping from the platform and the moment when the elastic rope begins to slow your fall. If you knew the length of the unstretched rope, you could calculate *g*.

There are easier methods for finding *g* that can be used in the laboratory. These are described in Practical Activity 2.2.

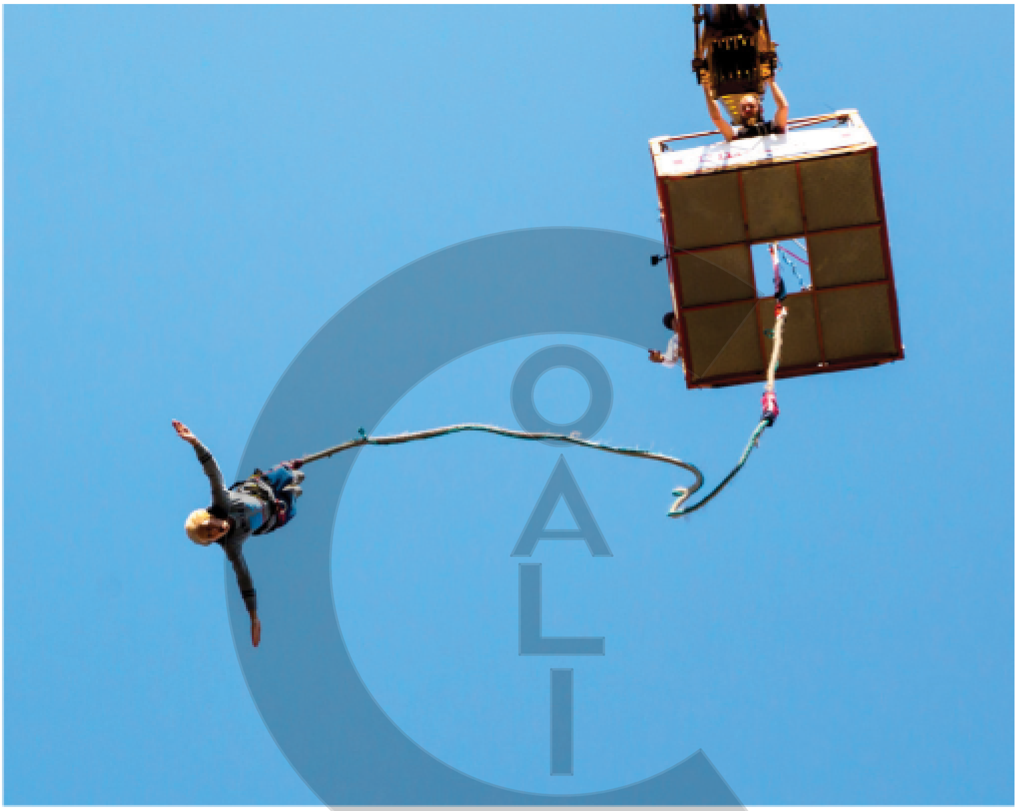


Figure 2.22: A bungee-jumper falls with initial acceleration *g*.

PRACTICAL ACTIVITY 2.2: LABORATORY MEASUREMENTS OF *g*

Measuring *g* using an electronic timer

In this method, a steel ball-bearing is held by an electromagnet (Figure 2.23). When the current to the magnet is switched off, the ball begins to fall and an electronic timer starts. The ball falls through a trapdoor, and this breaks a circuit to stop the timer. This tells us the time taken for the ball to fall from rest through the distance *h* between the bottom of the ball and the trapdoor.

Here is how we can use one of the equations of motion to find *g*:

displacement $s = h$

time taken = t

initial velocity $u = 0$

acceleration $a = g$

Substituting in $s = ut + \frac{1}{2}at^2$ gives:

$h = \frac{1}{2}gt^2$

and for any values of *h* and *t* we can calculate a value for *g*.

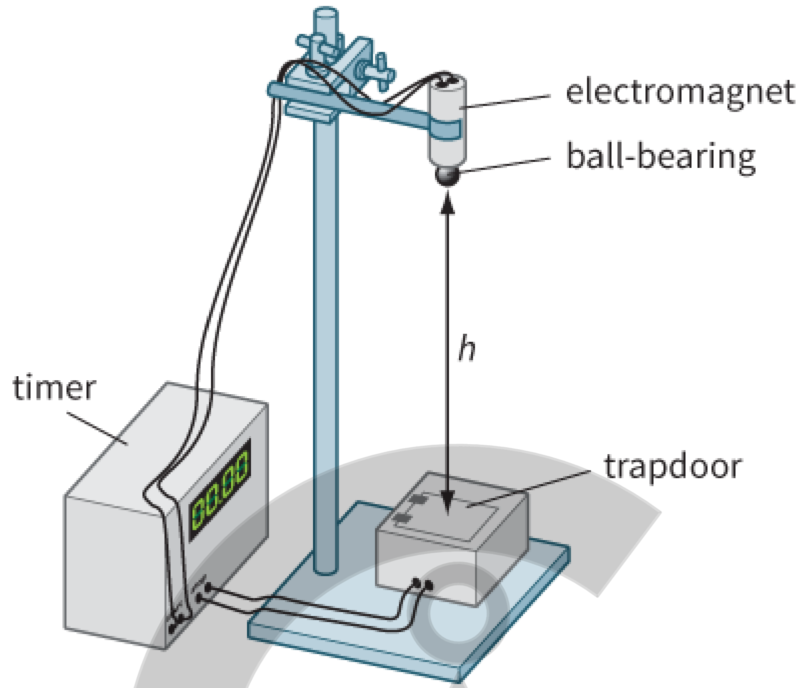


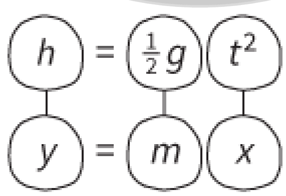
Figure 2.23: The timer records the time for the ball to fall through the distance *h*.

A more satisfactory procedure is to take measurements of *t* for several different values of *h*. The height of the ball bearing above the trapdoor is varied systematically, and the time of fall measured several times to calculate an average for each height. Table 2.4 and Figure 2.24 show some typical results. We can deduce *g* from the gradient of the graph of *h* against t^2.

The equation for a straight line through the origin is:

$$y = mx$$

In our experiment we have:



h / m	*t* / s	t^2 / s^2
0.27	0.25	0.063
0.39	0.30	0.090
0.56	0.36	0.130
0.70	0.41	0.168
0.90	0.46	0.212

Table 2.4: Data for Figure 2.24. These are mean values.

The gradient of the straight line of a graph of h against t^2 is equal to $\frac{g}{2}$

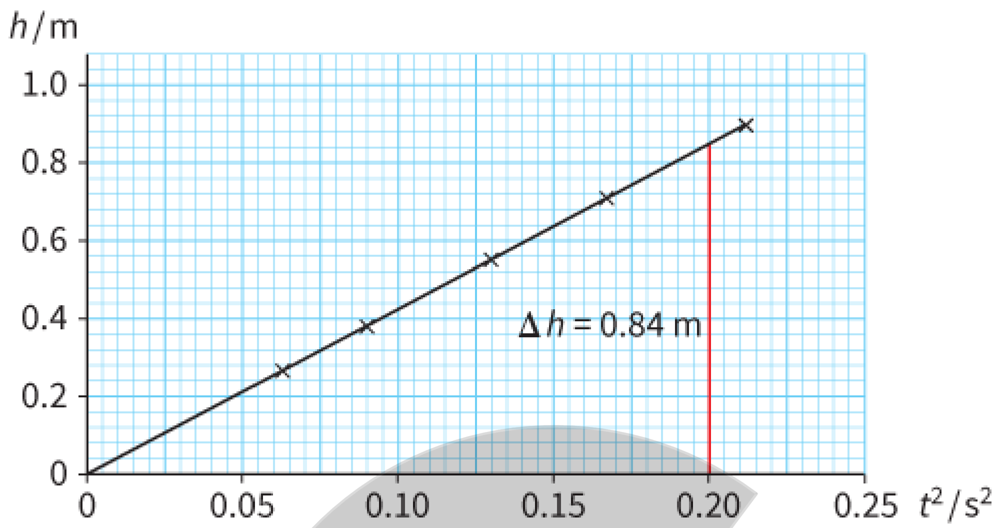


Figure 2.24: The acceleration of free fall can be determined from the gradient.

Therefore:

$$\text{gradient} = \frac{g}{2} = \frac{0.84}{0.20} = 4.2$$

$$g = 4.2 \times 2 = 8.4 \text{ m s}^{-2}$$

Sources of uncertainty

The electromagnet may retain some magnetism when it is switched off, and this may tend to slow the ball's fall. Consequently, the time t recorded by the timer may be longer than if the ball were to fall completely freely. From $h = \frac{1}{2}gt^2$ it follows that, if t is too great, the experimental value of g will be too small. This is an example of a systematic error – all the results are systematically distorted so that they are too great (or too small) as a consequence of the experimental design.

Measuring the height h is awkward. You can probably only find the value of h to within ±1 mm at best. So there is a random error in the value of h, and this will result in a slight scatter of the points on the graph, and a degree of uncertainty in the final value of g.

If you just have one value for h and the corresponding value for t you can use the uncertainty in h and t to find the uncertainty in g.

The percentage uncertainty in g = percentage uncertainty in h + 2 × percentage uncertainty in t.

For more about errors and combining uncertainties, see Chapter P1.

Measuring *g* using a ticker-timer

Figure 2.25 shows a weight falling. As it falls, it pulls a tape through a ticker-timer. The spacing of the dots on the tape increases steadily, showing that the weight is accelerating. You can analyse the tape to find the acceleration, as discussed in Practical Activity 2.1.

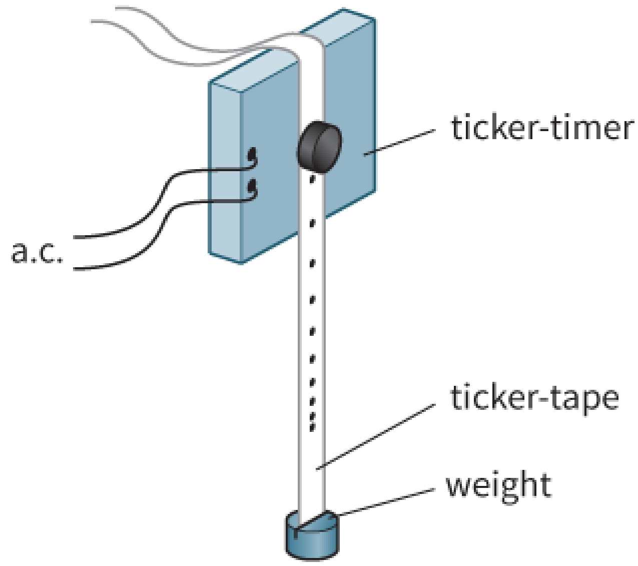


Figure 2.25: A falling weight pulls a tape through a ticker-timer.

This is not a very satisfactory method of measuring *g*. The main problem arises from friction between the tape and the ticker-timer. This slows the fall of the weight and so its acceleration is less than *g*. (This is another example of a systematic error.)

The effect of friction is less of a problem for a large weight, which falls more freely. If measurements are made for increasing weights, the value of acceleration gets closer and closer to the true value of *g*.

Measuring *g* using a light gate

Figure 2.26 shows how a weight can be attached to a card 'interrupt'. The card is designed to break the light beam twice as the weight falls. The computer can then calculate the velocity of the weight twice as it falls, and hence find its acceleration:

$$\text{initial velocity } u = \frac{x}{t_2 - t_1}$$

$$\text{final velocity } v = \frac{x}{t_4 - t_3}$$

Therefore:

$$\text{acceleration } a = \frac{v - u}{t_3 - t_1}$$

The weight can be dropped from different heights above the light gate. This allows you to find out whether its acceleration is the same at different points in its fall. This is an advantage over Method 1, which can only measure the acceleration from a stationary start.

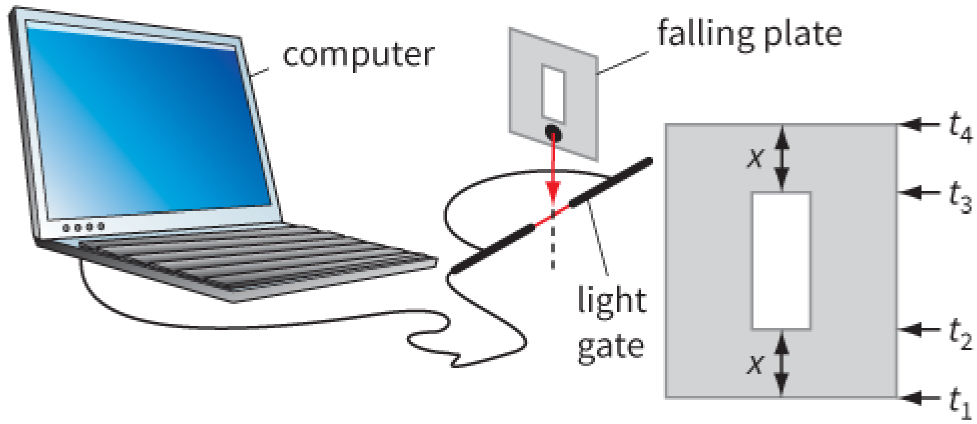


Figure 2.26: The weight accelerates as it falls. The upper section of the card falls more quickly through the light gate.

WORKED EXAMPLE

8 To get a rough value for *g*, a student dropped a stone from the top of a cliff. A second student timed the stone's fall using a stopwatch. Here are their results:

estimated height of cliff = 30 m

time of fall = 2.6 s

Use the results to estimate a value for *g*.

Step 1 Calculate the average speed of the stone:

$$\text{average speed of stone during fall} = \frac{30}{2.6} = 11.5 \text{ m s}^{-1}$$

Step 2 Find the values of *v* and *u*:

final speed $v = 2 \times 11.5 \text{ m s}^{-1} = 23.0 \text{ m s}^{-1}$

initial speed $u = 0 \text{ m s}^{-1}$

Step 3 Substitute these values into the equation for acceleration:

$$a = \frac{v-u}{t} = \frac{23.0}{2.6} = 8.8 \text{ m s}^{-2}$$

Note that you can reach the same result more directly using $s = ut + \frac{1}{2}at^2$ but you may find it easier to follow what is going on using the method given here. We should briefly consider why the answer is less than the expected value of $g = 9.81 \text{ m s}^{-2}$. It might be that the cliff was higher than the student's estimate. The timer may not have been accurate in switching the stopwatch on and off. There will have been air resistance that slowed the stone's fall.

Questions

18 A steel ball falls from rest through a height of 2.10 m. An electronic timer records a time of 0.67 s for the fall.

a Calculate the average acceleration of the ball as it falls.

b Suggest reasons why the answer is not exactly 9.81 m s^{-2}.

c Suppose the height is measured accurately but the time is measured to an uncertainty of ±0.02 s. Calculate the percentage uncertainty in the time and the percentage uncertainty in the average acceleration. You can do this by repeating the calculation for *g* using a time of 0.65 s. You can find out more about uncertainty in Chapter P1.

19 In an experiment to determine the acceleration due to gravity, a ball was timed electronically as it fell from rest through a height *h*. The times *t* shown in Table 2.5 were obtained.

a Plot a graph of *h* against t^2.

b From the graph, determine the acceleration of free fall *g*.

c Comment on your answer.

Height *h* / m	0.70	1.03	1.25	1.60	1.99
Time *t* / s	0.99	1.13	1.28	1.42	1.60

Table 2.5: Height h and time t data for Question 19.

20 In Chapter 1, we looked at how to use a motion sensor to measure the speed and position of a moving object. Suggest how a motion sensor could be used to determine *g*.

2.13 Motion in two dimensions: projectiles

A curved trajectory

A multiflash photograph can reveal details of the path, or trajectory, of a projectile. Figure 2.27 shows the trajectories of a projectile – a bouncing ball. Once the ball has left the child's hand and is moving through the air, the only force acting on it is its weight.

The ball has been thrown at an angle to the horizontal. It speeds up as it falls – you can see that the images of the ball become further and further apart. At the same time, it moves steadily to the right. You can see this from the even spacing of the images across the picture.

The ball's path has a mathematical shape known as a parabola. After it bounces, the ball is moving more slowly. It slows down, or decelerates, as it rises – the images get closer and closer together.

We interpret this picture as follows. The vertical motion of the ball is affected by the force of gravity, that is, its weight. When it rises it has a vertical deceleration of magnitude *g*, which slows it down, and when it falls it has an acceleration of *g*, which speeds it up. The ball's horizontal motion is unaffected by gravity. In the absence of air resistance, the ball has a constant velocity in the horizontal direction. We can treat the ball's vertical and horizontal motions separately, because they are independent of one another.

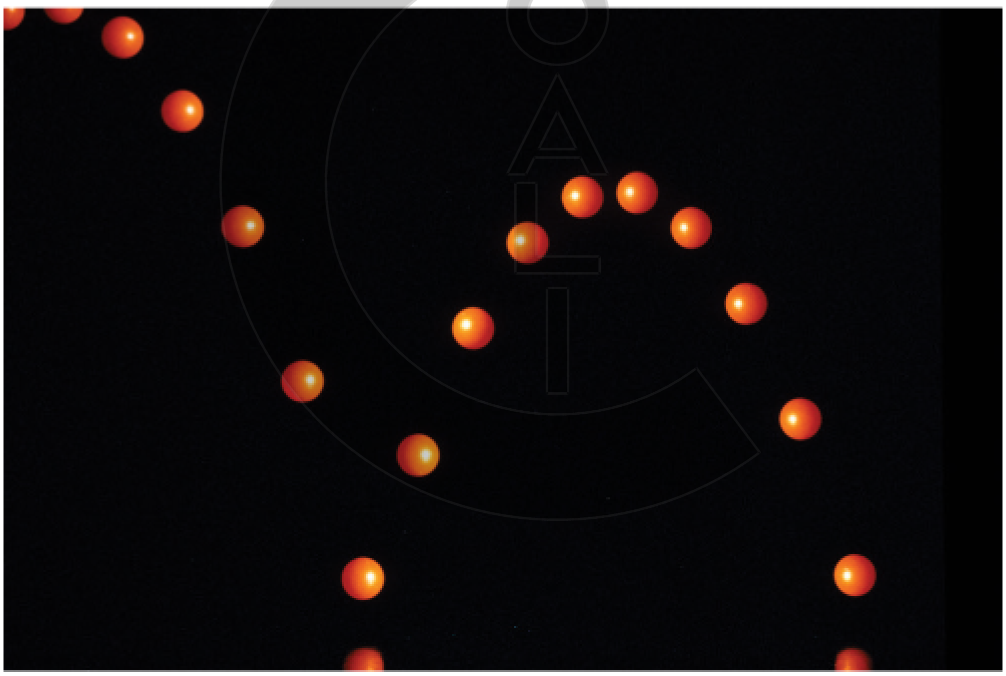


Figure 2.27: A bouncing ball is an example of a projectile. This multiflash photograph shows details of its motion that would escape the eye of an observer.

Components of a vector

In order to understand how to treat the velocity in the vertical and horizontal directions separately we start by considering a constant velocity.

If an aeroplane has a constant velocity v at an angle θ as shown in Figure 2.28, then we say that this velocity has two effects or **components**, v_N in a northerly direction and v_E in an easterly direction. These two components of velocity add up to make the actual velocity v.

This process of taking a velocity and determining its effect along another direction is known as **resolving** the velocity along a different direction. In effect, splitting the velocity into two components at right angles is the reverse of adding together two vectors – it is splitting one vector into two vectors along convenient directions.

KEY EQUATIONS

For a velocity v at an angle θ to the x-direction the components are:

x-direction: $v \cos \theta$

y-direction: $v \sin \theta$

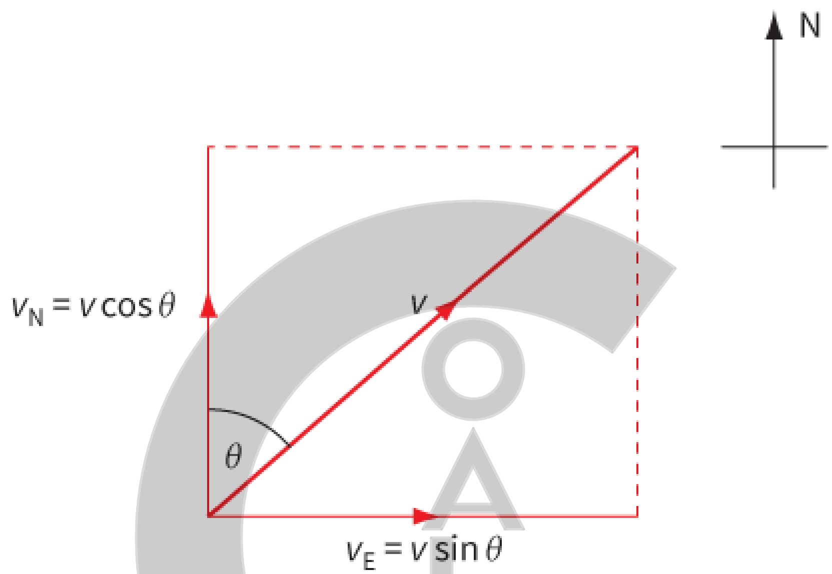


Figure 2.28: Components of a velocity. The component due north is $v_N = v \cos\theta$ and the component due east is $v_E = v \sin\theta$.

To find the component of any vector (for example, displacement, velocity, acceleration) in a particular direction, we can use the following strategy:

Step 1 Find the angle θ between the vector and the direction of interest.

Step 2 Multiply the vector by the cosine of the angle θ.

So the component of an object's velocity v at angle θ to v is equal to $v \cos \theta$ (Figure 2.28).

Question

21 Find the x- and y-components of each of the vectors shown in Figure 2.29. (You will need to use a protractor to measure angles from the diagram.)

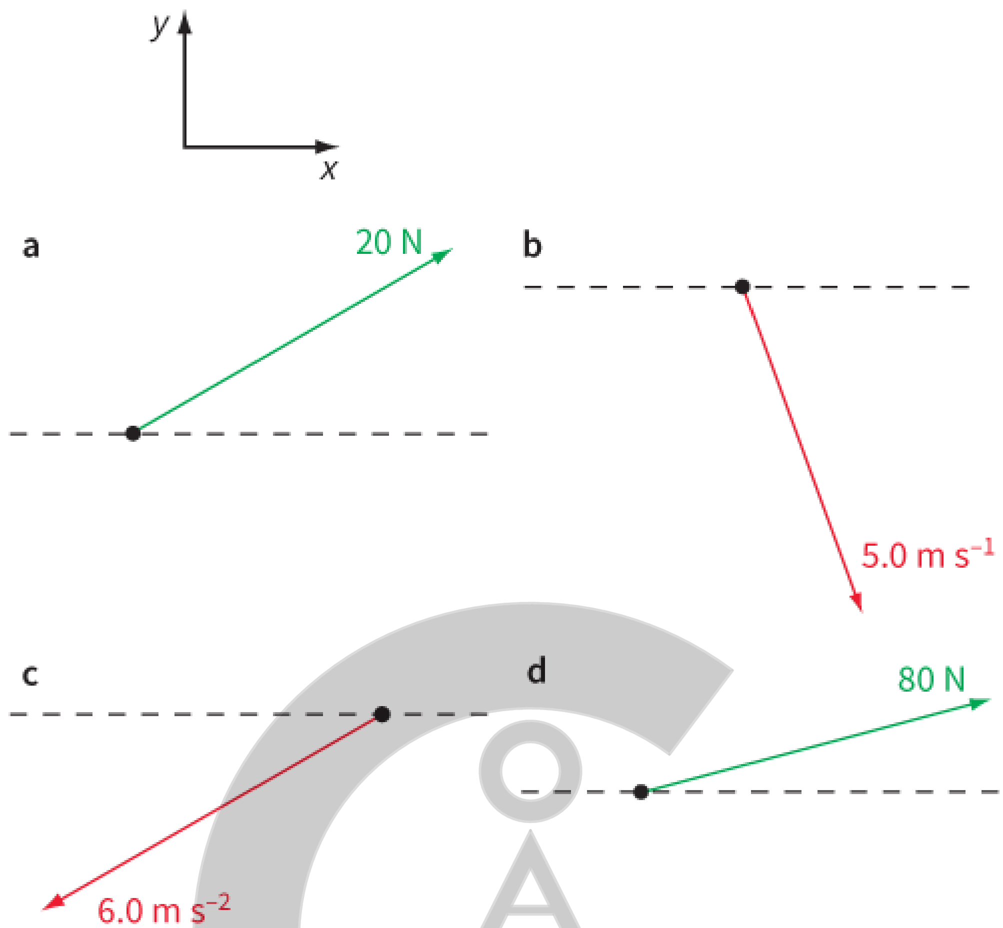


Figure 2.29: The vectors for Question 21.

2.14 Understanding projectiles

We will first consider the simple case of a projectile thrown straight up in the air, so that it moves vertically. Then we will look at projectiles that move horizontally and vertically at the same time.

Up and down

A stone is thrown upwards with an initial velocity of 20 m s^{-1}. Figure 2.30 shows the situation.

It is important to use a consistent sign convention here. We will take upwards as positive, and downwards as negative. So the stone's initial velocity is positive, but its acceleration *g* is negative. We can solve various problems about the stone's motion by using the equations of motion.

How high?

How high will the stone rise above ground level of the cliff?

As the stone rises upwards, it moves more and more slowly – it decelerates because of the force of gravity.

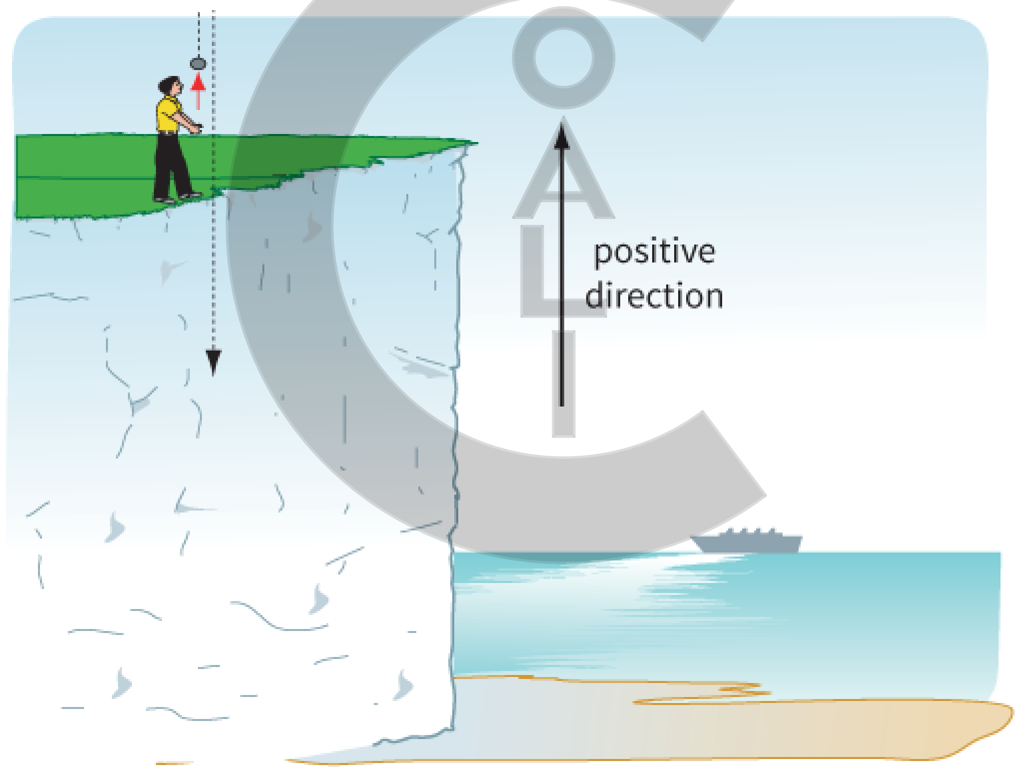


Figure 2.30: Standing at the edge of the cliff, you throw a stone vertically upwards. The height of the cliff is 25 m.

At its highest point, the stone's velocity is zero. So the quantities we know are:

initial velocity = u = 20 m s^{-1}

final velocity = v = 0 m s^{-1}

acceleration = a = −9.81 m s^{-2}

displacement = s = ?

The relevant equation of motion is $v^2 = u^2 + 2as$. Substituting values gives:

$$
\begin{aligned}
0^2 &= 20^2 + 2 \times (-9.81) \times s \\
0 &= 400 - 19.62\ s \\
s &= \frac{400}{19.62} \\
&= 20.4\ \text{m} \approx 20\ \text{m}
\end{aligned}
$$

The stone rises 20 m upwards before it starts to fall again.

How long?

How long will it take from leaving your hand for the stone to fall back to the clifftop?

When the stone returns to the point from which it was thrown, its displacement s is zero. So:

$$
\begin{aligned}
s &= 0 \\
u &= 20\ \text{m s}^{-1} \\
a &= -9.81\ \text{m s}^{-2} \\
t &= ?
\end{aligned}
$$

Substituting in $s = ut + \frac{1}{2}at^2$ gives:

$$
\begin{aligned}
0 &= 20t \times \tfrac{1}{2}(-9.81) \times t^2 \\
&= 20t - 4.905t^2 \\
&= (20 - 4.905t) \times t
\end{aligned}
$$

There are two possible solutions to this:

- t = 0 s; in other words, the stone had zero displacement at the instant it was thrown
- t = 4.1 s; in other words, the stone returned to zero displacement after 4.1 s, which is the answer we are interested in.

Falling further

The height of the cliff is 25 m. How long will it take the stone to reach the foot of the cliff?

This is similar to the last example, but now the stone's final displacement is 25 m below its starting point. By our sign convention, this is a negative displacement and $s = -25$ m.

Questions

22 In the example in 'Falling further', calculate the time it will take for the stone to reach the foot of the cliff.

23 A ball is fired upwards with an initial velocity of 30 m s^{-1}. Table 2.6 shows how the ball's velocity changes. (Take g = 9.81 m s^{-2}.)

- **a** Copy and complete the table.
- **b** Draw a graph to represent the data.
- **c** Use your graph to deduce how long the ball took to reach its highest point.

Velocity / m s^{-1}	30	20.19				
Time / s	0	1.0	2.0	3.0	4.0	5.0

Table 2.6: For Question 23.

Vertical and horizontal at the same time

Here is an example to illustrate what happens when an object travels vertically and horizontally at the same time.

In a toy, a ball-bearing is fired horizontally from a point 0.4 m above the ground. Its initial velocity is 2.5 m s^{-1}. Its positions at equal intervals of time have been calculated and are shown in Table 2.7. These results are also shown in Figure 2.31. Study the table and the graph. You should notice the following:

- The horizontal distance increases steadily. This is because the ball's horizontal motion is unaffected by the force of gravity. It travels at a steady velocity horizontally so we can use $v = \frac{s}{t}$
- The vertical distances do not show the same pattern. The ball is accelerating downwards so we must use the equations of motion. (These figures have been calculated using g = 9.81 m s^{-2}.)

Time / s	Horizontal distance / m	Vertical distance / m
0.00	0.00	0.000
0.04	0.10	0.008
0.08	0.20	0.031
0.12	0.30	0.071
0.16	0.40	0.126
0.20	0.50	0.196
0.24	0.60	0.283
0.28	0.70	0.385

Table 2.7: Data for the example of a moving ball, as shown in Figure 2.31.

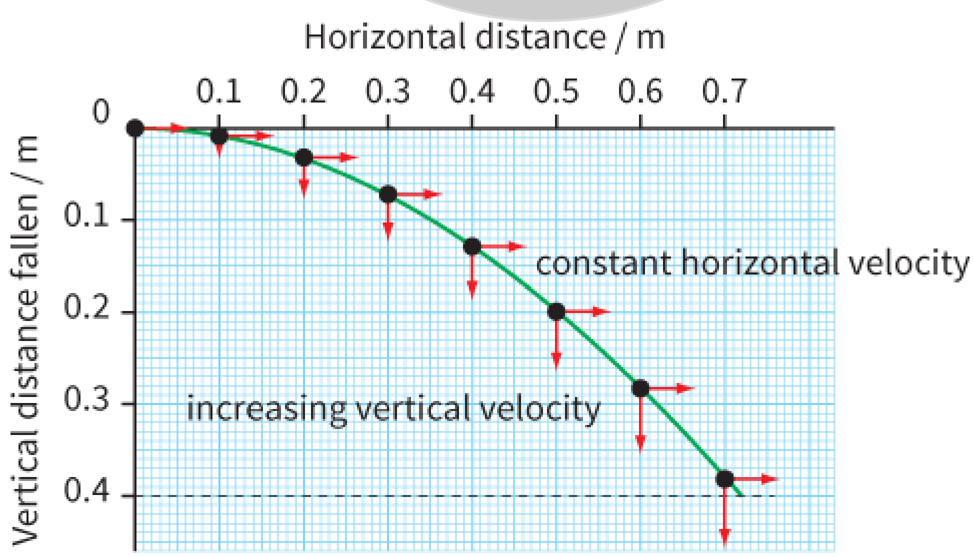


Figure 2.31: This sketch shows the path of the ball projected horizontally. The arrows represent the horizontal and vertical components of its velocity.

You can calculate the distance *s* fallen using the equation of motion $s = ut + \frac{1}{2}at^2$ (The initial vertical velocity $u = 0$.)

The horizontal distance is calculated using:

horizontal distance = 2.5 × *t*

The vertical distance is calculated using:

$$\text{vertical distance} = \frac{1}{2} \times 9.81 \times t^2$$

KEY IDEA

In the absence of air resistance, an object has constant velocity horizontally and constant acceleration vertically.

WORKED EXAMPLES

9 A stone is thrown horizontally with a velocity of 12 m s^{-1} from the top of a vertical cliff. Calculate how long the stone takes to reach the ground 40 m below and how far the stone lands from the base of the cliff.

Step 1 Consider the ball's vertical motion. It has zero initial speed vertically and travels 40 m with acceleration 9.81 m s^{-2} in the same direction.

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

$$40 = 0 + \frac{1}{2} \times 9.81 \times t^2$$

So, $t = 2.86$ s.

Step 2 Consider the ball's horizontal motion. The ball travels with a constant horizontal velocity, 12 m s^{-1}, as long as there is no air resistance.

distance travelled = $u \times t = 12 \times 2.86 = 34.3$ m

Hint: *You may find it easier to summarise the information like this:*

vertically $s = 40$ $u = 0$ $a = 9.81$ $t = ?$ $v = ?$

horizontally $u = 12$ $v = 12$ $a = 0$ $t = ?$ $s = ?$

10 A ball is thrown with an initial velocity of 20 m s^{-1} at an angle of 30° to the horizontal (Figure 2.32). Calculate the horizontal distance travelled by the ball (its **range**).

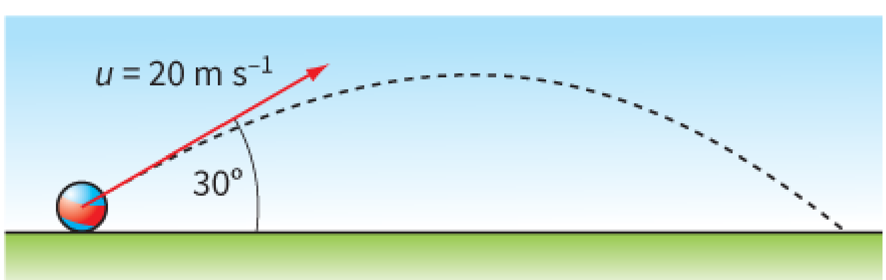


Figure 2.32: For Worked example 10.

Step 1 Split the ball's initial velocity into horizontal and vertical components:

initial velocity = u = 20 m s^{-1}

horizontal component of initial velocity v= $u \cos \theta = 20 \times \cos 30° = 17.3$ m s^{-1}

vertical component of initial velocity = $u \sin \theta = 20 \times \sin 30° = 10$ m s^{-1}

Step 2 Consider the ball's vertical motion. How long will it take to return to the ground? In other words, when will its displacement return to zero?

$u = 10$ m s^{-1} $a = -9.81$ m s^{-2} $s = 0$ $t = ?$

Using $s = ut + \frac{1}{2}at^2$ we have:

$0 = 10t - 4.905t^2$

This gives $t = 0$ s or $t = 2.04$ s.

So, the ball is in the air for 2.04 s.

Step 3 Consider the ball's horizontal motion. How far will it travel horizontally in the 2.04 s before it lands? This is simple to calculate, since it moves with a constant horizontal velocity of 17.3 m s^{-1}.

$$\begin{aligned} \text{horizontal displacement } s &= 17.3 \times 2.04 \\ &= 35.3 \text{ m} \end{aligned}$$

Hence the horizontal distance travelled by the ball (its range) is about 35 m.

Questions

24 A stone is thrown horizontally from the top of a vertical cliff and lands 4.0 s later at a distance 12.0 m from the base of the cliff. Ignore air resistance.

a Calculate the horizontal speed of the stone.

b Calculate the height of the cliff.

25 A stone is thrown with a velocity of 8.0 m s^{-1} into the air at an angle of 40° to the horizontal.

a Calculate the vertical component of the velocity.

b State the value of the vertical component of the velocity when the stone reaches its highest point. Ignore air resistance.

c Use your answers to part **a** and part **b** to calculate the time the stone takes to reach its highest point.

d Calculate the horizontal component of the velocity.

e Use your answers to part **c** and part **d** to find the horizontal distance travelled by the stone as it climbs to its highest point.

26 The range of a projectile is the horizontal distance it travels before it reaches the ground. The greatest range is achieved if the projectile is thrown at 45° to the horizontal.

A ball is thrown with an initial velocity of 40 m s^{-1}. Calculate its greatest possible range when air resistance is considered to be negligible.

REFLECTION

- Could you easily teach somebody a proof of the equations of motion? How would you do this?
- What do you find unexpected about projectile motion?

SUMMARY

Acceleration is equal to the rate of change of velocity. It is a vector, has units m s^{-2} and can be found from the gradient of a velocity–time graph. The area under this graph is the change in displacement.

Acceleration, velocity, displacement and time for a uniform acceleration are related by the equations of motion, which you should know how to derive and use.

The acceleration of free fall is taken as 9.81 m s^{-2} and you should know an experiment to measure this quantity.

Vector quantities can be resolved into components. Components at right angles to one another can be treated independently. For a velocity v at an angle θ to the x-direction the components are:

x-direction: $v \cos \theta$

y-direction: $v \sin \theta$

In the absence of air resistance, projectiles involve a constant acceleration downwards and a constant velocity horizontally. These can be treated independently of one another.

EXAM-STYLE QUESTIONS

1 An aircraft, starting from rest accelerates uniformly along a straight runway. It reaches a speed of 200 km h^{-1} and travels a distance of 1.4 km.

What is the acceleration of the aircraft along the runway? **[1]**

A 1.1 m s^{-2}

B 2.2 m s^{-2}

C 3.0 m s^{-2}

D 6.0 m s^{-2}

2 A ball is thrown with a velocity of 10 m s^{-1} at an angle of 30° to the horizontal. Air resistance has a negligible effect on the motion of the ball.

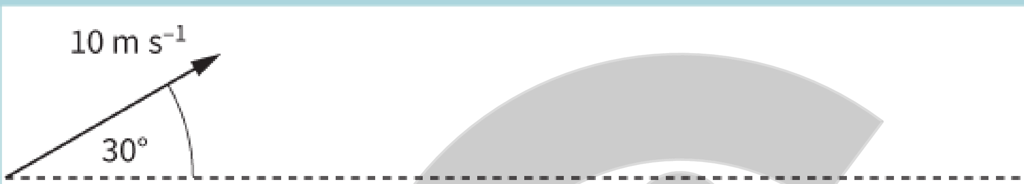


Figure 2.33

What is the velocity of the ball at the highest point in its path? **[1]**

A 0

B 5.0 m s^{-1}

C 8.7 m s^{-1}

D 10 m s^{-1}

3 A trolley travels along a straight track. The variation with time t of the velocity v of the trolley is shown. **[1]**

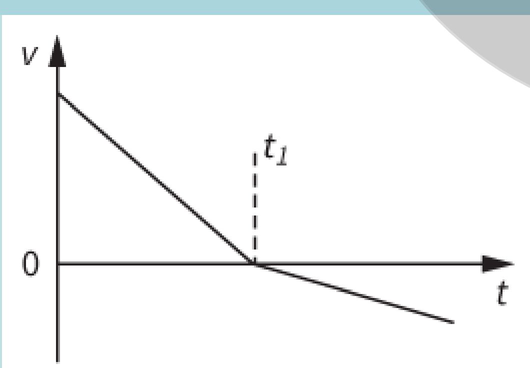
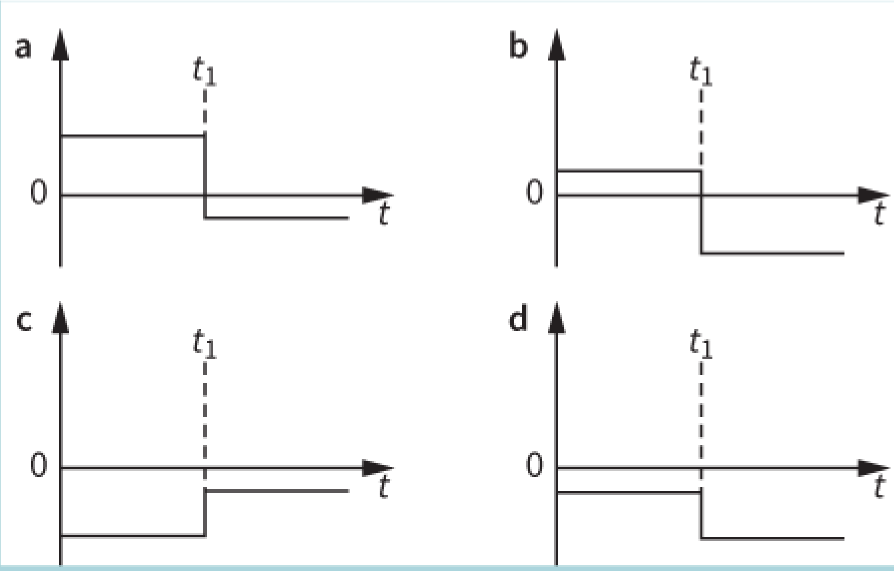


Figure 2.34

Which graph shows the variation with time of the acceleration a of the trolley?



4 A motorway designer can assume that cars approaching a motorway enter a slip road with a velocity of 10 m s^{-1} and reach a velocity of 30 m s^{-1} before joining the motorway. Calculate the minimum length for the slip road, assuming that vehicles have an acceleration of 4.0 m s^{-2}. **[4]**

5 A train is travelling at 50 m s^{-1} when the driver applies the brakes and gives the train a constant deceleration of magnitude 0.50 m s^{-2} for 100 s. Describe what happens to the train. Calculate the distance travelled by the train in 100 s. **[7]**

6 A boy stands on a cliff edge and throws a stone vertically upwards at time $t = 0$. The stone leaves his hand at 20 m s^{-1}. Take the acceleration of the ball as 9.81 m s^{-2}.

a Show that the equation for the displacement of the ball is:

$s = 20t - 4.9t^2$ **[2]**

b Calculate the height of the stone 2.0 s after release and 6.0 s after release. **[3]**

c Calculate the time taken for the stone return to the level of the boy's hand. You may assume the boy's hand does not move vertically after the ball is released. **[4]**

[Total: 9]

7 This graph shows the variation of velocity with time of two cars, A and B, which are travelling in the same direction over a period of time of 40 s.

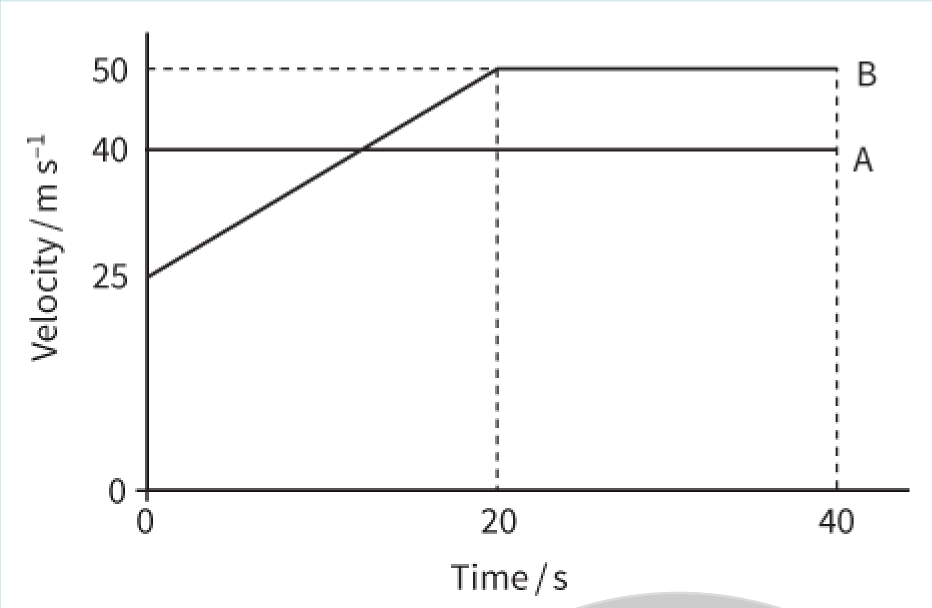


Figure 2.35

Car A, travelling at a constant velocity of 40 m s^{-1}, overtakes car B at time t = 0. In order to catch up with car A, car B immediately accelerates uniformly for 20 s to reach a constant velocity of 50 m s^{-1}. Calculate:

a the distance that A travels during the first 20 s **[2]**

b the acceleration and distance of travel of B during the first 20 s **[5]**

c the additional time taken for B to catch up with A **[2]**

d the distance each car will have then travelled since t = 0. **[2]**

[Total: 11]

8 An athlete competing in the long jump leaves the ground with a velocity of 5.6 m s^{-1} at an angle of 30° to the horizontal.

a Determine the vertical component of the velocity and use this value to find the time between leaving the ground and landing. **[4]**

b Determine the horizontal component of the velocity and use this value to find the horizontal distance travelled. **[4]**

[Total: 8]

9 This diagram shows an arrangement used to measure the acceleration of a metal plate as it falls vertically.

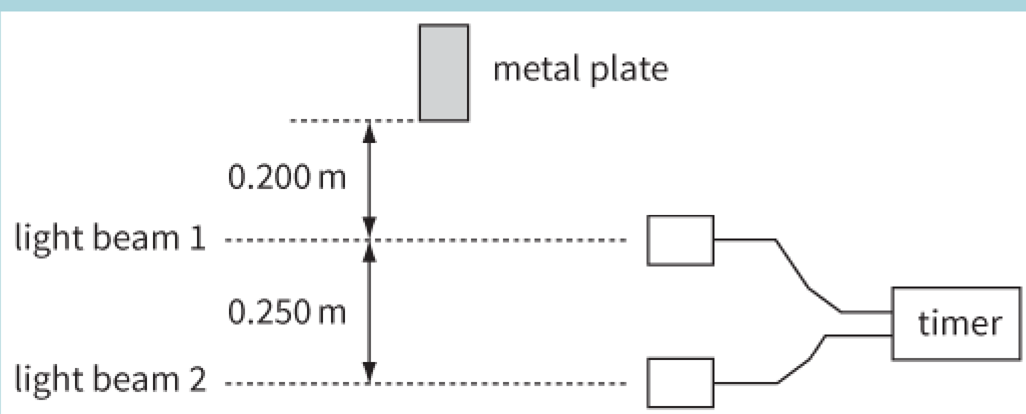


Figure 2.36

The metal plate is released from rest and falls a distance of 0.200 m before breaking light beam 1. It then falls a further 0.250 m before breaking light beam 2.

a Calculate the time taken for the plate to fall 0.200 m from rest. (You may assume that the metal plate falls with an acceleration equal to the acceleration of free fall.) **[2]**

b The timer measures the speed of the metal plate as it falls through each light beam. The speed as it falls through light beam 1 is 1.92 m s^{-1} and the speed as it falls through light beam 2 is 2.91 m s^{-1}.

i Calculate the acceleration of the plate between the two light beams. **[2]**

ii State and explain one reason why the acceleration of the plate is not equal to the acceleration of free fall. **[2]**

[Total: 6]

10 This is a velocity–time graph for a vertically bouncing ball.

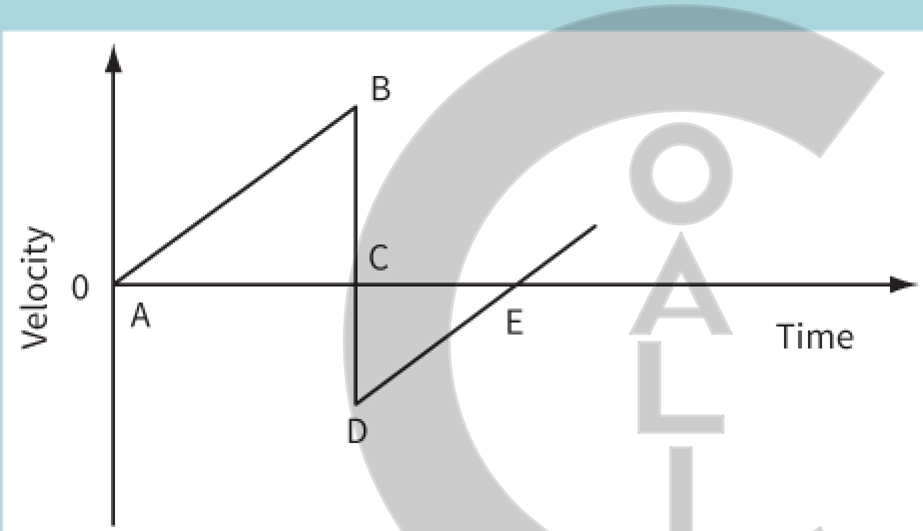


Figure 2.37

The ball is released at A and strikes the ground at B. The ball leaves the ground at D and reaches its maximum height at E. The effects of air resistance can be neglected.

a State:

i why the velocity at D is negative **[1]**

ii why the gradient of the line AB is the same as the gradient of line DE **[1]**

iii what is represented by the area between the line AB and the time axis **[1]**

iv why the area of triangle ABC is greater than the area of triangle CDE. **[1]**

b The ball is dropped from rest from an initial height of 1.2 m. After hitting the ground the ball rebounds to a height of 0.80 m. The ball is in contact with the ground between B and D for a time of 0.16 s.

Using the acceleration of free fall, calculate:

i the speed of the ball immediately before hitting the ground **[2]**

ii the speed of the ball immediately after hitting the ground **[2]**

iii the acceleration of the ball while it is in contact with the ground. State the direction of this acceleration. **[3]**

[Total: 11]

11 A student measures the speed v of a trolley as it moves down a slope. The variation of v with time t is shown in this graph.

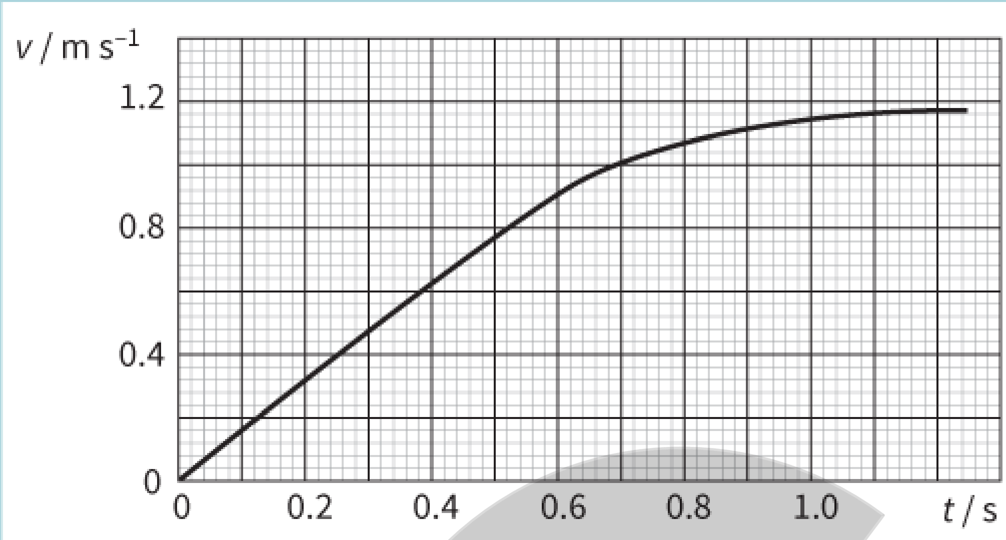


Figure 2.38

a Use the graph to find the acceleration of the trolley when t = 0.70 s. [2]

b State how the acceleration of the trolley varies between t = 0 and t = 1.0 s. Explain your answer by reference to the graph. [3]

c Determine the distance travelled by the trolley between t = 0.60 and t = 0.80 s. [3]

d The student obtained the readings for v using a motion sensor. The readings may have random errors and systematic errors. Explain how these two types of error affect the velocity–time graph. [2]

[Total: 10]

12 A car driver is travelling at speed v on a straight road. He comes over the top of a hill to find a fallen tree on the road ahead. He immediately brakes hard but travels a distance of 60 m at speed v before the brakes are applied. The skid marks left on the road by the wheels of the car are of length 140 m, as shown.

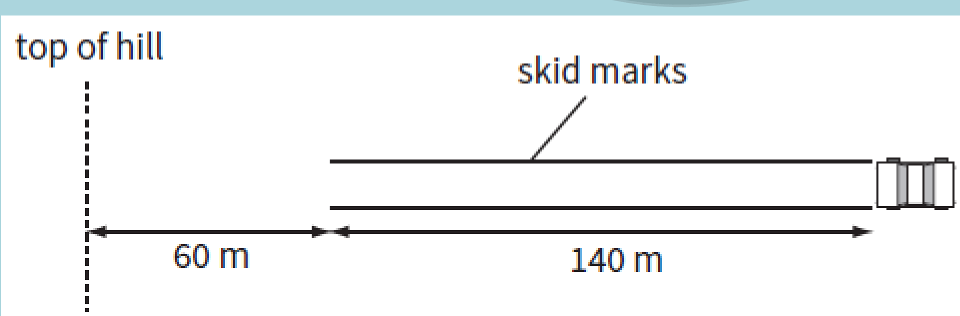


Figure 2.39

The police investigate whether the driver was speeding and establish that the car decelerates at 2.0 m s^{-2} during the skid.

a Determine the initial speed v of the car before the brakes are applied. [2]

b Determine the time taken between the driver coming over the top of the hill and applying the brakes. Suggest whether this shows whether the driver was [2] alert to the danger.

c The speed limit on the road is 100 km/h. Determine whether the driver was breaking the speed limit. **[2]**

[Total: 6]

13 A hot-air balloon rises vertically. At time $t = 0$, a ball is released from the balloon. This graph shows the variation of the ball's velocity v with t. The ball hits the ground at $t = 4.1$ s.

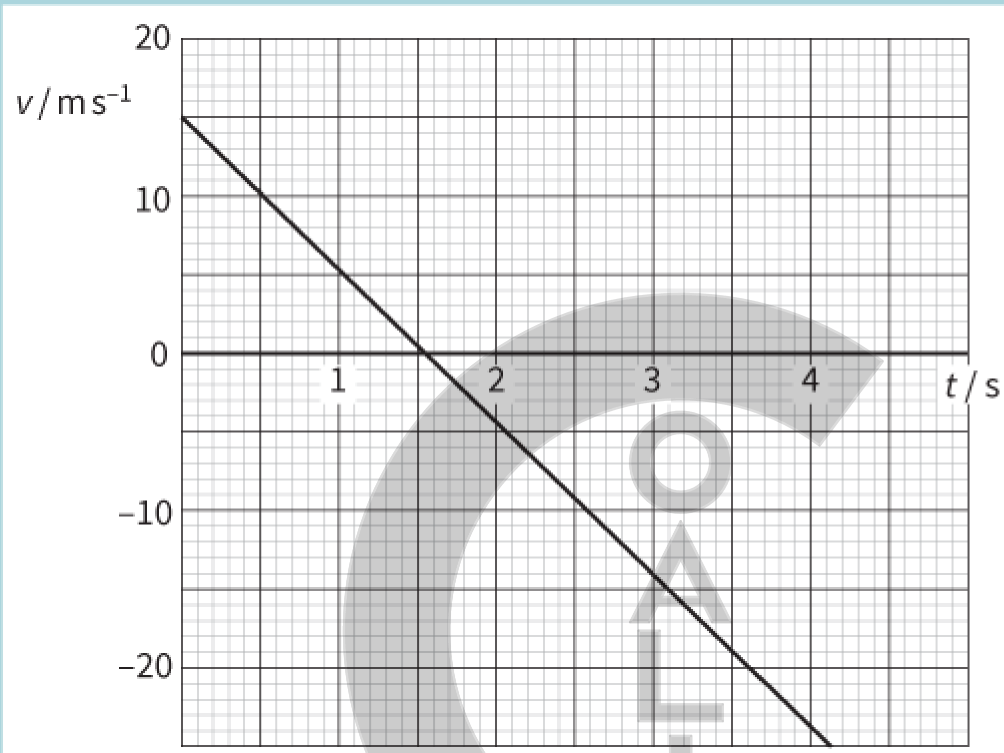


Figure 2.40

a Explain how the graph shows that the acceleration of the ball is constant. **[1]**

b Use the graph to:

i determine the time at which the ball reaches its highest point **[1]**

ii show that the ball rises for a further 12 m between release and its highest point **[2]**

iii determine the distance between the highest point reached by the ball and the ground. **[2]**

c The equation relating v and t is $v = 15 - 9.81t$. State the significance in the equation of:

i the number 15 **[1]**

ii the negative sign. **[1]**

[Total: 8]

14 An aeroplane is travelling horizontally at a speed of 80 m s^{-1} and drops a crate of emergency supplies.

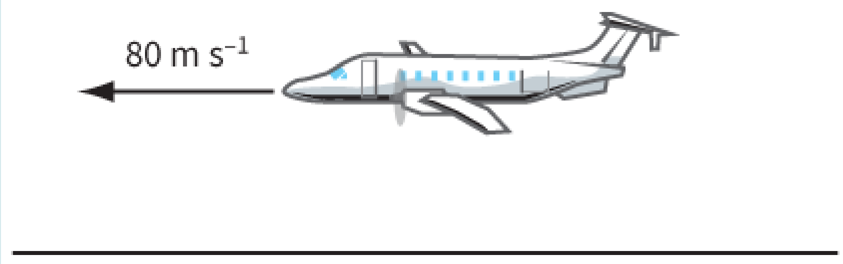


Figure 2.41

To avoid damage, the maximum vertical speed of the crate on landing is 20 m s^{-1}. You may assume air resistance is negligible.

a Calculate the maximum height of the aeroplane when the crate is dropped. **[2]**

b Calculate the time taken for the crate to reach the ground from this height. **[2]**

c The aeroplane is travelling at the maximum permitted height. Calculate the horizontal distance travelled by the crate after it is released from the aeroplane. **[1]**

[Total: 5]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define acceleration	2.1			
calculate displacement from the area under a velocity–time graph	2.5			
calculate velocity using the gradient of a displacement–time graph	2.6			
calculate acceleration using the gradient of a velocity–time graph	2.4			
derive and use the equations of uniformly accelerated motion	2.10			
describe an experiment to measure the acceleration of free fall, *g*	2.11, 2.12			
use perpendicular components to represent a vector	2.13			
explain projectile motion using uniform velocity in one direction and uniform acceleration in a perpendicular direction and do calculations on this motion.	2.14			

〉 Chapter 3

Dynamics: explaining motion

LEARNING INTENTIONS

In this chapter you will learn how to:

- recognise that mass is a property of an object that resists change in motion
- identify the forces acting on a body in different situations
- describe how the motion of a body is affected by the forces acting on it
- recall $F = ma$ and solve problems using it, understanding that acceleration and resultant force are always in the same direction
- state and apply Newton's first and third laws of motion
- recall that the weight of a body is equal to the product of its mass and the acceleration of free fall
- relate derived units to base units in the SI system and use base units to check the homogeneity of an equation
- recall and use a range of prefixes.

BEFORE YOU START

- Make a list of all the different types of force that you know about. Do you have the same list as someone else? Discuss any differences and describe the types of force to each other.
- What prefixes do you know that may be placed before a unit? For example, the 'c' in cm is the prefix 'centi' and means times 10^{-2}. Write down those that you know and what they mean then see if you are correct.

DYNAMIC AEROPLANES

Figure 3.1 shows a modern aeroplane. To decrease cost and the effect on the environment, such an aircraft must *reduce* air resistance and weight, yet be able to *use* air resistance and other forces to stop when landing. If you have ever flown in an aeroplane you will know how the back of the seat pushes you forwards when the aeroplane accelerates down the runway. The pilot must control many forces on the aeroplane in take-off, flying and landing.

In Chapters 1 and 2 we saw how motion can be described in terms of displacement, velocity, acceleration and so on. Now we are going to look at how we can explain how an object moves in terms of the forces that change its motion.

Apart from air resistance, see how many other forces you can discover that act on an aeroplane. Compare your list with someone else. What causes all these forces?

Figure 3.1: A modern aircraft flying over the ocean.

3.1 Force, mass and acceleration

Figure 3.2a shows how we represent the force that the motors on a train provide to cause it to accelerate. The resultant force is represented by a green arrow. The direction of the arrow shows the direction of the resultant force. The magnitude (size) of the resultant force of 20 000 N is also shown.

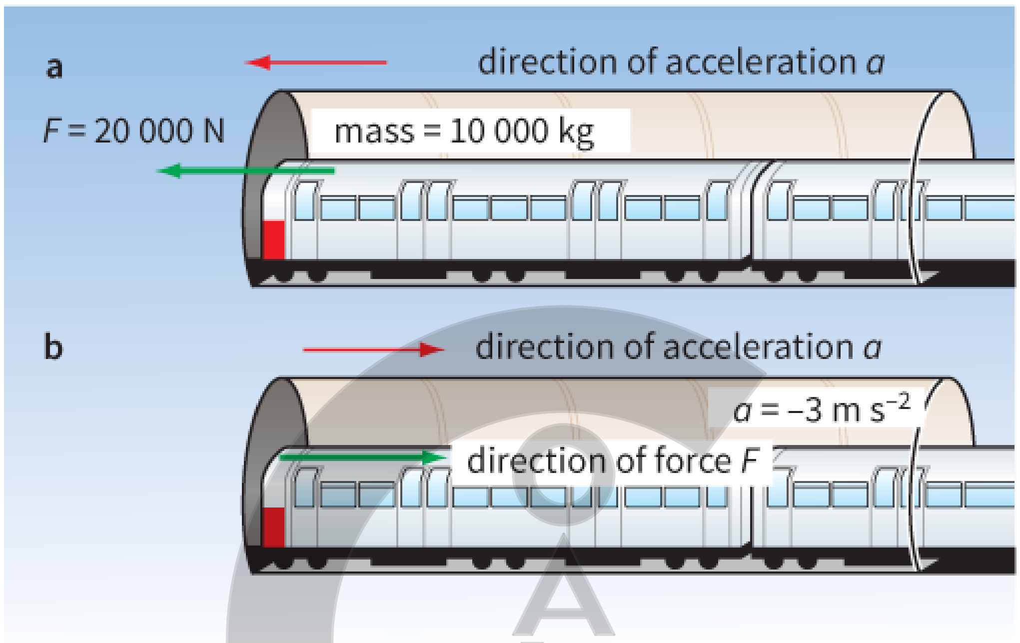


Figure 3.2: A force is needed to make the train *a* accelerate, and *b* decelerate.

To calculate the acceleration *a* of the train produced by the resultant force *F*, we must also know the train's mass *m* (Table 3.1). These quantities are related by:

$$a = \frac{F}{m} \text{ or } F = ma$$

KEY EQUATION

resultant force = mass × acceleration

$$F = ma$$

Quantity	Symbol	Unit
resultant force	*F*	N (newtons)
mass	*m*	kg (kilograms)
acceleration	*a*	$m\ s^{-2}$ (metres per second squared)

Table 3.1: The quantities related by *F* = *ma*.

In this example, we have F = 20 000 N and m = 10 000 kg, and so:

$$a = \frac{F}{m} = \frac{20\,000}{10\,000} = 2 \text{ m s}^{-2}$$

In Figure 3.2b, the train is decelerating as it comes into a station. Its acceleration is −3.0 m s^{-2}. What force must be provided by the braking system of the train?

$$F = ma = 10\,000 \times -3 = -30\,000 \text{ N}$$

The minus sign shows that the force must act towards the right in the diagram, in the opposite direction to the motion of the train.

Newton’s second law of motion

The equation we used, $F = ma$, is a simplified version of **Newton’s second law of motion**: For a body of constant mass, its acceleration is directly proportional to the resultant force applied to it.

An alternative form of Newton’s second law is given in Chapter 6, when you have studied momentum.

Since Newton’s second law holds for objects that have a constant mass, this equation can be applied to a train whose mass remains constant during its journey.

The equation $a = \frac{F}{m}$ relates acceleration, resultant force and mass. In particular, it shows that the bigger the force, the greater the acceleration it produces. You will probably feel that this is an unsurprising result. For a given object, the acceleration is directly proportional to the resultant force:

$$a \propto F$$

The equation also shows that the acceleration produced by a force depends on the mass of the object. The **mass** of an object is a measure of its **inertia**, or its ability to resist any change in its motion. The greater the mass, the smaller the acceleration that results. If you push your hardest against a small car (which has a small mass), you will have a greater effect than if you push against a more massive car (Figure 3.3). So, for a constant force, the acceleration is inversely proportional to the mass:

$$a \propto \frac{1}{m}$$

The train driver knows that when the train is full during the rush hour, it has a smaller acceleration. This is because its mass is greater when it is full of people. Similarly, it is more difficult to stop the train once it is moving. The brakes must be applied earlier to avoid the train overshooting the platform at the station.

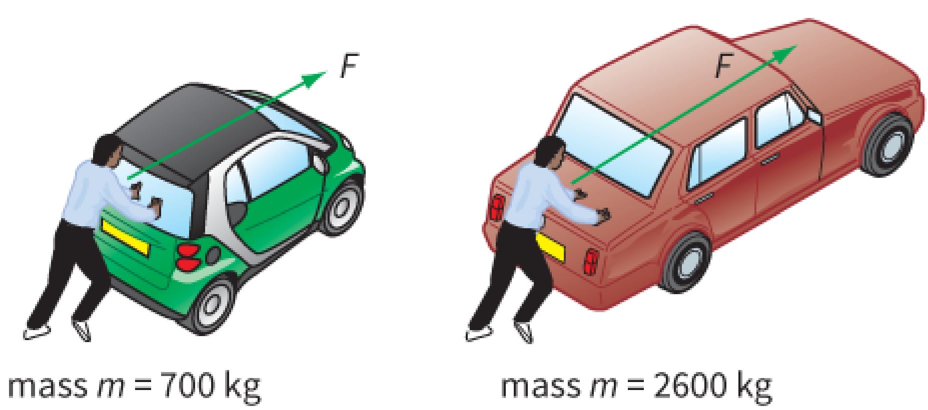


Figure 3.3: It is easier to make a small mass accelerate than a large mass.

WORKED EXAMPLES

1 A cyclist of mass 60 kg rides a bicycle of mass 20 kg. When starting off, the cyclist provides a force of 200 N. Calculate the initial acceleration.

Step 1 This is a straightforward example. First, we must calculate the combined mass *m* of the bicycle and its rider:

$m = 20 + 60 = 80$ kg

We are given the force *F*:

force causing acceleration $F = 200$ N

Step 2 Substituting these values gives:

$$a = \frac{F}{m} = \frac{200}{80} = 2.5 \text{ m s}^{-2}$$

So the cyclist's acceleration is 2.5 m s^{-2}.

2 A car of mass 500 kg is travelling at 20 m s^{-1}. The driver sees a red traffic light ahead, and slows to a halt in 10 s. Calculate the braking force provided by the car.

Step 1 In this example, we must first calculate the acceleration required. The car's final velocity is 0 m s^{-1}, so its change in velocity $\Delta v = 0 - 20 = -20 \text{ m s}^{-1}$

$$\text{acceleration } a = \frac{\text{change in velocity}}{\text{time taken}} = \frac{\Delta v}{\Delta t} = \frac{-20}{10} = -2 \text{ m s}^{-2}$$

Step 2 To calculate the force, we use:

$F = ma = 500 \times -2 = -1000$ N

So the brakes must provide a force of 1000 N. (The minus sign shows a force decreasing the velocity of the car.)

Questions

1 Calculate the force needed to give a car of mass 800 kg an acceleration of 2.0 m s^{-2}.

2 A rocket has a mass of 5000 kg. At a particular instant, the resultant force acting on the rocket is 200 000 N. Calculate its acceleration.

3 (In this question, you will need to make use of the equations of motion that you studied in Chapter 2.) A motorcyclist of mass 60 kg rides a bike of mass 40 kg. As she sets off from the lights, the forward force on the bike is 200 N. Assuming the resultant force on the bike remains constant, calculate the bike's velocity after 5.0 s.

3.2 Identifying forces

It is important to be able to identify the forces which act on an object. When we know what forces are acting, we can predict how the object will move. Table 3.2 shows some important forces, how they arise and how we represent them in diagrams.

3.3 Weight, friction and gravity

Now we need to consider some specific forces – such as **weight** and **friction**.

When Isaac Newton was confined to his rural home to avoid the plague which was spreading uncontrollably in other parts of England, he is said to have noticed an apple fall to the ground. From this, he developed his theory of gravity that relates the motion of falling objects here on Earth to the motion of the Moon around the Earth, and the planets around the Sun.

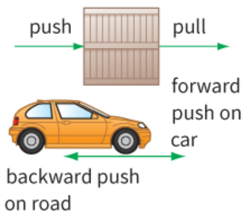
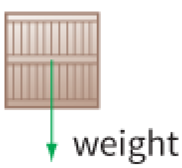
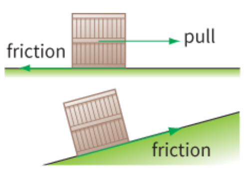
Diagram	Force	Important situations
	**Pushes and pulls.** You can make an object accelerate by pushing and pulling it. Your force is shown by an arrow pushing (or pulling) the object. The engine of a car provides a force to push backwards on the road. Frictional forces from the road on the tyre push the car forwards.	• pushing and pulling • lifting • force of car engine • attraction and repulsion by magnets and by electric charges
	**Weight.** This is the force of gravity acting on the object. It is usually shown by an arrow pointing vertically downwards from the object's centre of gravity.	• any object in a gravitational field • less on the Moon
	**Friction.** This is the force that arises when two surfaces rub over one another. If an object is sliding along the ground, friction acts in the opposite direction to its motion. If an object is stationary, but tending to slide – perhaps because it is on a slope – the force of friction acts up the slope to stop it from sliding down. Friction always acts along a surface, never at an angle to it.	• pulling an object along the ground • vehicles cornering or skidding • sliding down a slope
	**Drag.** This force is similar to friction. When an object moves through air, there is friction between it and the air. Also, the object has to push aside the air as it moves along. Together, these effects make up drag. Similarly, when an object moves through a liquid, it experiences a drag force. Drag acts to oppose the motion of an object; it acts in the opposite direction to the object's velocity. It can be reduced by giving the object a streamlined shape.	• vehicles moving • aircraft flying • parachuting • objects falling through air or water • ships sailing

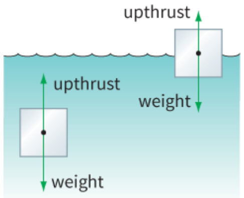
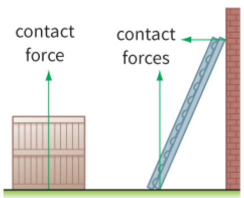
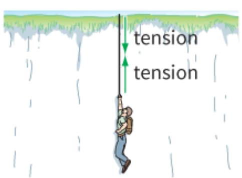
Diagram	Force	Important situations
	**Upthrust.** Any object placed in a fluid such as water or air experiences an upwards force. This is what makes it possible for something to float in water. Upthrust arises from the pressure that a fluid exerts on an object. The deeper you go, the greater the pressure. So there is more pressure on the lower surface of an object than on the upper surface, and this tends to push it upwards. If upthrust is greater than the object's weight, it will float up to the surface.	• boats and icebergs floating • people swimming divers surfacing a hot air balloon rising •
	**Contact force.** When you stand on the floor or sit on a chair, there is usually a force that pushes up against your weight, and which supports you so that you do not fall down. The contact force is sometimes known as the normal contact force of the floor or chair. (In this context, normal means 'perpendicular'.) The contact force always acts at right angles to the surface that produces it. The floor pushes straight upwards; if you lean against a wall, it pushes back against you horizontally.	• standing on the ground • one object sitting on top of another • leaning against a wall • one object bouncing off another
	**Tension.** This is the force in a rope or string when it is stretched. If you pull on the ends of a string, it tends to stretch. The tension in the string pulls back against you. It tries to shorten the string. Tension can also act in springs. If you stretch a spring, the tension pulls back to try to shorten the spring. If you squash (compress) the spring, the tension acts to expand the spring.	• pulling with a rope • squashing or stretching a spring

Table 3.2: Some important forces.

The force that caused the apple to accelerate was the pull of the Earth's gravity. Another name for this force is the **weight** of the apple. The force is shown as an arrow, pulling vertically downwards on the apple (Figure 3.4). It is usual to show the arrow coming from the centre of the apple – its **centre of gravity**. The centre of gravity of an object is defined as the point where its entire weight appears to act.

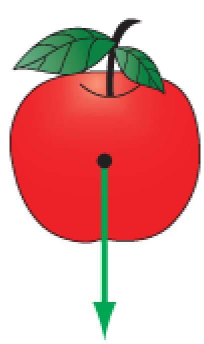


Figure 3.4: The weight of an object is a force caused by the Earth's gravity. It acts vertically down on the object.

Large and small

A large rock has a greater weight than a small rock, but if you push both rocks over a cliff at the same time, they will fall at the same rate. In other words, they have the **same** acceleration, regardless of their mass. This is a surprising result. Common sense may suggest that a heavier object will fall faster than a lighter one. It is said that Galileo dropped a large cannon ball and a small cannon ball from the top of the Leaning Tower of Pisa in Italy, and showed that they landed at the same time. The story illustrates that results are not always what you think they will be – if everyone thought that the two cannon balls would accelerate at the same rate, there would not have been any experiment or story.

In fact, we are used to lighter objects falling more slowly than heavy ones. A feather drifts down to the floor, while a stone falls quickly. But this is because of air resistance. The force of air resistance has a large effect on the falling feather, and almost no effect on the falling stone. When astronauts visited the Moon (where there is virtually no atmosphere and so no air resistance), they were able to show that a feather and a stone fell side-by-side to the ground.

As we saw in Chapter 2, an object falling freely close to the Earth's surface has an acceleration of roughly 9.81 m s^{-2}, the acceleration of free fall g.

We can find the force causing this acceleration using $F = ma$. This force is the object's weight. Hence, the weight W of an object is given by:

$$\text{weight} = \text{mass} \times \text{acceleration of free fall}$$

or

$$W = mg$$

KEY EQUATION

$$\text{weight} = \text{mass} \times \text{acceleration of free fall}$$
$$W = mg$$

Question

4 Estimate the mass and weight of each of the following at the surface of the Earth:

- **a** a kilogram of potatoes
- **b** an average student
- **c** a mouse
- **d** a 40-tonne truck.

(For estimates, use g = 10 m s^{-2}; 1 tonne = 1000 kg.)

On the Moon

The Moon is smaller and has less mass than the Earth, and so its gravity is weaker. If you were to drop a stone on the Moon, it would have a smaller acceleration. Your hand is about 1 m above ground level; a stone takes about 0.45 s to fall through this distance on the Earth, but about 1.1 s on the surface of the Moon. The acceleration of free fall on the Moon is about one-sixth of that on the Earth:

$$g_{\text{Moon}} = 1.6 \text{ m s}^{-2}$$

It follows that objects weigh less on the Moon than on the Earth. They are not completely weightless, because the Moon's gravity is not zero.

Mass and weight

We have now considered two related quantities, mass and weight. It is important to distinguish carefully between these (Table 3.3).

Quantity	Symbol	Unit	In terms of base units	Comment
mass	*m*	kg	kg	this does not vary from place to place
weight	*mg*	N	kg m s^{-2}	this is a force – it depends on the strength of gravity

Table 3.3: Distinguishing between mass and weight.

Figure 3.5 shows a vehicle used to travel on the moon, named a moon-buggy. If the moon-buggy breaks down, it will be no easier to get it moving on the Moon than on the Earth. This is because its mass does not change, because it is made from the same atoms and molecules wherever it is. From $F = ma$, it follows that if m does not change, you will need the same force F to start it moving.

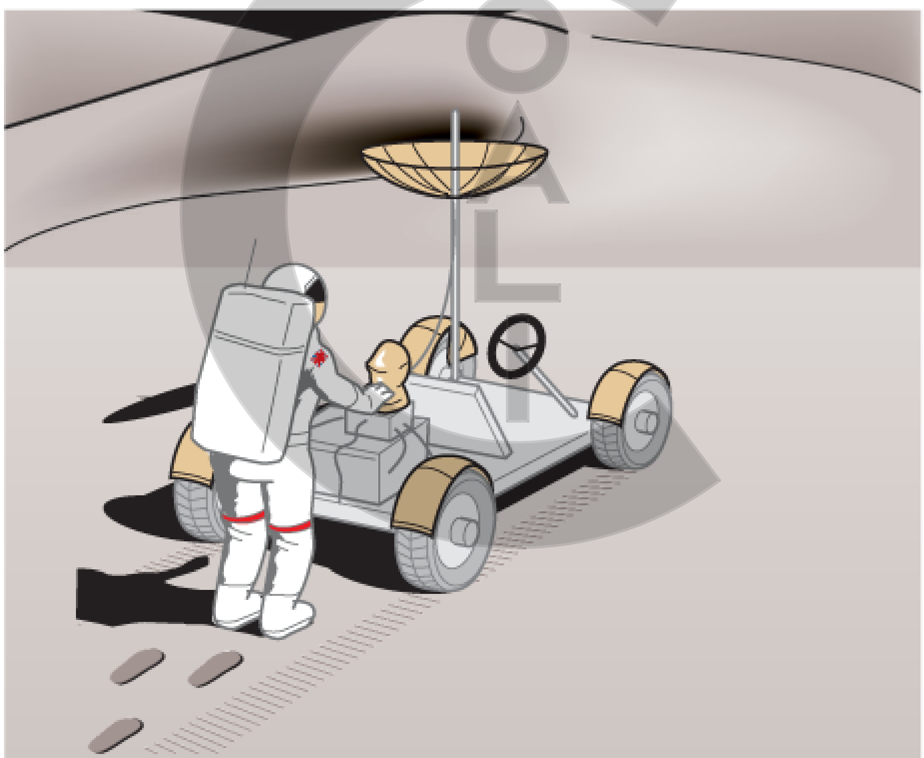


Figure 3.5: The mass of a moon-buggy is the same on the Moon as on the Earth, but its weight is smaller.

However, your moon-buggy will be easier to lift on the Moon, because its weight will be less. From $W = mg$, since g is less on the Moon, it has a smaller weight than when on the Earth.

3.4 Mass and inertia

It took a long time for scientists to develop correct ideas about forces and motion. We will start by thinking about some wrong ideas, and then consider why Galileo, Newton and others decided new ideas were needed.

Observations and ideas

Here are some observations to think about.

- The large tree trunk shown in Figure 3.6 is being pulled from a forest. The elephant provides the
- force needed to pull it along. If the elephant stops pulling, the tree trunk will stop moving.
- A horse is pulling a cart. If the horse stops pulling, the cart stops.
- You are riding a bicycle. If you stop pedaling, the bicycle will come to a halt.
- You are driving along the road. You must keep your foot on the accelerator pedal, otherwise the car will not keep moving.
- You kick a football. The ball rolls along the ground and gradually stops.

In each of these cases, there is a force that makes something move – the pull of the elephant or the horse, your push on the bicycle pedals, the force of the car engine, the push of your foot. Without the force, the moving object comes to a halt. So what conclusion might we draw?

A moving object needs a force to keep it moving.

This might seem a sensible conclusion to draw, but it is wrong. We have not thought about all the forces involved. The missing force is friction.

In each example, friction (or air resistance) makes the object slow down and stop when there is no force pushing or pulling it forwards. For example, if you stop pedaling your cycle, air resistance will slow you down. There is also friction at the axles of the wheels, and this too will slow you down. If you could lubricate your axles and cycle in a vacuum, you could travel along at a steady speed forever, without pedaling!

Figure 3.6: An elephant provides the force needed to pull this tree from the forest.

In the 17th century, astronomers began to use telescopes to observe the night sky. They saw that objects such as the planets could move freely through space. They simply kept on moving, without anything providing a force to push them. Galileo came to the conclusion that this was the natural motion of objects.

- An object at rest will stay at rest, unless a force causes it to start moving.
- A moving object will continue to move at a steady speed in a straight line, unless a force acts on it.

So objects move with a constant velocity, unless a force acts on them. (Being stationary is simply a particular case of this, where the velocity is zero.) Nowadays, it is much easier to appreciate this law of motion, because we have more experience of objects moving with little or no friction such as roller-skates with low-friction bearings, ice skates and spacecraft in empty space. In Galileo's day, people's everyday experience was of dragging things along the ground, or pulling things on carts with high-friction axles. Before Galileo, the orthodox scientific idea was that a force must act all the time to keep an object moving – this had been handed down from the time of the ancient Greek philosopher Aristotle. So it was a great achievement when scientists were able to develop a picture of a world without friction.

The idea of inertia

The tendency of a moving object to carry on moving is sometimes known as **inertia**.

- An object with a large mass is difficult to stop moving – think about catching a football, compared with a less massive tennis ball moving at the same speed.
- Similarly, a stationary object with a large mass is difficult to start moving – think about pushing a car to get it started.
- It is difficult to make a massive object change direction – think about the way a fully laden supermarket trolley tries to keep moving in a straight line.

All of these examples suggest another way to think of an object's mass; it is a measure of its inertia – how difficult it is to change the object's motion. **Uniform motion** is the natural state of motion of an object.

Here, uniform motion means 'moving with constant velocity' or 'moving at a steady speed in a straight line'.

Newton's first law of motion

The findings on inertia and uniform motion can be summarised as **Newton's first law of motion**:

In fact, this is already contained in the simple equation we have been using to calculate acceleration, $F = ma$. If no resultant force acts on an object ($F = 0$), it will not accelerate ($a = 0$). The object will either remain stationary or it will continue to travel at a constant velocity. If we rewrite the equation as $a = \frac{F}{m}$, we can see that the greater the mass m, the smaller the acceleration a produced by a force F.

Questions

5 Use the idea of inertia to explain why some large cars have power-assisted brakes.

6 A car crashes head-on into a brick wall. Use the idea of inertia to explain why the driver is more likely to come out through the windscreen if he or she is not wearing a seat belt.

Top speed

The vehicle shown in Figure 3.7 is capable of speeds as high as 760 mph, greater than the speed of sound. Its streamlined shape is designed to cut down air resistance and its jet engines provide a strong forwards force to accelerate it up to top speed.

All vehicles have a top speed. But why can't they go any faster? Why can't a car driver keep pressing on the accelerator pedal, and simply go faster and faster?

To answer this, we have to think about the two forces already mentioned: air resistance and the forwards thrust (force) of the engine. The vehicle will accelerate so long as the thrust is greater than the air resistance. When the two forces are equal, the resultant force on the vehicle is zero and the vehicle moves at a steady velocity.

Balanced and unbalanced forces

If an object has two or more forces acting on it, we have to consider whether or not they are 'balanced' (Figure 3.8). Forces on an object are balanced when the resultant force on the object is zero. The object will either remain at rest or have a constant velocity.

We can calculate the **resultant force** by adding up two (or more) forces that act in the same straight line. We must take account of the direction of each force. In the examples in Figure 3.8, forces to the right are positive and forces to the left are negative.

When a car travels slowly, it encounters little air resistance. However, the faster it goes, the more air it has to push out of the way each second and so the greater the air resistance. Eventually, the backwards force of air resistance equals the forwards force provided between the tyres and the road, and the forces on the car are balanced. It can go no faster–it has reached its top speed.

Free fall

Skydivers (Figure 3.9) are rather like cars–at first, they accelerate freely. At the start of the fall, the only force acting on the diver is his or her weight. The acceleration of the diver at the start must therefore be g. Then increasing air resistance opposes their fall and their acceleration decreases. Eventually, they reach a maximum velocity, known as the **terminal velocity**.

At the terminal velocity, the air resistance is equal to the weight. The terminal velocity is approximately 120 miles per hour (about 50 m s^{-1}), but it depends on the skydiver's weight and orientation. Head-first is fastest.

Figure 3.7: The Thrust SSC rocket car broke the world land-speed record in 1997. It achieved a top speed of 763 mph (just over 340 m s^{-1}) over a distance of 1 mile (1.6 km).

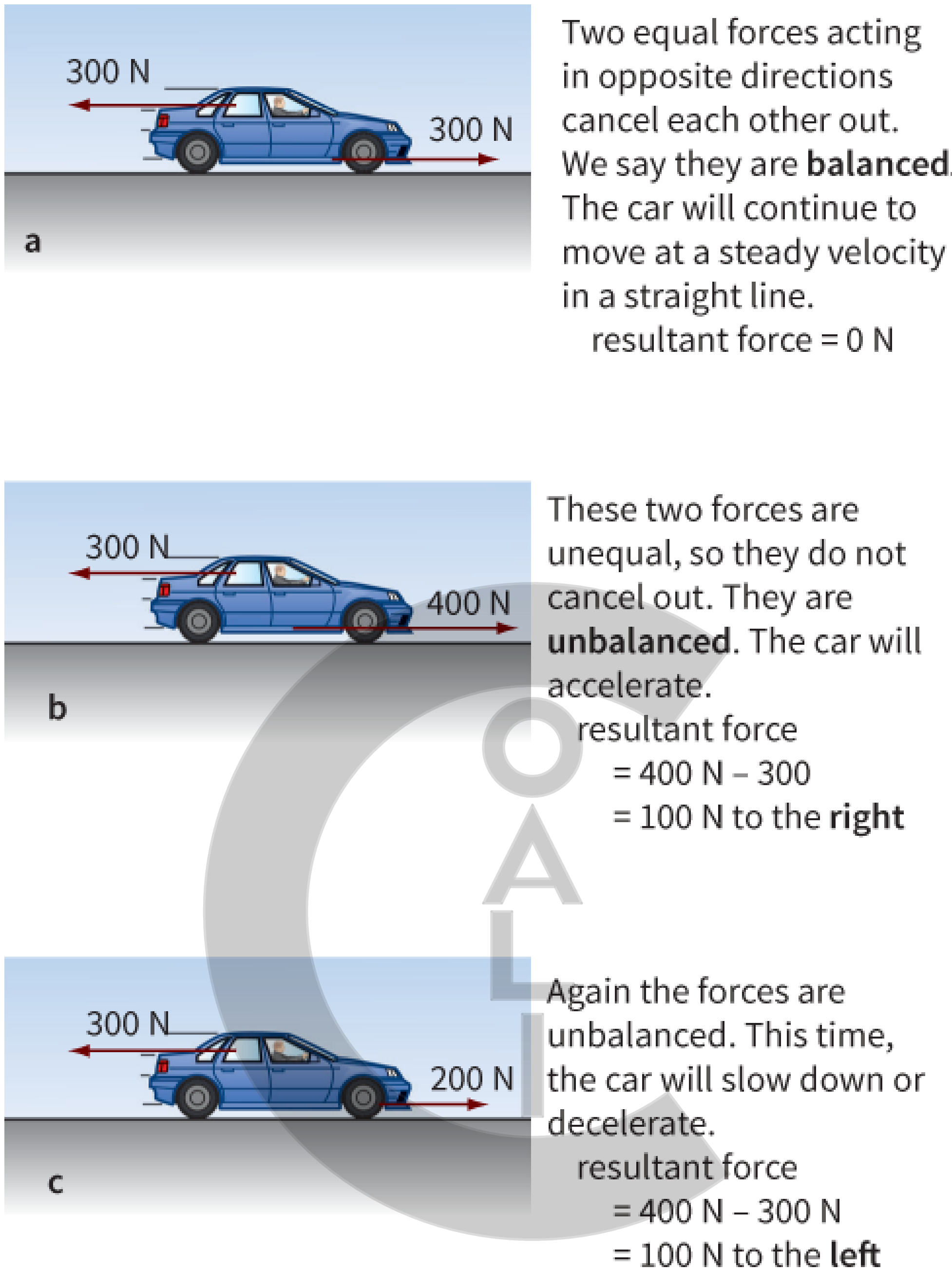


Figure 3.8: Balanced and unbalanced forces.

Figure 3.9: A skydiver falling freely.

The idea of a parachute is to greatly increase the air resistance. Then terminal velocity is reduced, and the parachutist can land safely. Figure 3.10 shows how a parachutist's velocity might change during descent.

Terminal velocity depends on the weight and surface area of the object. For insects, air resistance is much greater relative to their weight than for a human being and so their terminal velocity is quite low. Insects can be swept up several kilometres into the atmosphere by rising air streams. Later, they fall back to Earth uninjured. It is said that mice can survive a fall from a high building for the same reason.

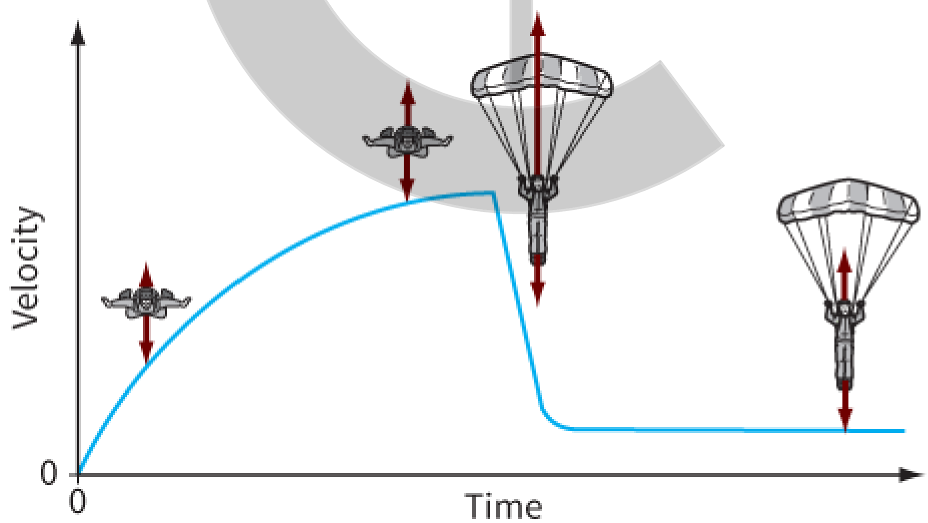


Figure 3.10: The velocity of a parachutist varies during a descent. The force arrows show weight (downwards) and air resistance (upwards).

3.5 Moving through fluids

Air resistance is just one example of the **resistive force** (or viscous force) that objects experience when they move through a fluid, a liquid or a gas. If you have ever run down the beach and into the sea, or tried to wade quickly through the water of a swimming pool, you will have experienced the force of **drag**. The deeper the water gets, the more it resists your movement and the harder you have to work to make progress through it. In deep water, it is easier to swim than to wade.

You can observe the effect of drag on a falling object if you drop a key or a coin into the deep end of a swimming pool. For the first few centimetres, it speeds up, but for the remainder of its fall, it has a steady speed. (If it fell through the same distance in air, it would accelerate all the way.) The drag of water means that the falling object reaches its terminal velocity very soon after it is released. Compare this with a skydiver, who has to fall hundreds of metres before reaching terminal velocity.

Moving through air

We rarely experience drag in air. This is because air is much less dense than water; its density is roughly that of water. At typical walking speed, we do not notice the effects of drag. However, if you want to move faster, the effects can be important. Racing cyclists, like the one shown in Figure 3.11, wear tight-fitting clothing and streamlined helmets.

Figure 3.11: A racing cyclist adopts a posture that helps to reduce drag. Clothing, helmet and even the cycle itself are designed to allow them to go as fast as possible.

Other athletes may take advantage of the drag of air. The runner in Figure 3.12 is undergoing resistance training. The parachute provides a backwards force against which his muscles must work. This should help to develop his muscles.

Figure 3.12: A runner making use of air resistance to build up his muscles.

WORKED EXAMPLES

3 A car of mass 500 kg is travelling along a flat road. The forward force provided between the car tyres and the road is 300 N and the air resistance is 200 N. Calculate the acceleration of the car.

Step 1 Start by drawing a diagram of the car, showing the forces mentioned in the question (Figure 3.13). Calculate the resultant force on the car; the force to the right is taken as positive:

resultant force = 300 − 200 = 100 N

Step 2 Now use $F = ma$ to calculate the car's acceleration:

$$a = \frac{F}{m} = \frac{100}{500} = 0.20\ \text{m s}^{-2}$$

So the car's acceleration is $0.20\ \text{m s}^{-2}$.

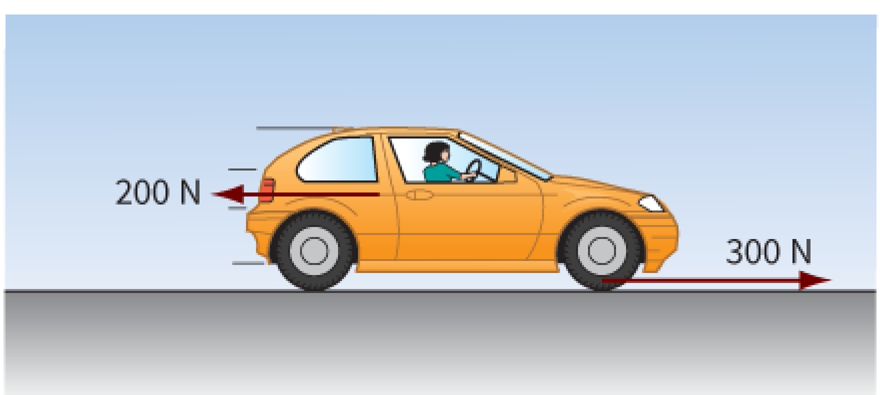


Figure 3.13: The forces on an accelerating car.

4 The maximum forward force a car can provide is 500 N. The air resistance F that the car experiences depends on its speed according to $F = 0.2v^2$, where v is the speed in m s^{-1}. Determine the top speed of the car.

Step 1 From the equation $F = 0.2v^2$, you can see that the air resistance increases as the car goes faster. Top speed is reached when the forward force equals the air resistance. So, at top speed:

$500 = 0.2v^2$

Step 2 Rearranging gives:

$$v^2 = \frac{500}{0.2}$$
$$= 2500$$
$$v = \sqrt{2500}$$
$$= 50 \text{ m s}^{-1}$$

So the car's top speed is 50 m s^{-1} (this is about 180 km h^{-1}).

Questions

7 If you drop a large stone and a small stone from the top of a tall building, which one will reach the ground first? Explain your answer.

8 In a race, downhill skiers want to travel as quickly as possible. They are always looking for ways to increase their top speed. Explain how they might do this. Think about:

a their skis

b their clothing

c their muscles

d the slope.

9 Skydivers jump from a plane at intervals of a few seconds. If two divers wish to join up as they fall, the second must catch up with the first.

a If one diver is more massive than the other, who should jump first? Use the idea of forces and terminal velocity to explain your answer.

b If both divers are equally massive, suggest what the second might do to catch up with the first.

Contact forces and upthrust

We will now think about the forces that act when two objects are in contact with each other. When two objects touch each other, each exerts a force on the other. These are called **contact forces**. For example, when you stand on the floor (Figure 3.14), your feet push downwards on the floor and the floor pushes back upwards on your feet. This is a vital force – the upward push of the floor prevents you from falling downwards under the pull of your weight.

Where do these contact forces come from? When you stand on the floor, the floor becomes slightly compressed. Its atoms are pushed slightly closer together, and the interatomic forces push back against the compressing force. At the same time, the atoms in your feet are also pushed together so that they push back in the opposite direction. (It is hard to see the compression of the floor when you stand on it, but if you stand on a soft material such as foam rubber or a mattress you will be able to see the compression clearly.)

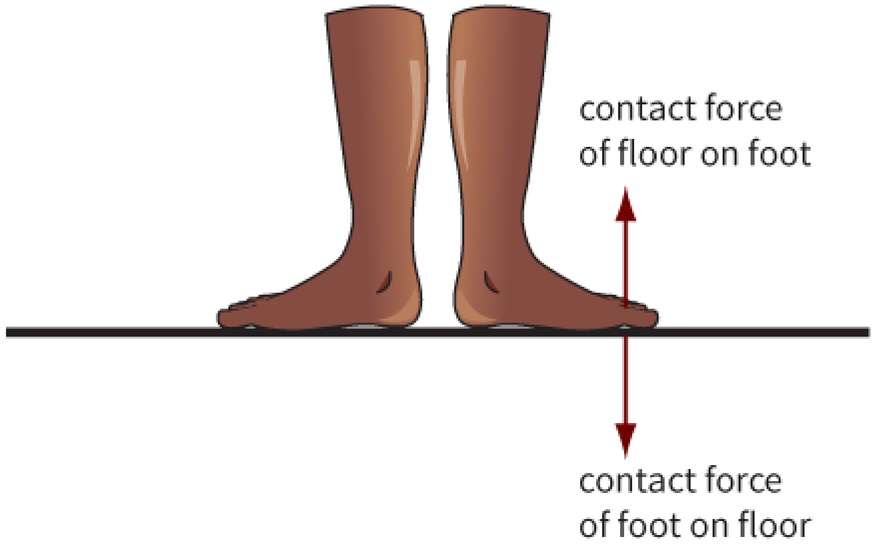


Figure 3.14: Equal and opposite contact forces act when you stand on the floor.

You can see from Figure 3.14 that the two contact forces act in opposite directions. They are also equal in magnitude. As we will see shortly, this is a consequence of Newton's third law of motion.

When an object is immersed in a fluid (a liquid or a gas), it experiences an upward force called **upthrust**. It is the upthrust of water that keeps a boat floating (Figure 3.15), and the upthrust of air that lifts a hot air balloon upwards.

The upthrust of water on a boat can be thought of as the contact force of the water on the boat. It is caused by the pressure of the water pushing upwards on the boat. Pressure arises from the motion of the water molecules colliding with the boat and the net effect of all these collisions is an upwards force.

An object in air, such as a ball, has a very small upthrust acting on it, because the density of the air around it is low. Molecules hit the top surface of the ball pushing down, but only a few more molecules push upwards on the bottom of the ball, so the resultant force upwards, or the upthrust, is low. If the ball is falling, air resistance is greater than this small upthrust but both these forces are acting upwards on the ball.

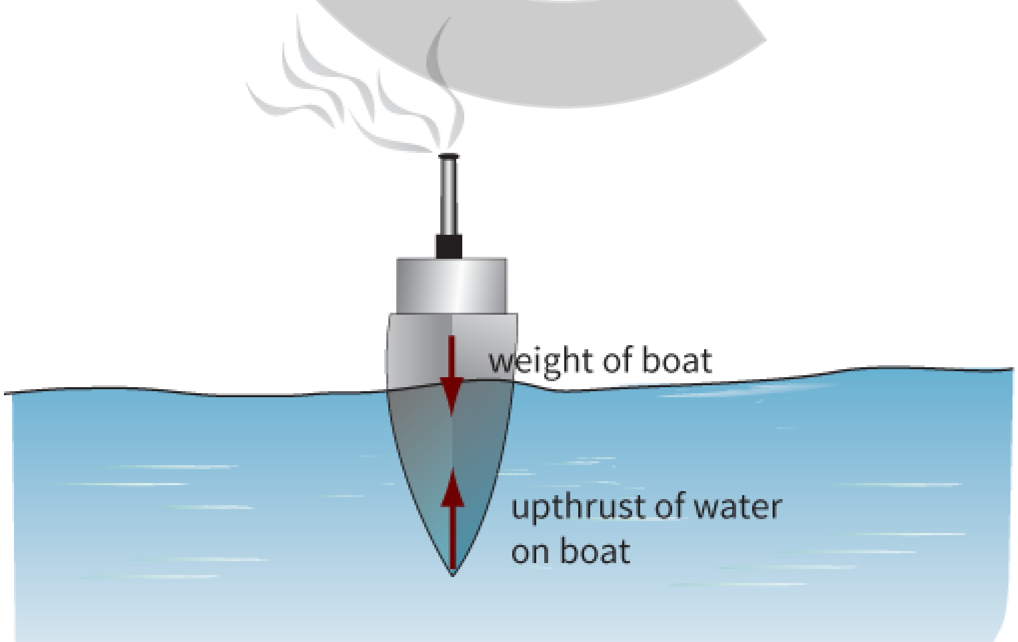


Figure 3.15: Without sufficient upthrust from the water, the boat would sink.

Questions

10 Name these forces:

- **a** the upward push of water on a submerged object
- **b** the force that wears away two surfaces as they move over one another
- **c** the force that pulled the apple off Isaac Newton's tree
- **d** the force that stops you falling through the floor
- **e** the force in a string that is holding up an apple
- **f** the force that makes it difficult to run through shallow water.

11 Draw a diagram to show the forces that act on a car as it travels along a level road at its top speed.

12 Imagine throwing a shuttlecock straight up in the air. Air resistance is more important for shuttlecocks than for a tennis ball. Air resistance always acts in the opposite direction to the velocity of an object.

Draw diagrams to show the two forces, weight and air resistance, acting on the shuttlecock:

- **a** as it moves upwards
- **b** as it falls back downwards.

3.6 Newton's third law of motion

For completeness, we should now consider **Newton's third law of motion**. (There is more about this in Chapter 6.)

When two objects interact, each exerts a force on the other. Newton's third law says that these forces are equal and opposite to each other:

When two bodies interact, the forces they exert on each other are equal in magnitude and opposite in direction.

(These two forces are sometimes described as action and reaction, but this is misleading as it sounds as though one force arises as a consequence of the other.
In fact, the two forces appear at the same time and we can't say that one caused the other.)

The two forces that make up a 'Newton's third law pair' have the following characteristics:

- They act on **different** objects.
- They are equal in magnitude.
- They are opposite in direction.
- They are forces **of the same type**.

What does it mean to say that the forces are 'of the same type'? We need to think about the type of interaction which causes the forces to appear.

- Two objects may attract each other because of the gravity of their masses – these are gravitational forces.
- Two objects may attract or repel because of their electrical charges – electrical forces.
- Two objects may touch – contact forces.
- Two objects may be attached by a string and pull on each other – tension forces.
- Two objects may attract or repel because of their magnetic fields – magnetic forces.

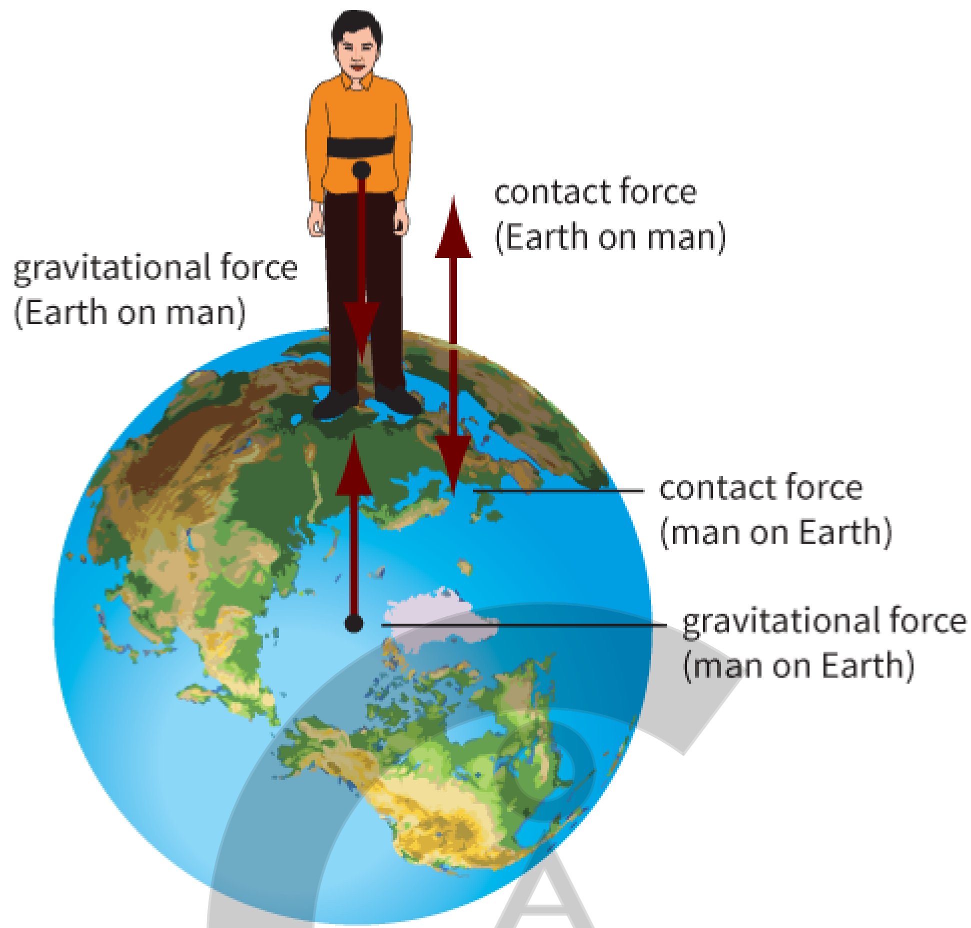


Figure 3.16: For each of the forces that the Earth exerts on you, an equal and opposite force acts on the Earth.

Figure 3.16 shows a person standing on the Earth's surface. The two gravitational forces are a Newton's third law pair, as are the two contact forces. Don't be misled into thinking that the person's weight and the contact force of the floor are a Newton's third law pair. Although they are 'equal and opposite', they do not act on different objects and they are not of the same type.

Question

13 Describe one 'Newton's third law pair' of forces involved in the following situations. In each case, state the object that each force acts on, the type of force and the direction of the force.

- **a** You step on someone's toe.
- **b** A car hits a brick wall and comes to rest.
- **c** A car slows down by applying the brakes.
- **d** You throw a ball upwards into the air.

3.7 Understanding SI units

Throughout physics, we calculate, measure and use many quantities. All quantities consist of a value and a unit. In physics, we mostly use units from the SI system. These units are all defined with extreme care, and for a good reason. In science and engineering, every measurement must be made on the same basis, so that measurements obtained in different laboratories can be compared. This is important for commercial reasons, too. Suppose an engineering firm in Taiwan is asked to produce a small part for the engine of a car that is to be assembled in India. The dimensions are given in millimetres and the part must be made with an accuracy of a tiny fraction of a millimetre. All concerned must know that the part will fit correctly – it would not be acceptable to use a different millimetre scale in Taiwan and India.

KEY IDEA

All physical quantities have a numerical magnitude (a numerical size) and a unit

Base units, derived units

The metre, kilogram and second are three of the seven SI **base units**. These are defined with great precision so that every standards laboratory can reproduce them correctly.

Other units, such as units of speed (m s^{-1}) and acceleration (m s^{-2}) are known as **derived units** because they are combinations of base units. Some derived units, such as the newton and the joule, have special names that are more convenient to use than giving them in terms of base units. The definition of the newton will show you how this works.

Defining the newton

Isaac Newton (1642–1727) played a significant part in developing the scientific idea of force. Building on Galileo’s earlier thinking, he explained the relationship between force, mass and acceleration, which we now write as $F = ma$. For this reason, the SI unit of force is named after him.

We can use the equation $F = ma$ to define the **newton** (N).

One newton is the force that will give a 1 kg mass an acceleration of 1 m s^{-2} in the direction of the force.

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

The seven base units

In mechanics (the study of forces and motion), the units we use are based on three base units: the metre, kilogram and second. As we move into studying electricity, we will need to add another base unit, the ampere. Heat requires another base unit, the kelvin (the unit of temperature).

Table 3.4 shows the seven base units of the SI system. Remember that all other units can be derived from these seven. The equations that relate them are the equations that you will learn as you go along (just as $F = ma$ relates the newton to the kilogram, metre and second). The unit of luminous intensity is not part of the AS & A Level courses.

Base unit	Symbol	Base unit
length	*x*, *l*, *s* and so on	m (metre)
mass	*m*	kg (kilogram)

Base unit	Symbol	Base unit
time	*t*	s (second)
electric current	*I*	A (ampere)
thermodynamic temperature	*T*	K (kelvin)
amount of substance	*n*	mol (mole)
luminous intensity	*I*	cd (candela)

Table 3.4: SI base quantities and units. In this course, you will learn about all of these except the candela.

KEY IDEA

Length, mass, time, current and temperature are base units in mechanics.

Question

14 The pull of the Earth's gravity on an apple (its weight) is about 1 newton. We could devise a new international system of units by defining our unit of force as the weight of an apple. State as many reasons as you can why this would not be a very useful definition.

Other SI units

Using only seven base units means that only this number of quantities have to be defined with great precision. It would be confusing if more units were also defined. For example, if the density of water were defined as exactly 1 g cm^{-3}, then 1000 cm^3 of a sample of water would have a mass of exactly 1 kg. However, it is unlikely that the mass of this volume of water would equal exactly the mass of the standard kilogram.

All other units can be derived from the base units. This is done using the definition of the quantity. For example, speed is defined as $\frac{\text{distance}}{\text{time}}$ and so the base units of speed in the SI system are **m s^{-1}**.

Since the defining equation for force is $F = ma$, the base units for force are **kg m s^{-2}**.

Equations that relate different quantities must have the same base units on each side of the equation. If this does not happen the equation must be wrong.

When each term in an equation has the same base units the equation is said to be **homogeneous**.

KEY IDEA

Base units on each side of a physics equation are the same.

WORKED EXAMPLE

5 It is suggested that the time *T* for one oscillation of a swinging pendulum is given by the equation $T^2 = 4\pi^2\left(\frac{l}{g}\right)$ where *l* is the length of the pendulum and *g* is the acceleration due to gravity. Show that this equation is homogeneous.

For the equation to be homogeneous, the term on the left-hand side must have the same base units as all the terms on the right-hand side.

Step 1 The base unit of time *T* is s. The base unit of the left-hand side of the equation is therefore s^2.

Step 2 The base unit of l is m. The base units of g are m s^{-2}. Therefore, the base unit of the right-hand side is $\frac{m}{(ms^{-2})} = s^2$

(Notice that the constant $4\pi^2$ has no units.)

Since the base units on the left-hand side of the equation are the same as those on the right, the equation is homogeneous.

Questions

15 Determine the base units of:

- **a** pressure $(= \frac{\text{force}}{\text{area}})$
- **b** energy (= force × distance)
- **c** density $(= \frac{\text{mass}}{\text{volume}})$

16 Use base units to prove that the following equations are homogeneous.

- **a** pressure = density × acceleration due to gravity × depth
- **b** distance travelled = initial speed × time + $\frac{1}{2}$acceleration × time^2 $(s = ut + \frac{1}{2}at^2)$

Prefixes

Each unit in the SI system can have **multiples** and **sub-multiples** to avoid using very high or low numbers. For example, 1 millimetre (mm) is one thousandth of a metre and 1 micrometre (μm) is one millionth of a metre.

The **prefix** comes before the unit. In the unit mm, the first m is the prefix milli and the second m is the unit metre. You will need to recognise a number of prefixes for the AS & A Level courses, as shown in Table 3.5.

You must take care when using prefixes.

Multiples			Sub-multiples		
Multiple	**Prefix**	**Symbol**	**Multiple**	**Prefix**	**Symbol**
10^3	kilo	k	10^{-1}	deci	d
10^6	mega	M	10^{-2}	centi	c
10^9	giga	G	10^{-3}	milli	m
10^{12}	tera	T	10^{-6}	micro	μ
			10^{-9}	nano	n
			10^{-12}	pico	p

Table 3.5: Multiples and sub-multiples.

Squaring or cubing prefixes

For example:

$$\begin{aligned} 1\ \text{cm} &= 10^{-2}\ \text{m} \\ \text{so } 1\ \text{cm}^2 &= (10^{-2}\ \text{m})^2 = 10^{-4}\ \text{m}^2 \\ \text{and } 1\ \text{cm}^3 &= (10^{-2}\ \text{m})^3 = 10^{-6}\ \text{m}^3 \end{aligned}$$

Writing units

You must leave a small space between each unit when writing a speed such as 3 m s^{-1}, because if you write it as 3 ms^{-1} it would mean 3 millisecond^{-1}.

WORKED EXAMPLE

6 The density of water is 1.0 g cm^{-3}. Calculate this value in kg m^{-3}.

Step 1 Find the conversions for the units:

$1\ \text{g} = 1 \times 10^{-3}\ \text{kg}$

$1\ \text{cm}^3 = 1 \times 10^{-6}\ \text{m}^3$

Step 2 Use these in the value for the density of water:

$$\begin{aligned} 1.0\ \text{gcm}^{-3} &= \frac{1.0 \times 1 \times 10^{-3}}{1 \times 10^{-6}} \\ &= 1.0 \times 10^{3}\ \text{kgm}^{-3} \end{aligned}$$

Questions

17 **a** Find the area of one page of this book in cm^2 and then convert your value to m^2.

b If the uncertainty in measuring one side of the page is 0.1 cm find the uncertainty in the area. This can be done by either taking the largest value of each side when you multiply them together and then finding the difference from you value in part **a** or using a combination of the percentage uncertainties (see Chapter P1). Try both methods.

18 Write down, in powers of ten, the values of these quantities:

a 60 pA

b 500 MW

c 20 000 mm.

REFLECTION

Did you find it difficult to understand that Newton's third law of motion relates forces that act on **different** bodies?

SUMMARY

An object will remain at rest or in a state of uniform motion unless it is acted on by an external force. This is Newton's first law of motion.
For a body of constant mass, the acceleration is directly proportional to the resultant force applied to it. Resultant force F, mass m and acceleration a are related by the equation: resultant force = mass × acceleration ($F = ma$)   This is a form of Newton's second law of motion.
When two bodies interact, the forces they exert on each other are equal in magnitude and opposite in direction. This is Newton's third law of motion.
The acceleration produced by a force is in the same direction as the force. Where there are two or more forces, we must determine the resultant force.
A newton (N) is the force required to give a mass of 1 kg an acceleration of 1 m s^{-2} in the direction of the force.
The greater the mass of an object, the more it resists changes in its motion. Mass is a measure of the object's inertia.
The weight of an object is a result of the pull of gravity on it: weight = mass × acceleration of free fall ($W = mg$)
Terminal velocity is reached when the fluid resistance is equal to the weight of the object.
Physics equations are homogenous and have the same base units on each side. The main base units are m, kg, s, A and K (the thermodynamic unit for temperature).

EXAM-STYLE QUESTIONS

1 Which list contains only SI base units? **[1]**

A ampere, kelvin, gram

B kilogram, metre, newton

C newton, second, ampere

D second, kelvin, kilogram

2 The speed *v* of a wave travelling a wire is given by the equation

$$v = \left(\frac{Tl}{m}\right)^n$$

where *T* is the tension in the wire that has mass *m* and length *l*.

In order for the equation to be homogenous, what is the value of *n*? **[1]**

A $\frac{1}{2}$

B 1

C 2

D 4

3 When a golfer hits a ball his club is in contact with the ball for about 0.000 50 s and the ball leaves the club with a speed of 70 m s^{-1}. The mass of the ball is 46 g.

a Determine the mean accelerating force. **[4]**

b What mass, resting on the ball, would exert the same force as in part **a**? **[2]**

[Total: 6]

4 The mass of a spacecraft is 70 kg. As the spacecraft takes off from the Moon, the upwards force on the spacecraft caused by the engines is 500 N. The acceleration of free fall on the Moon is 1.6 N kg^{-1}.

Determine:

a the weight of the spacecraft on the Moon **[2]**

b the resultant force on the spacecraft **[2]**

c the acceleration of the spacecraft. **[2]**

[Total: 6]

5 A metal ball is dropped into a tall cylinder of oil. The ball initially accelerates but soon reaches a terminal velocity.

a By considering the forces on the metal ball bearing, explain why it first accelerates but then reaches terminal velocity. **[3]**

b State how you would show that the metal ball reaches terminal velocity. Suggest one cause of random errors in your readings. **[4]**

[Total: 7]

6 Determine the speed in m s^{-1} of an object that travels:

a 3.0 μm in 5.0 ms **[2]**

b 6.0 km in 3.0 Ms **[2]**

c 8.0 pm in 4.0 ns. **[2]**

[Total: 6]

7 This diagram shows a man who is just supporting the weight of a box. Two of the

forces acting are shown in the diagram. According to Newton's third law, each of these forces is paired with **another** force.

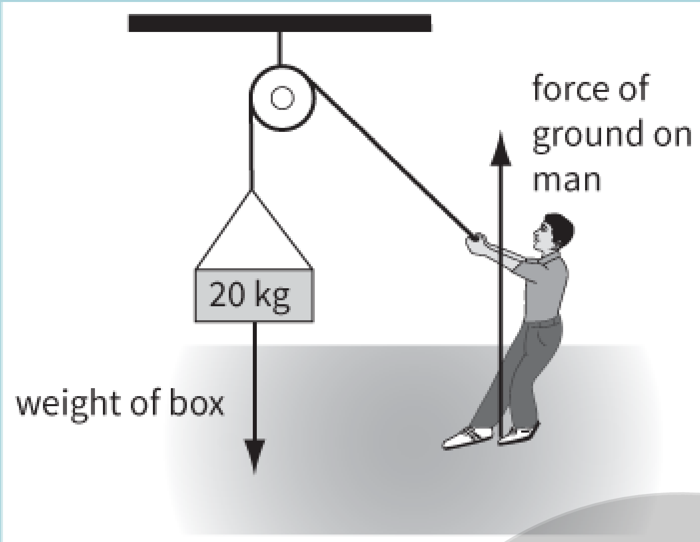


Figure 3.17

For **a** the weight of the box, and **b** the force of the ground on the man, state:

i the body that the other force acts upon **[2]**

ii the direction of the other force **[2]**

iii the type of force involved. **[2]**

[Total: 6]

8 A car starts to move along a straight, level road. For the first 10 s, the driver maintains a constant acceleration of 1.5 m s^{-2}. The mass of the car is 1.1×10^3 kg.

a Calculate the driving force provided by the wheels, when:

i the force opposing motion is negligible **[1]**

ii the total force opposing the motion of the car is 600 N. **[1]**

b Calculate the distance travelled by the car in the first 10 s. **[2]**

[Total: 4]

9 These are the speed–time graphs for two falling balls:

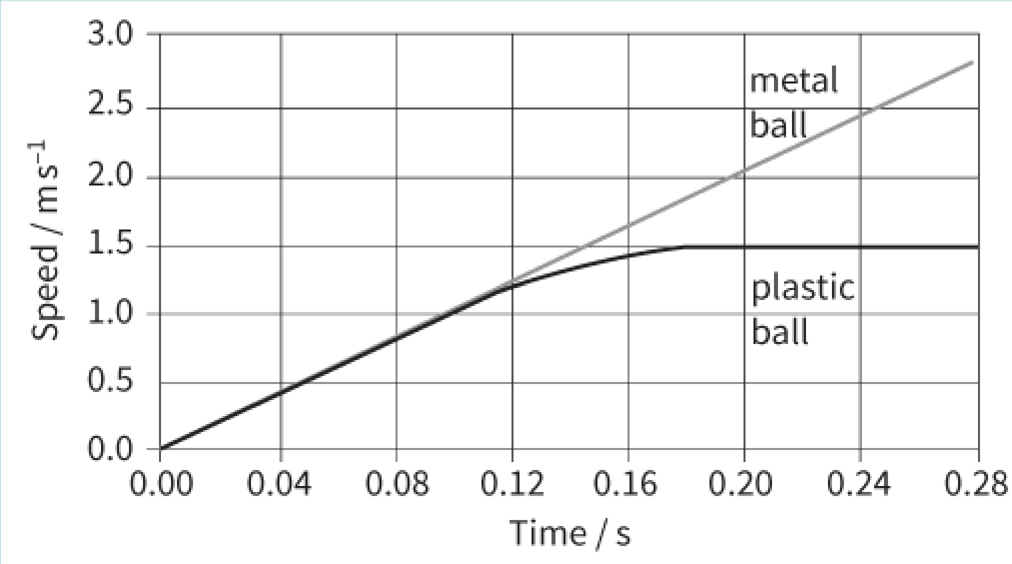


Figure 3.18

a Determine the terminal velocity of the plastic ball. **[1]**

b Both balls are of the same size and shape but the metal ball has a greater mass.

Explain, in terms of Newton's laws of motion and the forces involved, why the plastic ball reaches a constant velocity but the metal ball does not. **[3]**

c Explain why both balls have the same initial acceleration. **[2]**

[Total: 6]

10 A car of mass 1200 kg accelerates from rest to a speed of 8.0 m s^{-1} in a time of 2.0 s.

a Calculate the forward driving force acting on the car while it is accelerating. Assume that, at low speeds, all frictional forces are negligible. **[2]**

b At high speeds the resistive frictional force F produced by air on a body moving with velocity v is given by the equation $F = bv^2$, where b is a constant.

i Derive the base units of force in the SI system. **[1]**

ii Determine the base units of b in the SI system. **[1]**

iii The car continues with the same forward driving force and accelerates until it reaches a top speed of 50 m s^{-1}. At this speed the resistive force is given by the equation $F = bv^2$. Determine the value of b for the car. **[2]**

iv Use your value for b in **iii** and the driving force calculated in part **a** to calculate the acceleration of the car when the speed is 30 m s^{-1}. **[2]**

v **Sketch** a graph showing how the value of F varies with v over the range 0 to 50 m s^{-1}. Use your graph to describe what happens to the acceleration of the car during this time. **[2]**

[Total: 10]

11 a Explain what is meant by the mass of a body and the weight of a body. **[3]**

b State and explain one situation in which the weight of a body changes while its mass remains constant. **[2]**

c State the difference between the base units of mass and weight in the SI system. **[2]**

[Total: 7]

12 a State Newton's second law of motion in terms of acceleration. **[2]**

b When you jump from a wall on to the ground, it is advisable to bend your knees on landing.

i State how bending your knees affects the time it takes to stop when hitting the ground. **[1]**

ii Using Newton's second law of motion, explain why it is sensible to bend your knees. **[2]**

[Total: 5]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
recognise that mass is a property of an object that resists change in motion	3.1			
identify the forces acting on a body in different situations	3.2, 3.3			
recall $F = ma$ and solve problems using it	3.3			
state and apply Newton's first and third laws of motion	3.4, 3.6			
recall that the weight of a body is equal to the product of its mass and the acceleration of free fall	3.3			
relate derived units to base units in the SI system and use base units to check the homogeneity of an equation	3.7			
recall and use a range of prefixes.	3.7			


USA
CALI

〉Chapter 4

Forces: vectors and moments

LEARNING INTENTIONS

In this chapter you will learn how to:

- use a vector triangle to represent coplanar forces in equilibrium and add two or more coplanar forces
- resolve a force into perpendicular components
- represent the weight of a body as acting at a single point known as its centre of gravity
- define and apply the moment of a force and the torque of a couple
- state and apply the principle of moments
- use the idea that, when there is no resultant force and no resultant torque, a system is in equilibrium.

BEFORE YOU START

- Write down what a *vector* is. List some examples.
- Is force is a vector? Discuss with a partner.

SAILING AHEAD

Force is a vector quantity. Sailors know a lot about the vector nature of forces. For example, they can sail 'into the wind'. The sails of a yacht can be angled to provide a 'component' of force (in other words, an effect of the force in the forward direction) and the boat can then sail at almost 45° to the wind. The boat tends to 'heel over' and the crew sit on the side of the boat to provide a turning effect in the opposite direction (Figure 4.1). If the wind has an effect forwards, what stops the boat from moving sideways due to the 'component' of the wind sideways? (Hint: find out about the shape of the bottom of the boat.)

Figure 4.1: Sailing into the wind.

4.1 Combining forces

You will have learned that a vector quantity has both magnitude and direction. An object may have two or more forces acting on it and, since these are vectors, we must use vector addition (Chapter 1) to find their combined effect (their resultant).

There are several forces acting on the car (Figure 4.2) as it struggles up the steep hill. They are:

- its weight W (= mg)
- the normal contact force N of the road
- air resistance D
- the forward force F caused by friction between the car tyres and the road.

If we knew the magnitude and direction of each of these **forces**, we could work out their combined effect on the car. Will it accelerate up the hill? Or will it slide backwards down the hill?

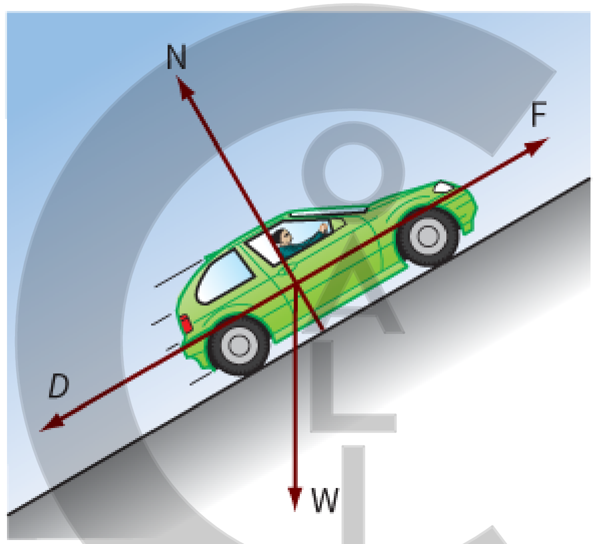


Figure 4.2: Four forces act on this car as it moves uphill.

The combined effect of several forces is known as the **resultant force**. To see how to work out the resultant of two or more forces, we will start with a relatively simple example.

Two forces in a straight line

We saw some examples in Chapter 3 of two forces acting in a straight line. For example, a falling tennis ball may be acted on by two forces: its weight mg, downwards, and air resistance D, upwards (Figure 4.3). The resultant force is then:

resultant force = $mg - D = 1.0 - 0.2 = 0.8$ N

When adding two or more forces that act in a straight line, we have to take account of their directions. A force may be positive or negative; we adopt a **sign convention** to help us decide which is which. In setting up the sign convention you decide for yourself which direction is positive. In Figure 4.3, for example, we have taken the direction downwards as positive so the weight is +1.0 N, a positive force, and the force upwards is −0.2 N, a negative force. The resultant is +0.8 N, which tells us the resultant is downwards.

You might choose the upwards direction as positive, but if you apply a sign convention correctly, the sign of your final answer will tell you the direction of the resultant force (and hence acceleration).

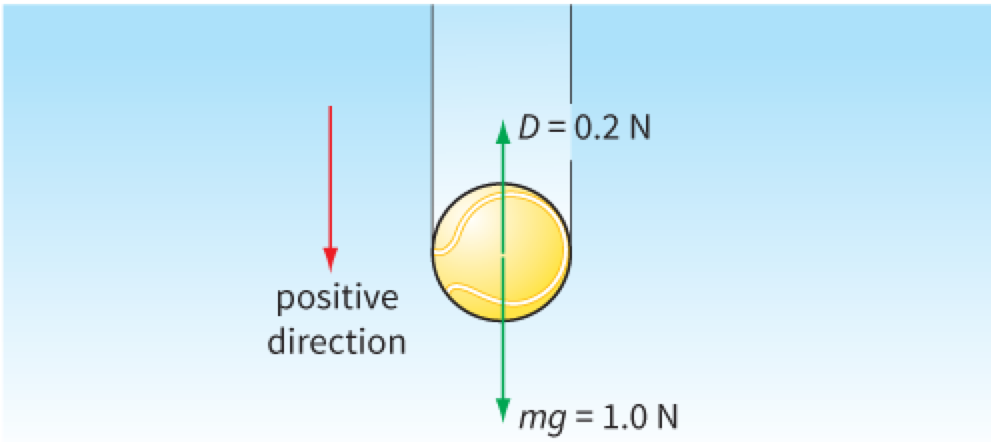


Figure 4.3: Two forces on a falling tennis ball.

Two forces at right angles

Figure 4.4 shows a shuttlecock falling on a windy day. There are two forces acting on the shuttlecock: its weight vertically downwards, and the horizontal push of the wind. (It helps if you draw the force arrows of different lengths, to show which force is greater.) We must add these two forces together to find the resultant force acting on the shuttlecock.

We add the forces by drawing two arrows, head-to-tail, as shown on the right of Figure 4.4.

- First, draw a horizontal arrow to represent the 6.0 N push of the wind.

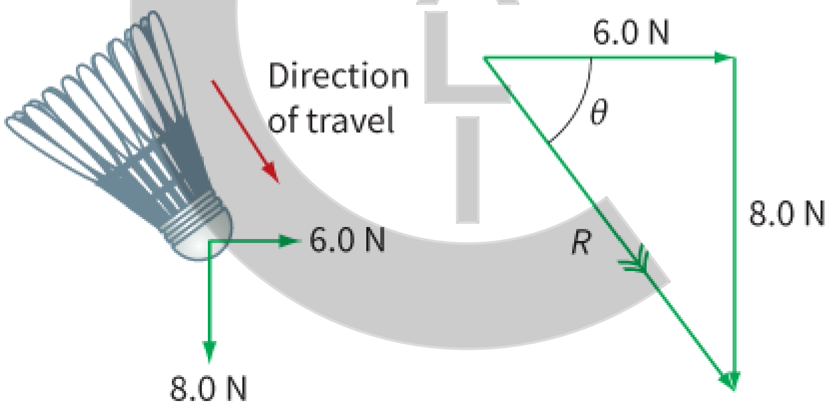


Figure 4.4: Two forces act on this shuttlecock as it travels through the air; the vector triangle shows how to find the resultant force.

- Next, starting from the end of this arrow, draw a second arrow, downwards, representing the weight of 8.0 N.
- Now, draw a line from the start of the first arrow to the end of the second arrow. This arrow represents the resultant force *R*, in both magnitude and direction.

The arrows are added by drawing them end-to-end; the end of the first arrow is the start of the second arrow. Now we can find the resultant force either by scale drawing or by calculation. In this case, we have a 3–4–5 right-angled triangle, so calculation is simple:

$$R^2 = 6.0^2 + 8.0^2 = 36 + 64$$
$$= 100$$
$$R = \sqrt{100}$$
$$= 10 \text{ N}$$
$$\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{8.0}{6.0} = \frac{4}{3}$$
$$\theta = \tan^{-1}\frac{4}{3} \approx 53°$$

So the resultant force is 10 N, at an angle of 53° below the horizontal. This is a reasonable answer; the weight is pulling the shuttlecock downwards and the wind is pushing it to the right. The angle is greater than 45° because the downward force is greater than the horizontal force.

KEY IDEA

When you draw a scale drawing you should:

- state the scale used
- draw a large diagram to reduce the uncertainty.

Three or more forces

The spider shown in Figure 4.5 is hanging by a thread. It is blown sideways by the wind. The diagram shows the three forces acting on it:

- weight acting downwards
- the tension in the thread
- the push of the wind.

The diagram also shows how these can be added together. In this case, we arrive at an interesting result. Arrows are drawn to represent each of the three forces, end-to-end. The end of the third arrow coincides with the start of the first arrow, so the three arrows form a closed triangle. This tells us that the resultant force R on the spider is zero, that is, $R = 0$. The closed triangle in Figure 4.5 is known as a **triangle of forces**.

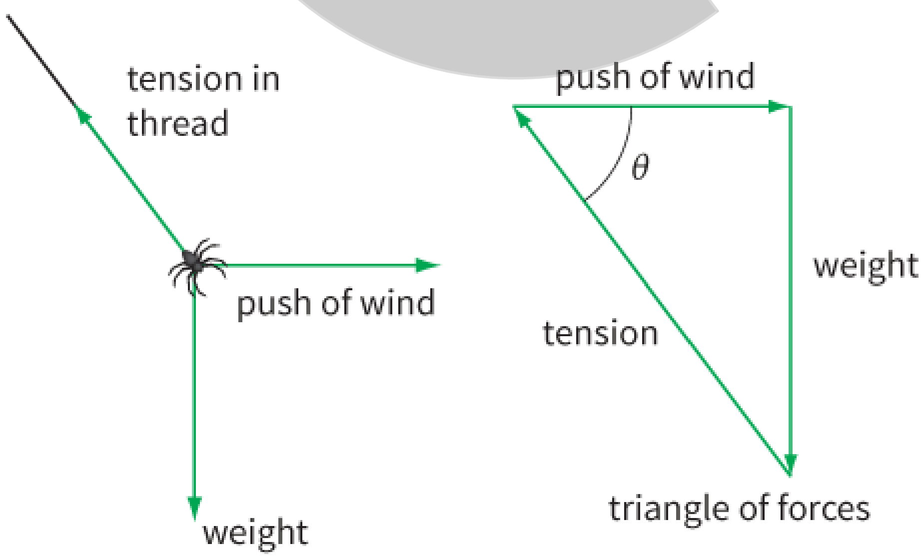


Figure 4.5: Blowing in the wind–this spider is hanging in equilibrium.

So there is no resultant force. The forces on the spider balance each other out, and we say that the spider is in **equilibrium**. If the wind blew a little harder, there would be an unbalanced force on the spider, and it would move off to the right.

We can use this idea in two ways:

- If we work out the resultant force on an object and find that it is zero, this tells us that the object is in equilibrium.
- If we know that an object is in equilibrium, we know that the forces on it must add up to zero. We can use this to work out the values of one or more unknown forces.

Questions

1 A parachutist weighs 1000 N. When she opens her parachute, it pulls upwards on her with a force of 2000 N.

  a Draw a diagram to show the forces acting on the parachutist.

  b Calculate the resultant force acting on her.

  c What effect will this force have on her?

2 The ship shown in Figure 4.6 is travelling at a constant velocity.

  a Is the ship in equilibrium (in other words, is the resultant force on the ship equal to zero)? How do you know?

  b What is the upthrust *U* of the water?

  c What is the drag *D* of the water?

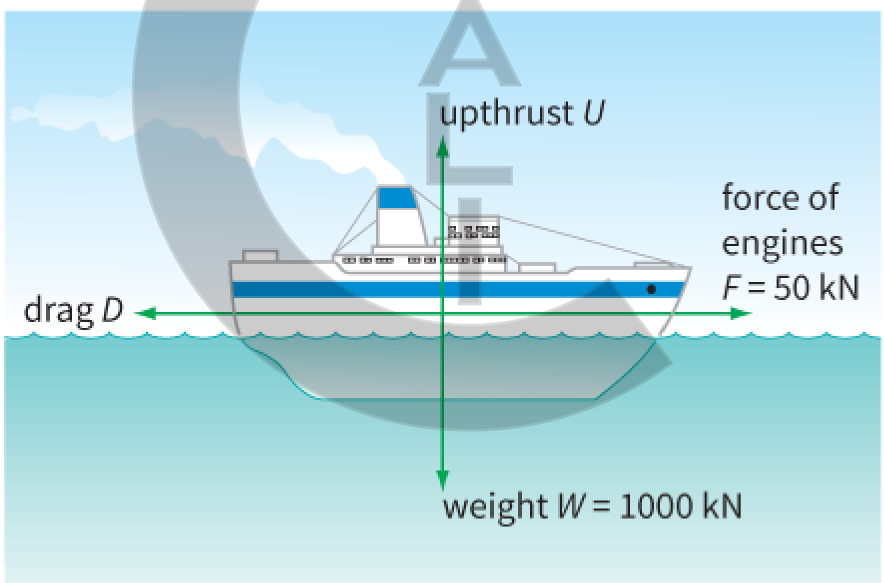


Figure 4.6: For Question 2. The force *D* is the frictional drag of the water on the boat. Like air resistance, drag is always in the opposite direction to the object's motion.

3 A stone is dropped into a fast-flowing stream. It does not fall vertically because of the sideways push of the water (Figure 4.7).

  a Calculate the resultant force on the stone.

  b Is the stone in equilibrium?

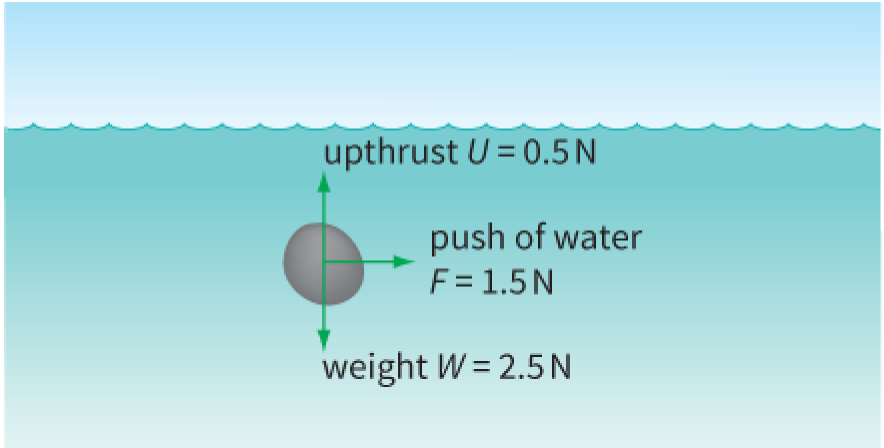


Figure 4.7: For Question 3.

4.2 Components of vectors

Look back to Figure 4.5. The spider is in equilibrium, even though three forces are acting on it. We can think of the tension in the thread as having two effects. It is:

- pulling upwards, to counteract the downward effect of gravity
- pulling to the left, to counteract the effect of the wind.

We can say that this force has two effects or **components**: an upwards (vertical) component and a sideways (horizontal) component. It is often useful to split up a vector quantity into components like this, just as we did with velocity in Chapter 2. The components are in two directions at right angles to each other, often horizontal and vertical. The process is called **resolving** the vector.

Then we can think about the effects of each component separately; we say that the perpendicular components are independent of one another. Because the two components are at 90° to each other, a change in one will have no effect on the other. Figure 4.8 shows how to resolve a force F into its horizontal and vertical components. These are:

$$\text{horizontal component of } F, F_x = F \cos \theta$$

$$\text{vertical component of } F, F_y = F \sin \theta$$

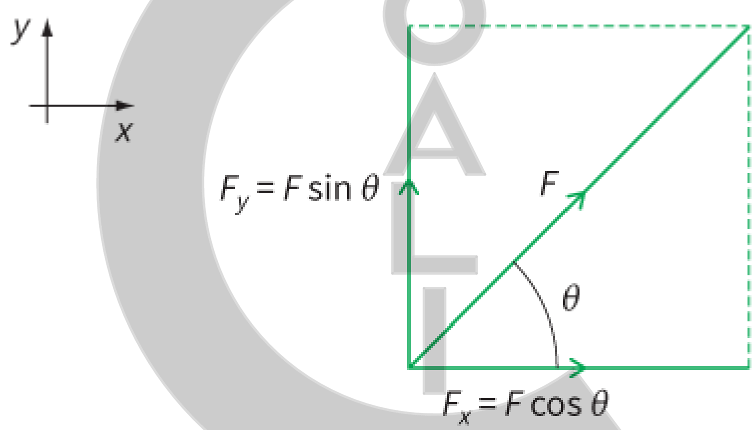


Figure 4.8: Resolving a vector into two components at right angles.

Making use of components

When the trolley shown in Figure 4.9 is released, it accelerates down the ramp. This happens because of the weight of the trolley. The weight acts vertically downwards, although this by itself does not determine the resulting motion. However, the weight has a component that acts down the slope. By calculating the component of the trolley's weight down the slope, we can determine its acceleration.

Figure 4.9: This student is investigating the acceleration of a trolley down a sloping ramp.

Figure 4.10 shows the forces acting on the trolley. To simplify the situation, we will assume there is no friction. The forces are:

- the weight of the trolley, W, which acts vertically downwards
- the contact force of the ramp, N, which acts at right angles to the ramp.

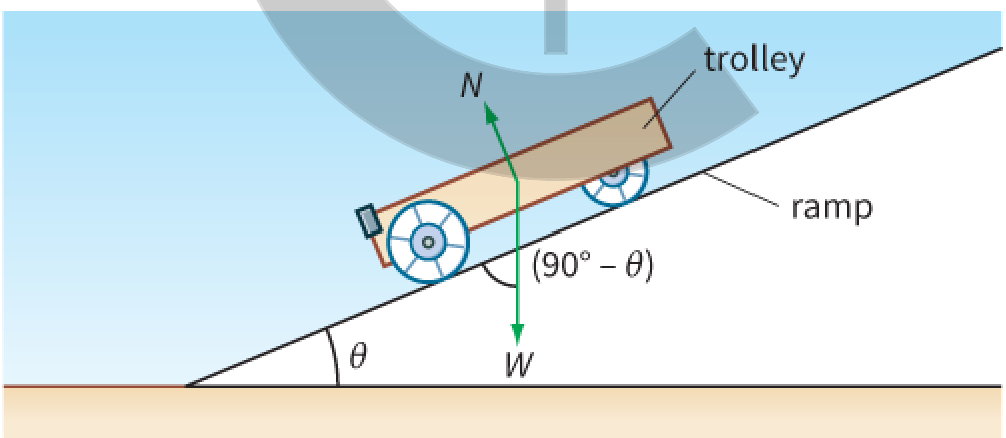


Figure 4.10: A force diagram for a trolley on a ramp.

You can see at once from Figure 4.10 that the forces cannot be balanced, since they do not act in the same straight line.

To find the component of W down the slope, we need to know the angle between W and the slope. The slope makes an angle θ with the horizontal, and from the diagram we can see that the angle between the weight and the ramp is $(90° - \theta)$. Using the rule for calculating the component of a vector given previously, we have:

$$\text{component of } W \text{ down the slope} = W\cos(90° - \theta) = W\sin\theta$$

(It is helpful to recall that $\cos(90° - \theta) = \sin\theta$; you can see this from Figure 4.10.)

Does the contact force N help to accelerate the trolley down the ramp? To answer this, we must calculate its component down the slope. The angle between N and the slope is 90°. So:

$$\text{component of } N \text{ down the slope} = N\cos 90° = 0$$

The cosine of 90° is zero, and so N has no component down the slope. This shows why it is useful to think in terms of the components of forces; we don't know the value of N, but, since it has no effect down the slope, we can ignore it.

(There's no surprise about this result. The trolley runs down the slope because of the influence of its weight, not because it is pushed by the contact force N.)

Changing the slope

If the students in Figure 4.9 increase the slope of their ramp, the trolley will move down the ramp with greater acceleration. They have increased θ, and so the component of W down the slope will have increased.

Now we can work out the trolley's acceleration. If the trolley's mass is m, its weight is mg. So the force F making it accelerate down the slope is:

$$F = mg\sin\theta$$

Since from Newton's second law for constant mass we have $a = \frac{F}{m}$, the trolley's acceleration a is given by:

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

We could have arrived at this result simply by saying that the trolley's acceleration would be the component of g down the slope (Figure 4.11). The steeper the slope, the greater the value of $\sin\theta$, and hence the greater the trolley's acceleration.

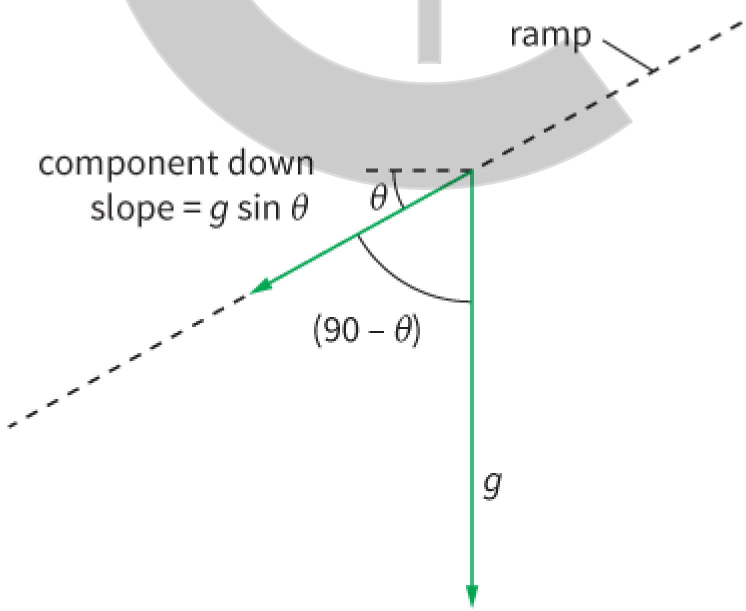


Figure 4.11: Resolving g down the ramp.

Questions

4 The person in Figure 4.12 is pulling a large box using a rope. Use the idea of components of a force to explain why they are more likely to get the box to move if the rope is horizontal (as in **a**) than if it is sloping upwards (as in **b**).

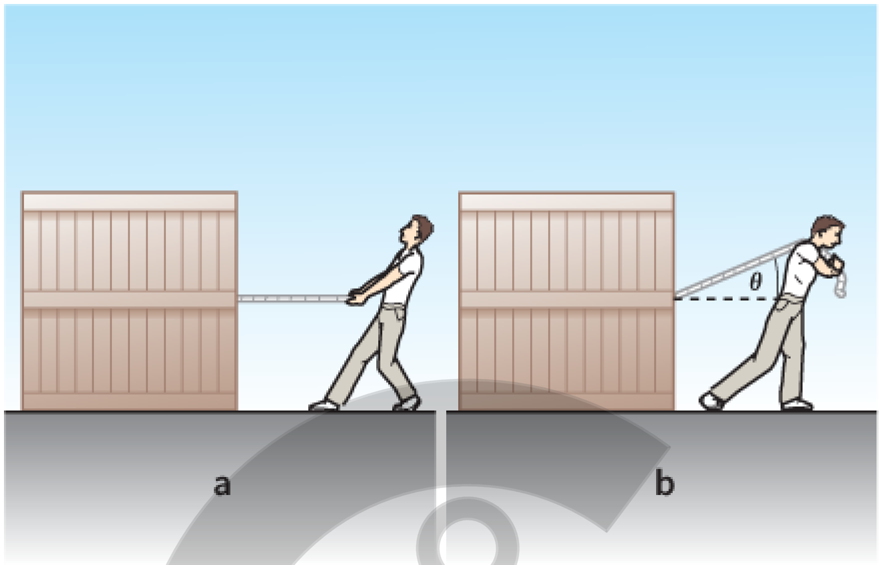


Figure 4.12: Why is it easier to move the box with the rope horizontal? For Question 4.

5 A crate is sliding down a slope. The weight of the crate is 500 N. The slope makes an angle of 30° with the horizontal.

- **a** Draw a diagram to show the situation. Include arrows to represent the weight of the crate and the contact force of the slope acting on the crate.
- **b** Calculate the component of the weight down the slope.
- **c** Explain why the contact force of the slope has no component down the slope.
- **d** What third force might act to oppose the motion? In which direction would it act?

Solving problems by resolving forces

A force can be resolved into two components at right angles to each other; these can then be treated independently of one another. This idea can be used to solve problems, as illustrated in Worked example 1.

WORKED EXAMPLE

1 A boy of mass 40 kg is on a waterslide that slopes at 30° to the horizontal. The frictional force up the slope is 120 N. Calculate the boy's acceleration down the slope. Take the acceleration of free fall g to be $9.81\ \mathrm{m\ s^{-2}}$.

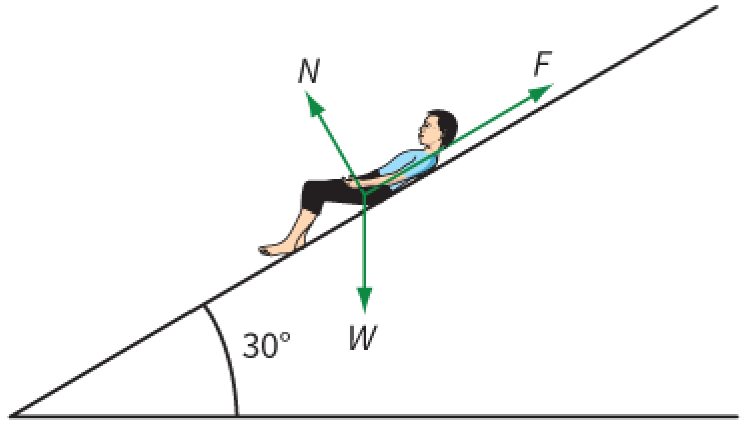


Figure 4.13: For Worked example 1.

Step 1 Draw a labelled diagram showing all the forces acting on the object of interest (Figure 4.13). This is known as a **free-body force diagram**. The forces are:

the boy's weight $W = 40 \times 9.81 = 392$ N

the frictional force up the slope $F = 120$ N

the contact force N at 90° to the slope.

Step 2 We are trying to find the resultant force on the boy that makes him accelerate down the slope. We resolve the forces down the slope, i.e., we find their components in that direction.

component of W down the slope = 392 × sin 30° = 196 N

component of F down the slope = −120 N (negative because F is directed up the slope)

component of N down the slope = 0 (because it is at 90° to the slope)

It is convenient that N has no component down the slope, since we do not know the value of N.

Step 3 Calculate the resultant force on the boy:

resultant force = 196 − 120 = 76 N

Step 4 Calculate his acceleration:

$$\text{acceleration} = \frac{\text{resultant force}}{\text{mass}} = \frac{76}{40} = 1.9\ \text{m s}^{-2}$$

So the boy's acceleration down the slope is 1.9 m s^{-2}. We could have arrived at the same result by resolving vertically and horizontally, but that would have led to two simultaneous equations from which we would have had to eliminate the unknown force N. It often helps to resolve forces at 90° to an unknown force.

Question

6 A child of mass 40 kg is on a water slide. The slide slopes down at 25° to the horizontal. The acceleration of free fall is 9.81 m s^{-2}. Calculate the child's acceleration down the slope:

a when there is no friction and the only force acting on the child is his weight

b if a frictional force of 80 N acts up the slope.

COALI

4.3 Centre of gravity

We have weight because of the force of gravity of the Earth on us. Each part of our body – arms, legs, head, for example – experiences a force, caused by the force of gravity. However, it is much simpler to picture the overall effect of gravity as acting at a single point. This is our **centre of gravity** – the point where all the weight of the object may be considered to act.

For a person standing upright, the centre of gravity is roughly in the middle of the body, behind the navel. For a sphere, it is at the centre. It is much easier to solve problems if we simply indicate an object's weight by a single force acting at the centre of gravity, rather than a large number of forces acting on each part of the object. Figure 4.14 illustrates this point. The athlete performs a complicated manoeuvre. However, we can see that his centre of gravity follows a smooth, parabolic path through the air, just like the paths of projectiles we discussed in Chapter 2.

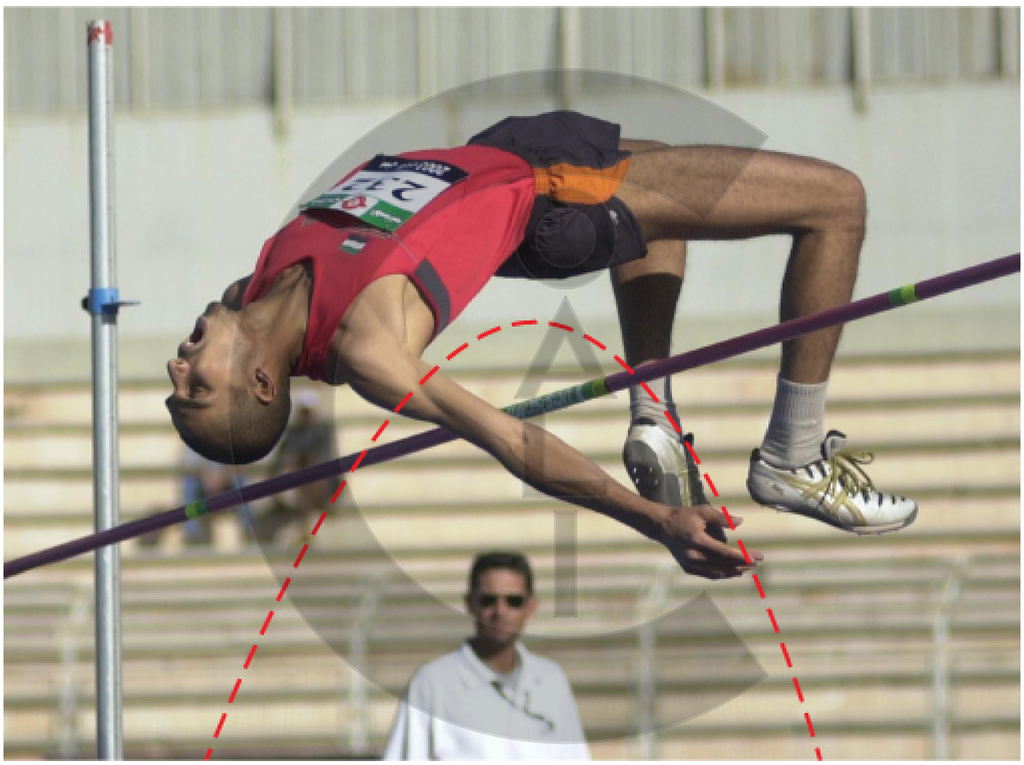


Figure 4.14: The dashed line indicates the path of the athlete's centre of gravity, which follows a smooth trajectory through the air. With his body curved like this, the athlete's centre of gravity is actually outside his body, just below the small of his back. At no time is the whole of his body above the bar.

PRACTICAL ACTIVITY 4.1

Finding the centre of gravity

The centre of gravity of a thin sheet, or lamina, of cardboard or metal can be found by suspending it freely from two or three points (Figure 4.15).

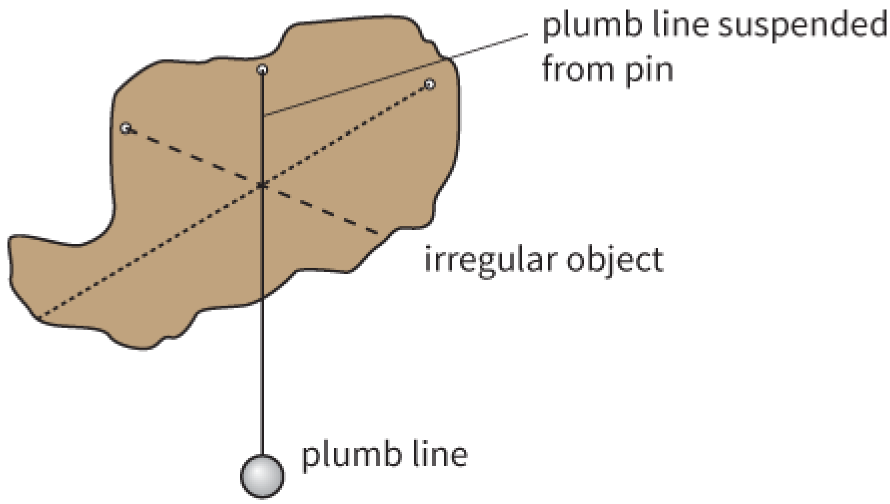


Figure 4.15: The centre of gravity is located at the intersection of the lines.

Small holes are made round the edge of the irregularly shaped object. A pin is put through one of the holes and held firmly in a clamp and stand so the object can swing freely. A length of string is attached to the pin. The other end of the string has a heavy mass attached to it. This arrangement is called a *plumb line*.

The object will stop swinging when its centre of gravity is vertically below the point of suspension. A line is drawn on the object along the vertical string of the plumb line. The centre of gravity must lie on this line. To find the position of the centre of gravity, the process is repeated with the object suspended from different holes. The centre of gravity will be at the point of intersection of the lines drawn on the object.

4.4 The turning effect of a force

Forces can make things accelerate. They can do something else as well: they can make an object turn round. We say that they can have a turning effect. Figure 4.16 shows how to use a spanner to turn a nut (a fastener with a threaded hole).

To maximise the turning effect of his force, the operator pulls close to the end of the spanner, as far as possible from the pivot (the centre of the nut) and at 90° to the spanner.

Figure 4.16: A mechanic turns a nut.

Moment of a force

The quantity that tells us about the turning effect of a force is its **moment**. The moment of a force depends on two quantities, the:

- magnitude of the force (the bigger the force, the greater its moment)
- perpendicular distance of the force from the pivot (the further the force acts from the pivot, the greater its moment).

The moment of a force = force × perpendicular distance of the pivot from the line of action of the force.

Figure 4.17a shows these quantities. The force F_1 is pushing down on the lever, at a perpendicular distance x_1 from the pivot. The moment of the force F_1 about the pivot is then given by:

$$\begin{aligned} \text{moment} &= \text{force} \times \text{distance from pivot} \\ &= F_1 \times x_1 \end{aligned}$$

The unit of moment is the newton metre (N m). This is a unit that does not have a special name. You can also determine the moment of a force in N cm.

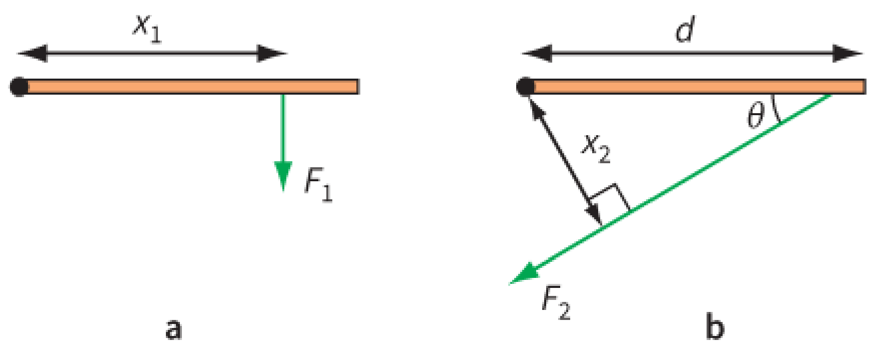


Figure 4.17: The quantities involved in calculating the moment of a force.

Figure 4.17b shows a slightly more complicated situation. F_2 is pushing at an angle θ to the lever, rather than at 90°. This makes it have less turning effect. There are two ways to calculate the moment of the force.

Method 1

Draw a perpendicular line from the pivot to the line of the force.

Find the distance x_2. Calculate the moment of the force, $F_2 \times x_2$. From the right-angled triangle, we can see that:

$$x_2 = d \sin \theta$$

Hence:

moment of force = $F_2 \times d \sin \theta = F_2 d \sin \theta$

Method 2

Calculate the component of F_2 that is at 90° to the lever.

This is $F_2 \sin \theta$. Multiply this by d.

$$\text{moment} = F_2 \sin \theta \times d$$

We get the same result as Method 1:

$$\text{moment of force} = F_2 d \sin \theta$$

Note that any force (such as the component $F_2 \cos \theta$) that passes through the pivot has no turning effect, because the distance from the pivot to the line of the force is zero.

Note also that we can calculate the moment of a force about any point, not just the pivot. However, in solving problems, it is often most convenient to take moments about the pivot as there is often an unknown force acting through the pivot (its contact force on the object).

Balanced or unbalanced?

We can use the idea of the moment of a force to solve two sorts of problem. We can:

- check whether an object will remain balanced or start to rotate
- calculate an unknown force or distance if we know that an object is balanced.

We can use the **principle of moments** to solve problems. The principle of moments states that, for any object that is in **equilibrium**, the sum of the clockwise moments about any point provided by the forces acting on the object equals the sum of the anticlockwise moments about that same point.

Note that, for an object to be in equilibrium, we also require that no resultant force acts on it. Worked examples 2, 3 and 4 illustrate how we can use these ideas to determine unknown forces.

WORKED EXAMPLES

2 Is the see-saw shown in Figure 4.18 in equilibrium (balanced), or will it start to rotate?

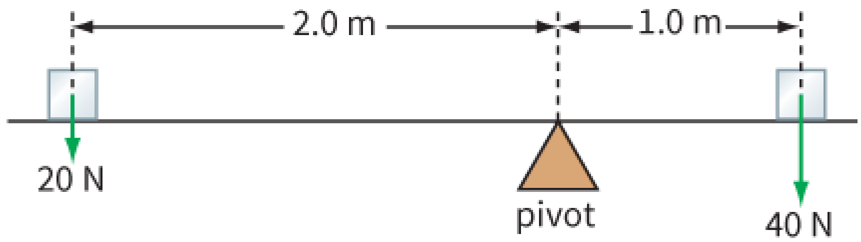


Figure 4.18: Will these forces make the see-saw rotate, or are their moments balanced?

The see-saw will remain balanced, because the 20 N force is twice as far from the pivot as the 40 N force.

To prove this, we need to think about each force individually. Which direction is each force trying to turn the see-saw, clockwise or anticlockwise? The 20 N force is tending to turn the see-saw anticlockwise, while the 40 N force is tending to turn it clockwise.

Step 1 Determine the anticlockwise moment:

moment of anticlockwise force = 20 × 2.0 = 40 N m

Step 2 Determine the clockwise moment:

moment of clockwise force = 40 × 1.0 = 40 N m

Step 3 We can see that:

clockwise moment = anticlockwise moment

So the see-saw is balanced and therefore does not rotate. The see-saw is in equilibrium.

3 The beam shown in Figure 4.19 is in equilibrium. Determine the force *X*.

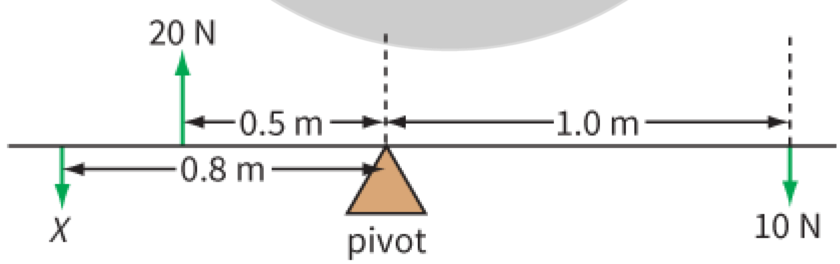


Figure 4.19: For Worked example 3.

The unknown force *X* is tending to turn the beam anticlockwise. The other two forces (10 N and 20 N) are tending to turn the beam clockwise. We will start by calculating their moments and adding them together.

Step 1 Determine the clockwise moments:

$$\text{sum of moments of clockwise forces} = (10 \times 1.0) + (20 \times 0.5) = 10 + 10 = 20 \text{ N m}$$

Step 2 Determine the anticlockwise moment:

moment of anticlockwise force = $X \times 0.8$

Step 3 Since we know that the beam must be balanced, we can write:

sum of clockwise moments = sum of anticlockwise moments

$$20 = X \times 0.8$$

$$X = \frac{20}{0.8} = 25 \text{ N}$$

So a force of 25 N at a distance of 0.8 m from the pivot will keep the beam still and prevent it from rotating (keep it balanced).

4 Figure 4.20 shows the internal structure of a human arm holding an object. The biceps is a muscle attached to one of the bones of the forearm. This muscle provides an upwards force.

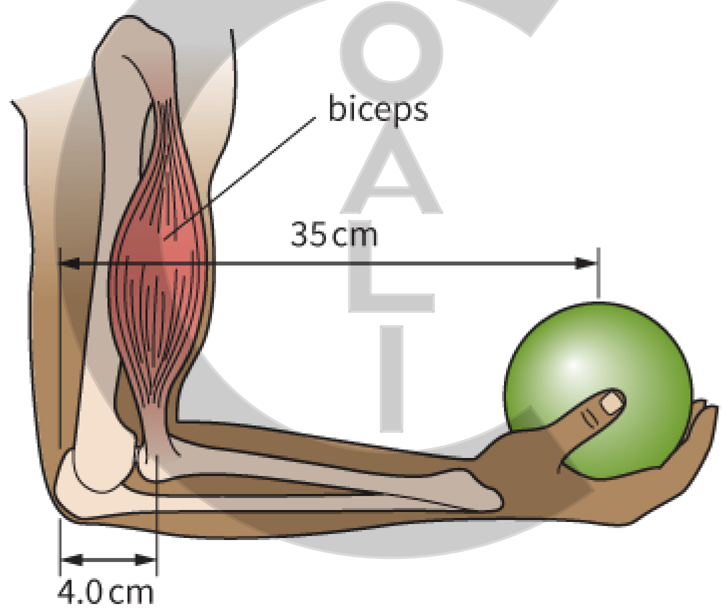


Figure 4.20: The human arm. For Worked example 4.

An object of weight 50 N is held in the hand with the forearm at right angles to the upper arm. Use the principle of moments to determine the muscular force F provided by the biceps, given the following data:

weight of forearm = 15 N

distance of biceps from elbow = 4.0 cm

distance of centre of gravity of forearm from elbow = 16 cm

distance of object in the hand from elbow = 35 cm

Step 1 There is a lot of information in this question. It is best to draw a simplified diagram of the forearm that shows all the forces and the relevant distances (Figure 4.21). All distances must be from the pivot, which in this case is the elbow.

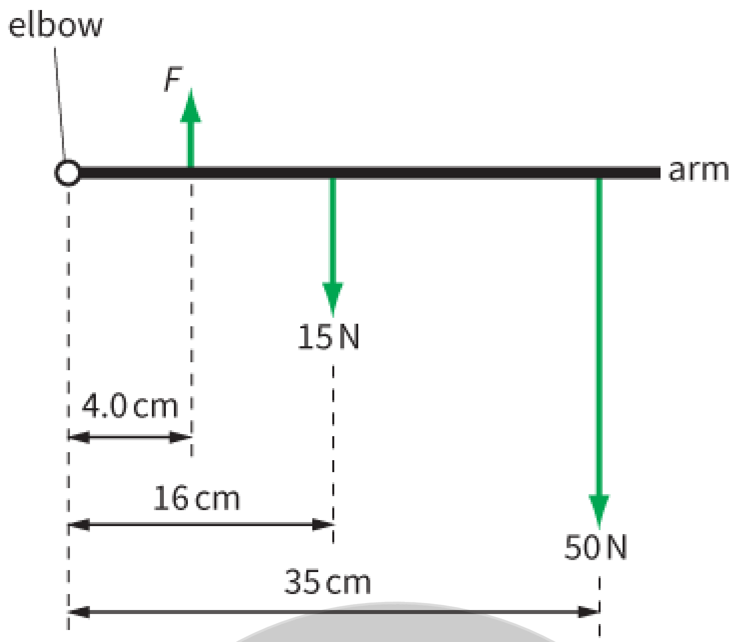


Figure 4.21: Simplified diagram showing forces on the forearm. For Worked example 4. Note that another force acts on the arm at the elbow; we do not know the size or direction of this force but we can ignore it by taking moments about the elbow.

Step 2 Determine the clockwise moments:

$$\begin{aligned}\text{sum of moments of clockwise forces} &= (15 + 0.16) + (50 \times 0.35)\\ &= 2.4 + 17.5\\ &= 19.9 \text{ N m}\end{aligned}$$

Step 3 Determine the anticlockwise moment:

moment of anticlockwise force = $F \times 0.04$

Step 4 Since the arm is in balance, according to the principle of moments we have:

sum of clockwise moments = sum of anticlockwise moments

$$\begin{aligned}19.9 &= 0.04F\\ F &= \frac{19.9}{0.04}\\ &= 497.5 \text{ N} \approx 500 \text{ N}\end{aligned}$$

The biceps provides a force of 500 N–a force large enough to lift 500 apples!

Questions

7 A wheelbarrow is loaded as shown in Figure 4.22.

a Calculate the force that the person needs to exert to hold the wheelbarrow's legs off the ground.

b Calculate the force exerted by the ground on the legs of the wheelbarrow (taken both together) when the gardener is not holding the handles.

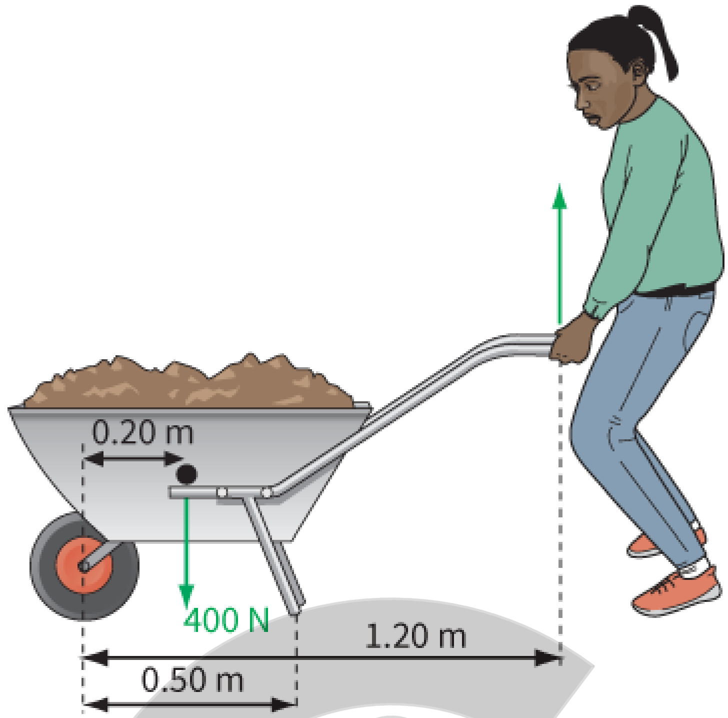


Figure 4.22: For Question 7.

8 A traditional pair of scales uses sliding masses of 10 g and 100 g to achieve a balance. A diagram of the arrangement is shown in Figure 4.23. The bar itself is supported with its centre of gravity at the pivot.

a Calculate the value of the mass *M*, attached at X.

b State **one** advantage of this method of measuring mass.

c Determine the upwards force of the pivot on the bar.

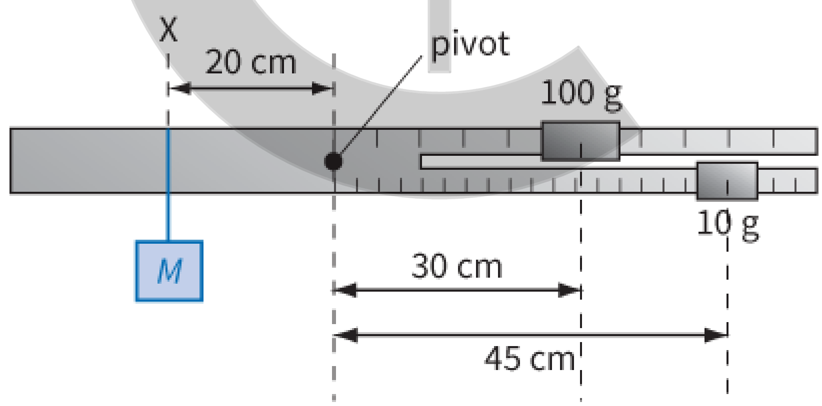


Figure 4.23: For Question 8.

9 Figure 4.24 shows a beam with four forces acting on it.

a For each force, calculate the moment of the force about point P.

b State whether each moment is clockwise or anticlockwise.

c State whether or not the moments of the forces are balanced.

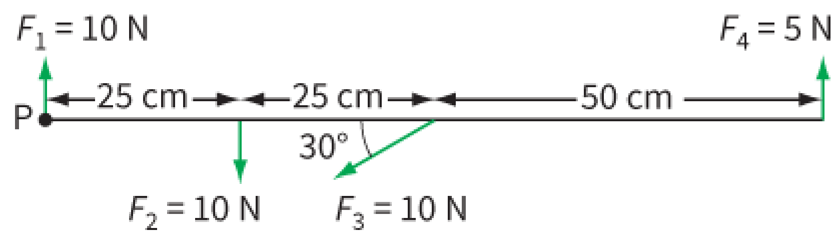


Figure 4.24: For Question 9.

4.5 The torque of a couple

Figure 4.25 shows the forces needed to turn a car's steering wheel. The two forces balance up and down (15 N up and 15 N down), so the wheel will not move up, down or sideways. However, the wheel is not in equilibrium. The pair of forces will cause it to rotate.

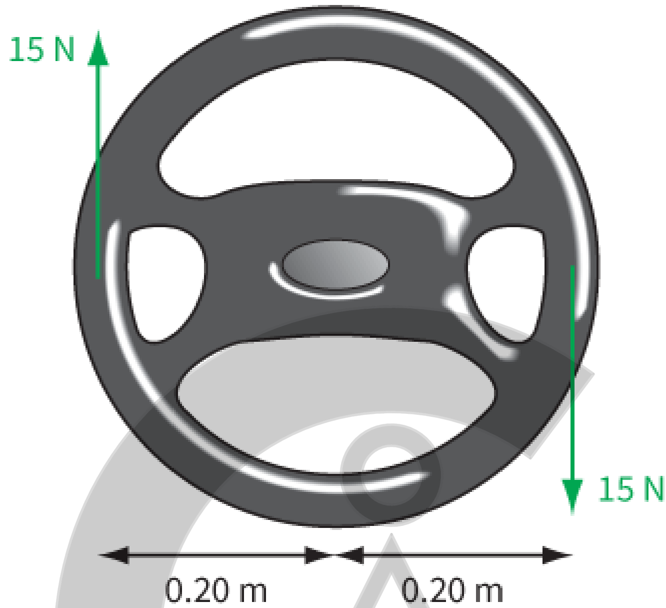


Figure 4.25: Two forces act on this steering wheel to make it turn.

A pair of forces like that in Figure 4.25 is known as a **couple**.

A couple has a turning effect, but does not cause an object to accelerate. To form a couple, the two forces must be:

- equal in magnitude
- parallel, but opposite in direction
- separated by a distance *d*.

The turning effect or moment of a couple is known as its **torque**.

We can calculate the torque of the couple in Figure 4.25 by adding the moments of each force about the centre of the wheel:

$$\begin{aligned} \text{torque of couple} &= (15 \times 0.20) + (15 \times 0.20) \\ &= 6.0\ \text{N m} \end{aligned}$$

We could have found the same result by multiplying one of the forces by the perpendicular distance between them:

torque of a couple = 15 × 0.4 = 6.0 N m

The torque of a couple is defined as follows:

torque of a couple = one of the forces × perpendicular distance between the forces

Question

10 The driving wheel of a car travelling at a constant velocity has a torque of 137 N m applied to it by the axle that drives the car (Figure 4.26). The radius of the tyre is 0.18 m. Calculate the driving force provided by this wheel.

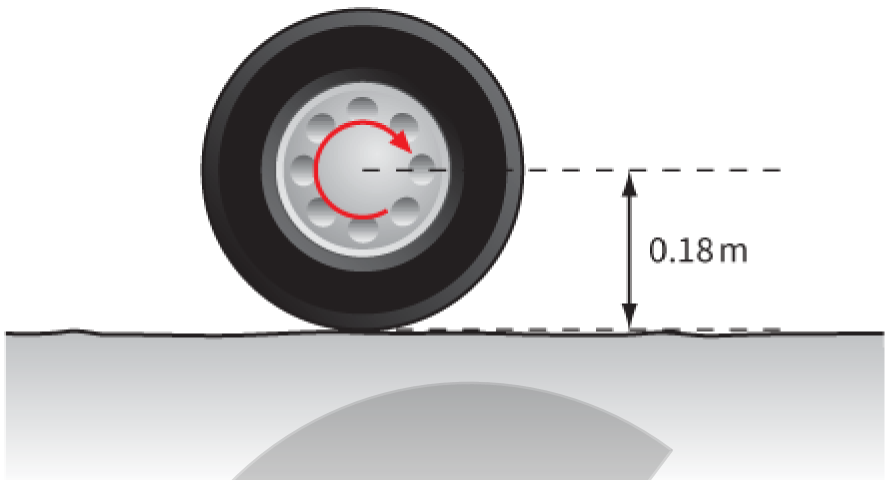


Figure 4.26: For Question 10.

Pure turning effect

When we calculate the moment of a single force, the result depends on the point or pivot about which the moment acts. The further the force is from the pivot, the greater the moment. A couple is different; the moment of a couple does not depend on the point about which it acts, only on the perpendicular distance between the two forces. A single force acting on an object will tend to make the object accelerate (unless there is another force to balance it). A couple, however, is a pair of equal and opposite forces, so it will not make the object accelerate. This means we can think of a couple as a pure 'turning effect', the size of which is given by its torque.

For an object to be in equilibrium, two conditions must be met at the same time:

- The resultant force acting on the object is zero.
- The resultant moment is zero.

KEY IDEA

If a body is in equilibrium, there is no resultant force and no resultant torque or moment about any point.

REFLECTION

Are there any things that you need more help with to fully understand vectors and moments?

Work out a simple way for yourself to remember which component is which. Check your method by explaining it to someone else with lots of examples.

SUMMARY

Forces are vector quantities that can be added by means of a vector triangle. Their resultant can be determined using trigonometry or by scale drawing.

Forces can be resolved into components. Components at right angles to one another can be treated independently of one another. For a force F at an angle θ to the x-direction, the components are:

$$x\text{-direction: } F\cos\theta$$

$$y\text{-direction: } F\sin\theta$$

The moment of a force = force × perpendicular distance of the pivot from the line of action of the force.

The principle of moments states that, for any object in equilibrium, the sum of the clockwise moments about any point provided by the forces acting on the object is equal to the sum of the anticlockwise moments about that same point.

A couple is a pair of equal, parallel but opposite forces whose effect is to produce a turning effect on a body without giving it linear acceleration.

torque of a couple = one of the forces × perpendicular distance between the forces

For an object in equilibrium, the resultant force acting on the object must be zero and the resultant moment must be zero.

EXAM-STYLE QUESTIONS

1 A force F is applied at a distance d from the hinge H and an angle x to the door.

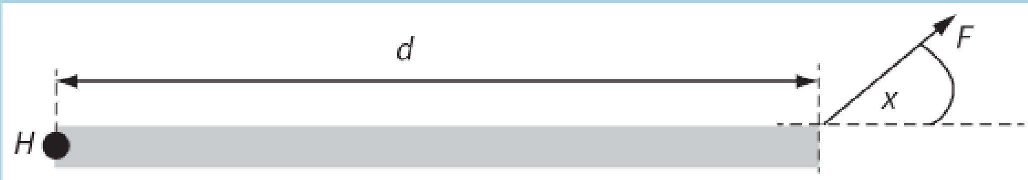


Figure 4.27

What is the moment of the force F about the point H? **[1]**

A $Fd \cos x$

B $\frac{Fd}{\cos x}$

C $Fd \sin x$

D $\frac{Fd}{\sin x}$

2 The angle between two forces, each of magnitude 5.0 N, is 120°.

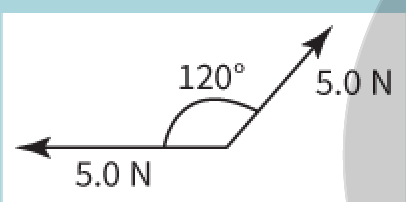


Figure 4.28

What is the magnitude of the resultant of these two forces? **[1]**

A 1.7 N

B 5.0 N

C 8.5 N

D 10 N

3 A ship is pulled at a constant speed by two small boats, A and B, as shown. The engine of the ship does not produce any force.

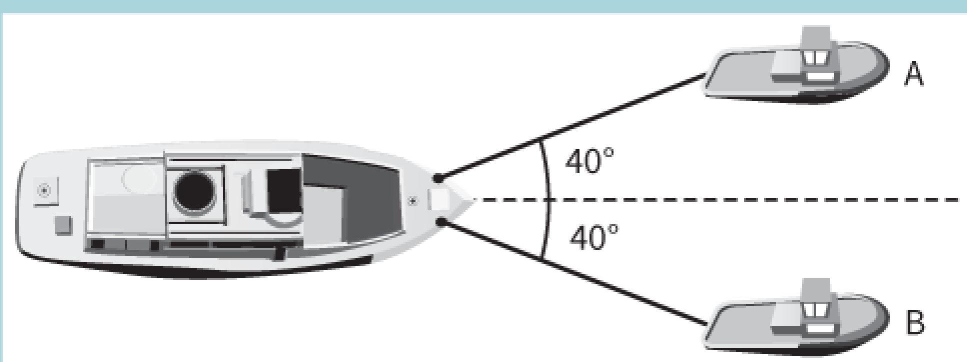


Figure 4.29

The tension in each cable between A and B and the ship is 4000 N.

a Draw a free-body diagram showing the three horizontal forces acting on the ship. **[2]**

b Draw a vector diagram to scale showing these three forces and use your diagram to find the value of the drag force on the ship. **[2]**

[Total: 4]

4 A block of mass 1.5 kg is at rest on a rough surface which is inclined at 20° to the horizontal as shown.

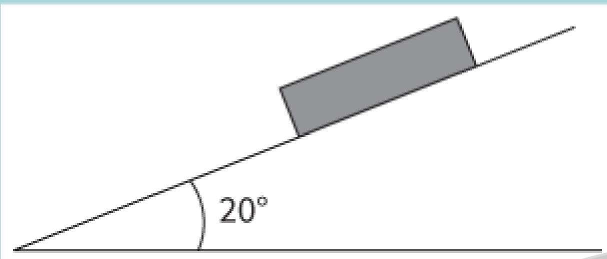


Figure 4.30

a Draw a free-body diagram showing the three forces acting on the block. **[2]**

b Calculate the component of the weight that acts down the slope. **[2]**

c Use your answer to part **b** to determine the force of friction that acts on the block. **[2]**

d If the angle of the surface is actually measured as 19° and 21° determine the absolute uncertainty in this angle and the uncertainty this produces in the value for part **b**. **[3]**

e Determine the normal contact force between the block and the surface. **[3]**

[Total: 12]

5 This free-body diagram shows three forces that act on a stone hanging at rest from two strings.

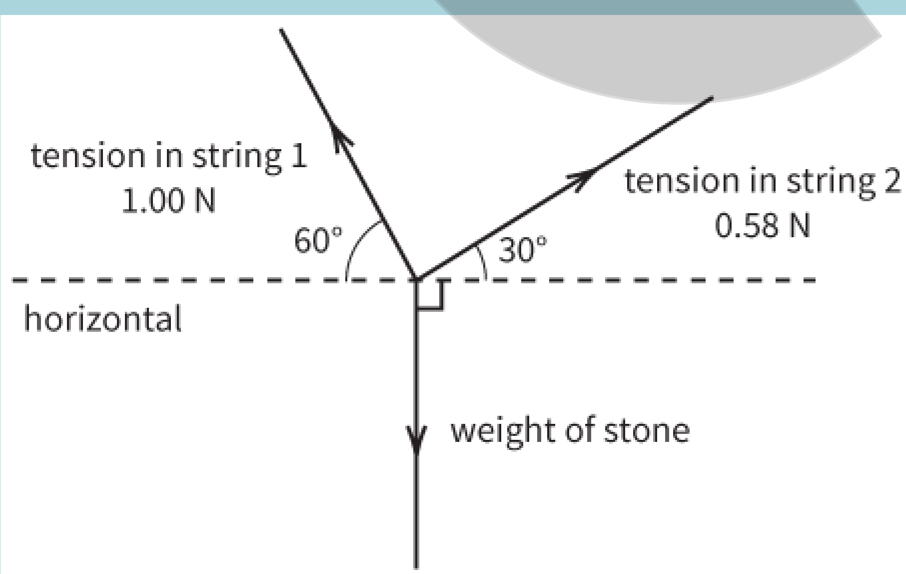


Figure 4.31

a Calculate the horizontal component of the tension in each string. State why **[5]**

these two components are equal in magnitude?

b Calculate the vertical component of the tension in each string. **[4]**

c Use your answer to part **b** to calculate the weight of the stone. **[2]**

d Draw a vector diagram of the forces on the stone. This should be a triangle of forces. **[1]**

e Use your diagram in part **d** to calculate the weight of the stone. **[2]**

[Total: 14]

6 The force *F* shown here has a moment of 40 N m about the pivot. Calculate the magnitude of the force *F*. **[4]**

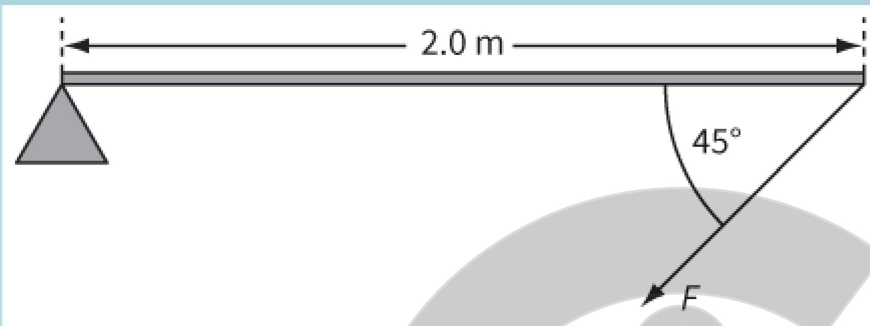


Figure 4.32

7 The asymmetric bar shown has a weight of 7.6 N and a centre of gravity that is 0.040 m from the wider end, on which there is a load of 3.3 N. It is pivoted a distance of 0.060 m from its centre of gravity. Calculate the force *P* that is needed at the far end of the bar in order to maintain equilibrium. **[4]**

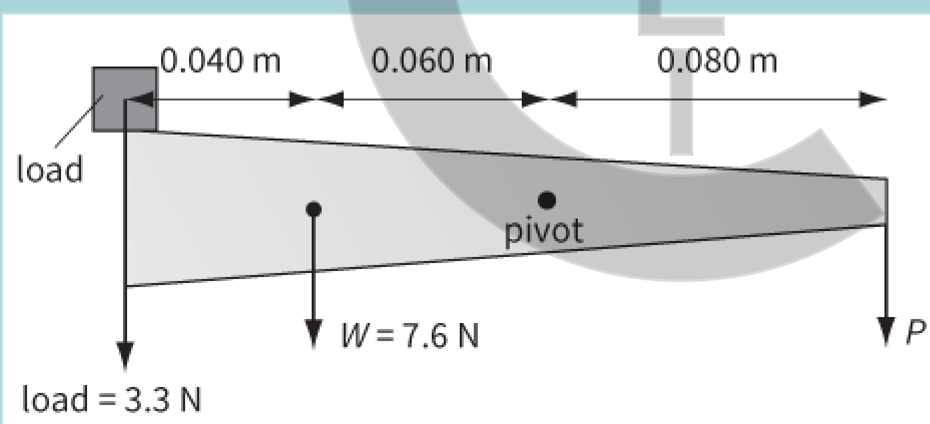


Figure 4.33

8 **a** State what is meant by:

i a couple **[1]**

ii torque. **[2]**

b The engine of a car produces a torque of 200 N m on the axle of the wheel in contact with the road. The car travels at a constant velocity towards the right:

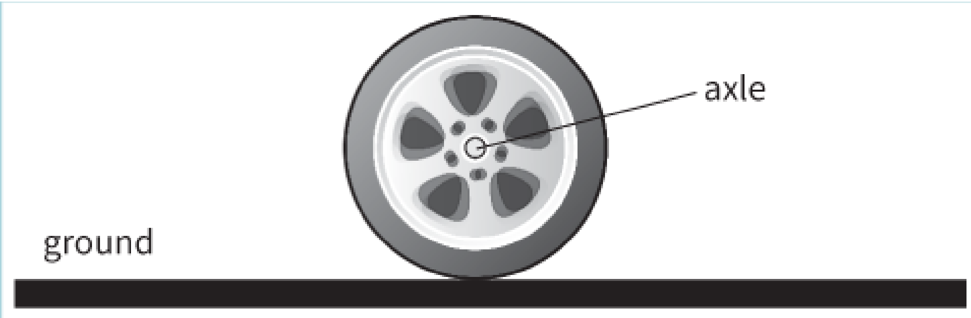


Figure 4.34

i Copy the diagram of the wheel and show the direction of rotation of the wheel, and the horizontal component of the force that the road exerts on the wheel. **[2]**

ii State the resultant torque on the wheel. Explain your answer. **[2]**

iii The diameter of the car wheel is 0.58 m. Determine the value of the horizontal component of the force of the road on the wheel. **[1]**

[Total: 8]

9 a Explain what is meant by the centre of gravity of an object. **[2]**

b A flagpole of mass 25 kg is held in a horizontal position by a cable as shown. The centre of gravity of the flagpole is at a distance of 1.5 m from the fixed end.

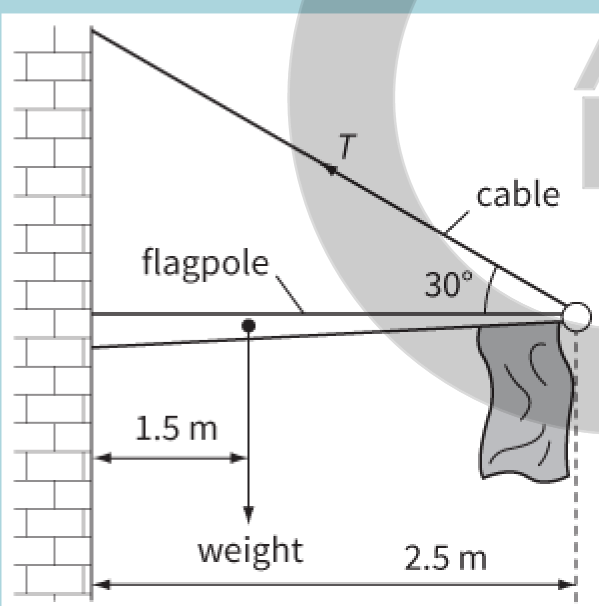


Figure 4.35

i Write an equation to represent taking moments about the left-hand end of the flagpole. Use your equation to find the tension T in the cable. **[4]**

ii Determine the vertical component of the force at the left-hand end of the flagpole. **[2]**

[Total: 8]

10 a State the two conditions necessary for an object to be in equilibrium. **[2]**

b A metal rod of length 90 cm has a disc of radius 24 cm fixed rigidly at its centre, as shown. The assembly is pivoted at its centre.

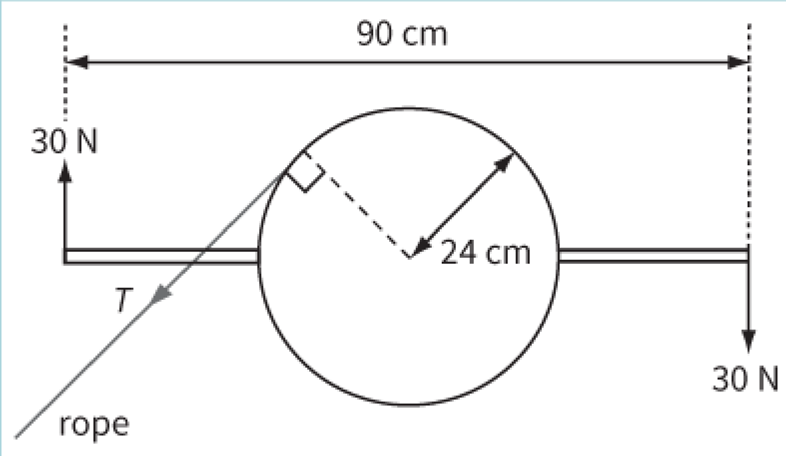


Figure 4.36

Two forces, each of magnitude 30 N, are applied normal to the rod at each end so as to produce a turning effect on the rod. A rope is attached to the edge of the disc to prevent rotation.

Calculate:

i the torque of the couple produced by the 30 N forces **[1]**

ii the tension T in the rope. **[3]**

[Total: 6]

11 a State what is meant by the torque of a couple. **[2]**

b Three strings, A, B and C, are attached to a circular ring, as shown in Figure 4.35.

The strings and the ring all lie on a smooth horizontal surface and are at rest. The tension in string A is 8.0 N. Calculate the tension in strings B and C. **[4]**

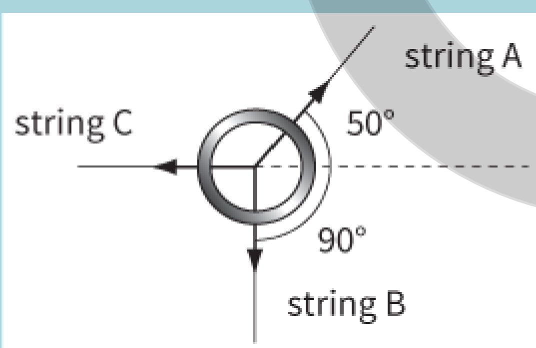


Figure 4.37

[Total: 6]

12 This diagram shows a picture hanging symmetrically by two cords from a nail fixed to a wall. The picture is in equilibrium.

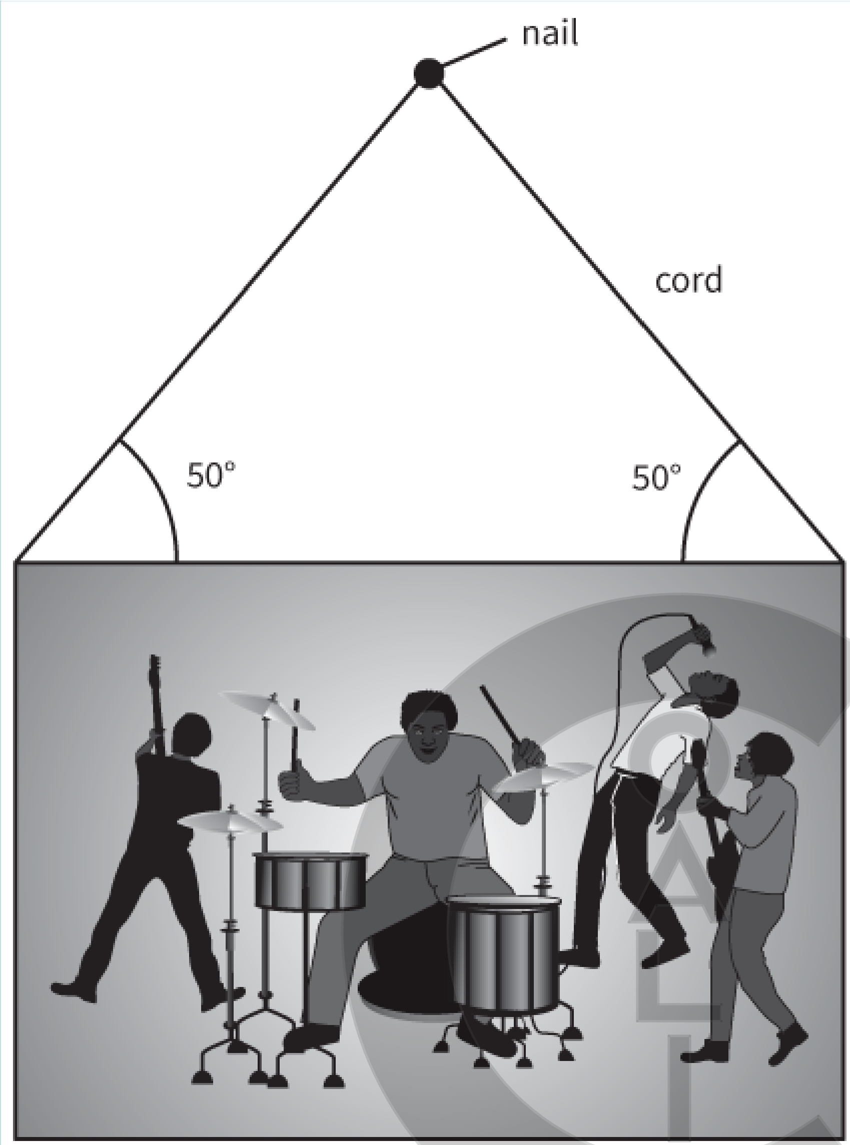


Figure 4.38

a Explain what is meant by equilibrium. **[2]**

b Draw a vector diagram to represent the three forces acting on the picture in the vertical plane. Label each force clearly with its name and show the direction of each force with an arrow. **[2]**

c The tension in the cord is 45 N and the angle that each end of the cord makes with the horizontal is 50°. Calculate:

i the vertical component of the tension in the cord **[1]**

ii the weight of the picture. **[1]**

[Total: 6]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
add forces using a vector triangle	4.1			
resolve forces into perpendicular components	4.2			
represent the weight of a body at a single point known as its centre of gravity	4.3			
define and apply the moment of a force and the torque of a couple	4.4, 4.5			
state and apply the principle of moments	4.4			
use the idea that, when there is no resultant force and no resultant torque, a system is in equilibrium.	4.5			

⟩ Chapter 5

Work, energy and power

LEARNING INTENTIONS

In this chapter you will learn how to:

- use the concept of work and energy
- recall and apply the principle of conservation of energy
- recall and understand that the efficiency of a system is the ratio of useful energy output from the system to the total energy input
- use the concept of efficiency to solve problems
- define and use the equation for power using $P = \frac{W}{t}$ and derive $P = Fv$
- derive and use the formulae for kinetic energy and gravitational potential energy.

BEFORE YOU START

- Write down definitions for energy, work and power.
- Write down all that you know about these topics and share your ideas with someone else. Be prepared to discuss your answers with the rest of the class.

THE IDEA OF ENERGY

The Industrial Revolution started in the late 18th century in Britain. Today, many other countries have become or are becoming industrialised (Figure 5.1). Industrialisation is the development of new machines capable of doing the work of hundreds of craftsmen and labourers. At first, people used water and wind to power machines. Water stored behind a dam was used to turn a wheel, which turned many machines. Steam engines were developed, initially for pumping water out of mines. Steam engines use a fuel such as coal; there is much more energy stored in 1 kg of coal than in 1 kg of water held behind a dam.

Nowadays, most factories rely on electrical power, generated by burning coal or gas at a power station. High-pressure steam is generated, and this turns a turbine that turns a generator. Even in the most efficient coal-fired power station, only about 40% of the energy from the fuel is transferred to the electrical energy that the station supplies to the electricity grid.

Figure 5.1: Anshan steel works, China.

Engineers worked hard to develop machines that made the most efficient use of the energy supplied to them. At the same time, scientists were working out the basic ideas of energy transfer and energy transformations. The idea of energy itself had to be developed; it was not obvious at first that heat, light, electrical energy were all forms of the same thing: energy. What is the history of your country in developing the use of machines, generating electrical power and increasing efficiency?

The earliest steam engines had very low efficiencies–many converted less than 1% of the energy supplied to them into useful work. The understanding of the relationship between work and energy led to many clever ways of making the most of the energy supplied by fuel.

Figure 5.2: The jet engines of this aircraft are designed to make efficient use of their fuel. If they were less efficient, their thrust might only be sufficient to lift the empty aircraft and the passengers would have to be left behind.

CALI

5.1 Doing work, transferring energy

The weight-lifter shown in Figure 5.3 has powerful muscles. They can provide the force needed to lift a large weight above her head – about 2 m above the ground. The force exerted by the weight-lifter transfers energy from her to the weights. We know that the weights have gained energy because, when the athlete releases them, they come crashing down to the ground.

As the athlete lifts the weights and transfers energy to them, we say that her lifting force is doing work. ‘Doing work’ is a way of transferring energy from one object to another. In fact, if you want to know the scientific meaning of the word ‘energy’, we have to say it is ‘that which is transferred when a force moves through a distance’. So, work and energy are two closely linked concepts.

In physics, we often use an everyday word but with a special meaning. **Work** is an example of this.

Figure 5.3: It is hard work being a weight-lifter.

Doing work	Not doing work
Pushing a car to start it moving: your force transfers energy to the car. The car's kinetic energy (that is, 'movement energy') increases.	Pushing a car but it does not budge: no energy is transferred, because your force does not move it. The car's kinetic energy does not change.
Lifting weights: you are doing work as the weights move upwards. The gravitational potential energy of the weights increases.	Holding weights above your head: you are not doing work on the weights (even though you may find it tiring) because the force you apply is not moving them. The gravitational potential energy of the weights is not changing.
A falling stone: the force of gravity is doing work. The stone's kinetic energy is increasing.	The Moon orbiting the Earth: the force of gravity is not doing work. The Moon's kinetic energy is not changing.
Writing an essay: you are doing work because you need a force to move your pen across the page, or to press the keys on the keyboard.	Reading an essay: this may seem like 'hard work', but no force is involved, so you are not doing any work.

Table 5.1: The meaning of 'doing work' in physics.

Table 5.1 describes some situations that illustrate the meaning of **doing work** in physics.

It is important to understand that our bodies sometimes mislead us. If you hold a heavy weight above your head for some time, your muscles will get tired. However, you are not doing any work **on the weights**, because you are not transferring energy to the weights once they are above your head. Your muscles get tired because they are constantly relaxing and contracting, and this uses energy, but none of the energy is being transferred to the weights.

Calculating work done

Because **doing work** defines what we mean by **energy**, we start this chapter by considering how to calculate **work done**.

There is no doubt that you do work if you push a car along the road. A force transfers energy from you to the car. But how much work do you do? Figure 5.4 shows the two factors involved:

- the size of the force F – the bigger the force, the greater the amount of work you do
- the distance s you push the car – the further you push it, the greater the amount of work done.

So, the bigger the force, and the further it moves, the greater the amount of work done.

KEY IDEA

Work is done on a body when a force moves (displaces) the body in the direction of the force. Energy is then transferred from one body to another

The work done by a force is defined as the product of the force and the distance moved in the direction of the force:

$$W = F \times s$$

where s is the distance moved in the direction of the force.

In the example shown in Figure 5.4, F = 300 N and s = 5.0 m, so:

$$\text{work done } W = F \times s = 300 \times 5.0 = 1500 \text{ J}$$

KEY EQUATION

$$\text{work done} = \text{force} \times \text{distance}$$
$$W = Fs$$

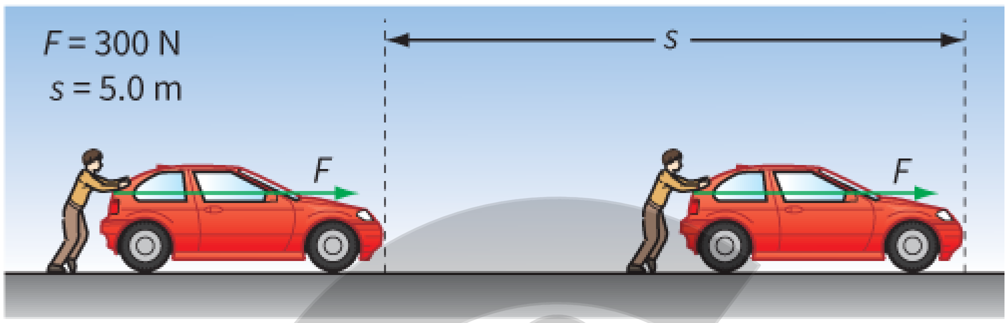


Figure 5.4: You have to do work to start the car moving.

Energy transferred

Doing work is a way of transferring energy. For both energy and work the correct SI unit is the joule (J).

The amount of work done, calculated using $W = F \times s$, shows the amount of energy transferred:

$$\text{work done} = \text{energy transferred}$$

KEY IDEA

work done = energy transferred

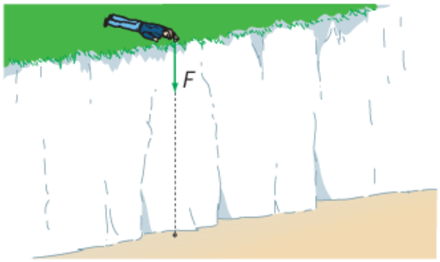
Newtons, metres and joules

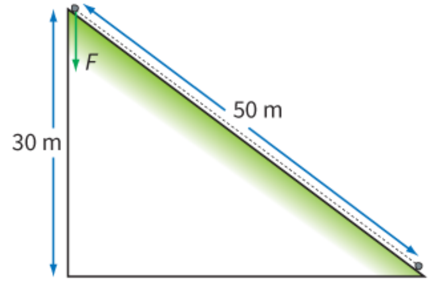
From the equation $W = Fs$ we can see how the unit of force (the newton), the unit of distance (the metre) and the unit of work or energy (the joule) are related.

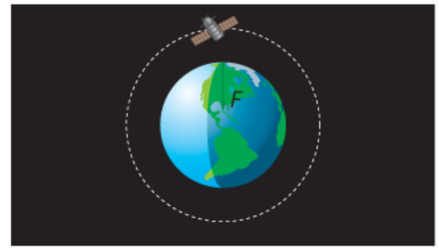
The joule is defined as the amount of work done when a force of 1 newton moves a distance of 1 metre in the direction of the force. Since **work done** = **energy transferred**, it follows that a joule is also the amount of energy transferred when a force of 1 newton moves a distance of 1 metre in the direction of the force.

Force, distance and direction

It is important to understand that, for a force to do work, there must be movement in the direction of the force. Both the force F and the distance s moved in the direction of the force are vector quantities, so you should know that their directions are likely to be important. To illustrate this, we will consider three examples involving gravity (Figure 5.5). In the equation for work done, $W = F \times s$, the distance moved s is the displacement in the direction of the force.







1 A stone weighing 5.0 N is dropped from the top of a 50 m high cliff.	**2** A stone weighing 5.0 N rolls 50 m down a slope.	**3** A satellite orbits the Earth at a constant height and at a constant speed. The weight of the satellite at this height is 500 N.
What is the work done by the force of gravity?	What is the work done by the force of gravity?	What is the work done by the force of gravity?
force on stone F = pull of gravity = weight of stone = 5.0 N vertically downwards	force on stone F = pull of gravity = weight of stone = 5.0 N vertically downwards	force on satellite F = pull of gravity = weight of satellite = 500 N towards centre of Earth
distance moved by stone in direction of force s = 50 m vertically downwards	distance moved by stone down slope = 50 m, **but** distance moved in direction of force s = 30 m	distance moved by satellite towards centre of Earth (that is, in the direction of force) s = 0. The satellite remains at a constant distance from the Earth. It does not move in the direction of F.
work done $= F \times s$ $= 5.0 \times 50$ $= 250$ J	work done $= F \times s$ $= 5.0 \times 30$ $= 150$ J	work done $= F \times s$ $= 500 \times 0$ $= 0$ J

Figure 5.5: Three examples involving gravity.

Questions

1 In each of the following examples, explain whether or not any work is done by the force mentioned.

   a You pull a heavy sack along rough ground.

   b The force of gravity pulls you downwards when you fall off a wall.

   c The tension in a string pulls on a stone when you whirl it around in a circle at a steady speed.

   d The contact force of the bedroom floor stops you from falling into the room below.

2 A man of mass 70 kg climbs stairs of vertical height 2.5 m. Calculate the work done against the force of gravity. (Take g = 9.81 m s^{-2}.)

3 A stone of weight 10 N falls from the top of a 250 m high cliff.

   a Calculate how much work is done by the force of gravity in pulling the stone to the foot of the cliff.

   b How much energy is transferred to the stone if air resistance is ignored?

Suppose that the force F moves through a distance s that is at an angle θ to F, as shown in Figure 5.6. To determine the work done by the force, it is simplest to determine the component of F in the direction of s. This component is $F \cos \theta$, and so we have:

$$\text{work done} = (F\cos\theta) \times s$$

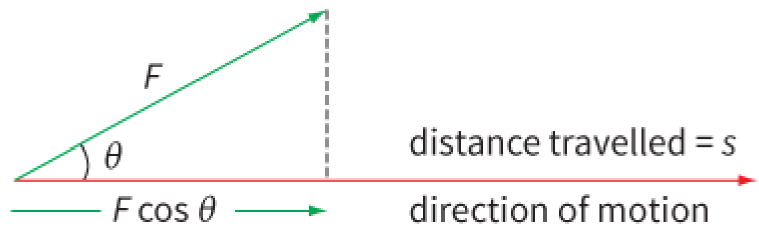


Figure 5.6: The work done by a force depends on the angle between the force and the distance it moves.

or simply:

$$\text{work done} = Fs\cos\theta$$

Worked example 1 shows how to use this.

KEY EQUATION

$$\text{work done} = Fs\cos\theta$$

WORKED EXAMPLE

1 A man pulls a box along horizontal ground using a rope (Figure 5.7). The force provided by the rope is 200 N, at an angle of 30° to the horizontal.

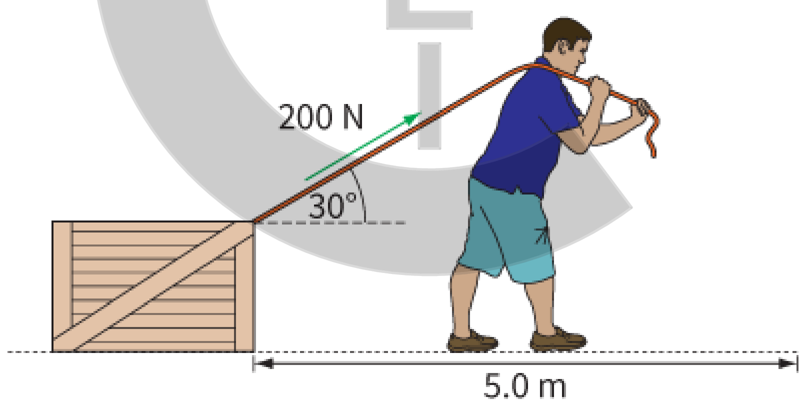


Figure 5.7: For Worked example 1.

Calculate the work done if the box moves 5.0 m along the ground.

Step 1 Calculate the component of the force in the direction in which the box moves. This is the horizontal component of the force:

horizontal component of force = 200 cos 30° ≈ 173 N

Hint: *F cos θ is the component of the force at an angle θ to the direction of motion.*

Step 2 Now calculate the work done:

work done = force × distance moved

= 173 × 5.0 = 865 J

Hint: *Note that we could have used the equation work done = Fs cos θ to combine the two steps into one.*

Questions

4 The crane shown in Figure 5.8 lifts its 500 N load to the top of the building from A to B. Distances are as shown on the diagram. Calculate how much work is done by the crane.

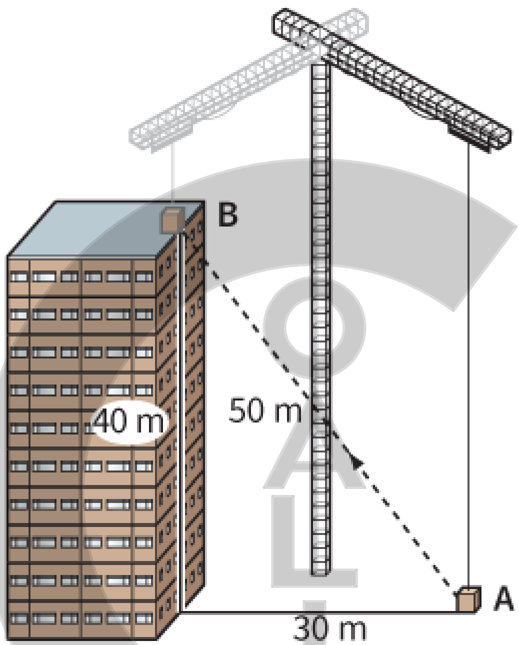


Figure 5.8: For Question 4. The dotted line shows the track of the load as it is lifted by the crane.

5 Figure 5.9 shows the forces acting on a box that is being pushed up a slope. Calculate the work done by each force if the box moves 0.50 m up the slope.

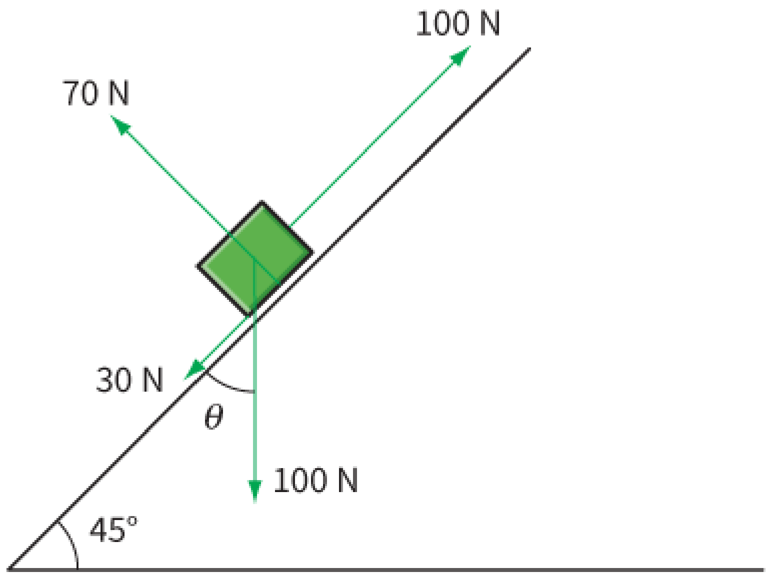


Figure 5.9: For Question 5.

5.2 Gravitational potential energy

If you lift a heavy object, you do work. You are providing an upwards force to overcome the downwards force of gravity on the object. The force moves the object upwards, so the force is doing work.

In this way, energy is transferred from you to the object. You lose energy, and the object gains energy. We say that the **gravitational potential energy, E_p** of the object has increased.

Worked example 2 shows how to calculate a change in gravitational potential energy (g.p.e.).

WORKED EXAMPLE

2 A weight-lifter raises weights with a mass of 200 kg from the ground to a height of 1.5 m. Calculate how much work he does. By how much does the g.p.e. of the weights increase?

Step 1 As shown in Figure 5.10, the downward force on the weights is their weight, $W = mg$. An equal, upward force F is required to lift them.

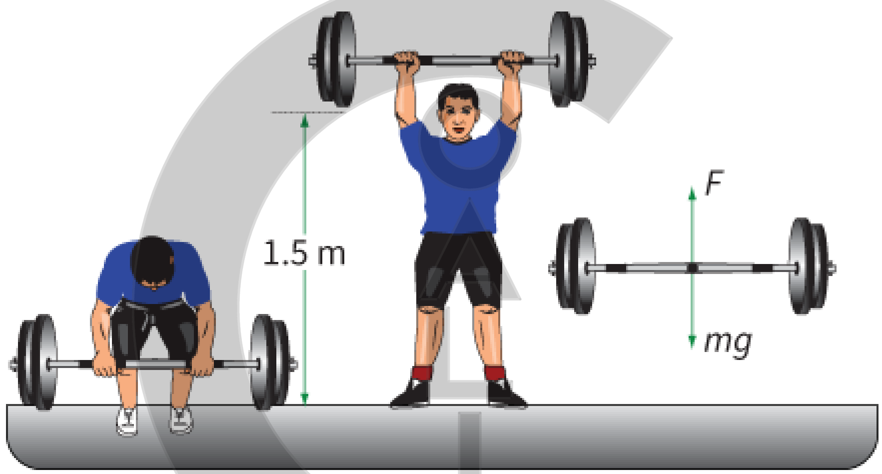


Figure 5.10: For Worked example 2.

$W = F = mg = 200 \times 9.81 = 1962$ N

Hint: *It helps to draw a diagram of the situation.*

Step 2 Now we can calculate the work done by the force F:

$$\begin{aligned} \text{work done} &= \text{force} \times \text{distance moved} \\ &= 1962 \times 1.5 \approx 2940 \text{ J} \end{aligned}$$

Note that the distance moved is in the same direction as the force. So the work done on the weights is about 2940 J. This is also the value of the increase in their g.p.e.

An equation for gravitational potential energy

The change (Δ) in the gravitational potential energy (g.p.e.) of an object, E_p, depends on the change in its height, h. We can calculate E_p using this equation:

$$\begin{aligned} \text{change in g.p.e} &= \text{weigth} \times \text{change in height} \\ \Delta E_p &= (m \times g) \times \Delta h \\ \Delta E_p &= mg\Delta h \end{aligned}$$

It should be clear where this equation comes from. The force needed to lift an object is equal to its weight *mg*, where *m* is the mass of the object and *g* is the acceleration of free fall or the gravitational field strength on the Earth's surface. The work done by this force is given by force × distance moved, or weight × change in height. You might feel that it takes a force greater than the weight of the object being raised to lift it upwards, but this is not so. Provided the force is equal to the weight, the object will move upwards at a steady speed.

KEY EQUATION

$$\begin{aligned} \text{change in g.p.e} &= \text{weight} \times \text{change in height} \\ \Delta E_p &= mg\Delta h \end{aligned}$$

You must learn how to derive this equation.

Note that *h* stands for the vertical height through which the object moves. Note also that we can only use the equation $\Delta E_p = mg\Delta h$ for relatively small changes in height. It would not work, for example, in the case of a satellite orbiting the Earth. Satellites orbit at a height of at least 200 km and *g* has a smaller value at this height.

Other forms of potential energy

Potential energy is the energy an object has because of its position or shape. So, for example, an object's gravitational potential energy changes when it moves through a gravitational field. (There is much more about gravitational fields in Chapter 17.)

We can identify other forms of potential energy. An electrically charged object has electric potential energy when it is placed in an electric field (see Chapter 21). An object may have elastic potential energy when it is stretched, squashed or twisted–if it is released it goes back to its original shape (see Chapter 7).

Questions

6 Calculate how much gravitational potential energy is gained if you climb a flight of stairs. Assume that you have a mass of 52 kg and that the height you lift yourself is 2.5 m.

7 A climber of mass 100 kg (including the equipment she is carrying) ascends from sea level to the top of a mountain 5500 m high. Calculate the change in her gravitational potential energy.

8 **a** A toy car works by means of a stretched rubber band. What form of potential energy does the car store when the band is stretched?

b A bar magnet is lying with its north pole next to the south pole of another bar magnet. A student pulls them apart. Why do we say that the magnets' potential energy has increased? Where has this energy come from?

5.3 Kinetic energy

As well as lifting an object, a force can make it accelerate. Again, work is done by the force and energy is transferred to the object. In this case, we say that it has gained kinetic energy, E_k. The faster an object is moving, the greater its kinetic energy (k.e.).

For an object of mass *m* travelling at a speed *v*, we have:

$$\text{kinetic energy} = \frac{1}{2} \times \text{mass} \times \text{speed}^2$$
$$E_k = \frac{1}{2}mv^2$$

Deriving the formula for kinetic energy

KEY EQUATION

$$\text{kinetic energy} = \frac{1}{2} \times \text{mass} \times \text{speed}^2$$
$$E_k = \frac{1}{2}mv^2$$

You must learn how to derive this equation.

The equation for kinetic energy, $E_k = \frac{1}{2}mv^2$ is related to one of the equations of motion. We imagine a car being accelerated from rest ($u = 0$) to velocity v. To give it acceleration a, it is pushed by a force F for a distance s. Since $u = 0$, we can write the equation $v^2 = u^2 + 2as$ as:

$$v^2 = 2as$$

Multiplying both sides by $\frac{1}{2}m$ gives:

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

Now, *ma* is the force *F* accelerating the car, and *mas* is the force × the distance it moves (that is, the work done by the force). So we have:

$$\frac{1}{2}mv^2 = \text{work done by force } F$$

This is the energy transferred to the car, and hence its kinetic energy.

WORKED EXAMPLE

3 Calculate the increase in kinetic energy of a car of mass 800 kg when it accelerates from 20 m s^{-1} to 30 m s^{-1}.

Step 1 Calculate the initial k.e. of the car:

$$E_k = \frac{1}{2}mv^2$$
$$= \frac{1}{2} \times 800 \times (20)^2$$
$$= 160\,000\text{ J} \equiv 160\text{kJ}$$

Step 2 Calculate the final k.e. of the car:

$$E_k = \tfrac{1}{2}mv^2$$
$$= \tfrac{1}{2} \times 800 \times (30)^2$$
$$= 360\,000\text{ J} \equiv 360\text{kJ}$$

Step 3 Calculate the change in the car's k.e.:

change in k.e. = 360 − 160 = 200 kJ

Hint: *Take care! You can't calculate the change in k.e. by squaring the change in speed. In this example, the change in speed is 10 m s*$^{-1}$*, and this would give an incorrect value for the change in k.e.*

Questions

9 Which has more k.e., a car of mass 500 kg travelling at 15 m s^{-1} or a motorcycle of mass 250 kg travelling at 30 m s^{-1}?

10 Calculate the change in kinetic energy of a ball of mass 200 g when it bounces. Assume that it hits the ground with a speed of 15.8 m s^{-1} and leaves it at 12.2 m s^{-1}.

5.4 Gravitational potential to kinetic energy transformations

A motor drags the roller-coaster car to the top of the first hill. The car runs down the other side, picking up speed as it goes (see Figure 5.11). It is moving just fast enough to reach the top of the second hill, slightly lower than the first. It accelerates downhill again. Everybody screams!

Figure 5.11: The roller-coaster car accelerates as it comes downhill. It's even more exciting if it runs through water.

The motor provides a force to pull the roller-coaster car to the top of the hill. It transfers energy to the car. But where is this energy when the car is waiting at the top of the hill? The car now has gravitational potential energy; as soon as it is given a small push to set it moving, it accelerates. It gains kinetic energy and at the same time it loses g.p.e.

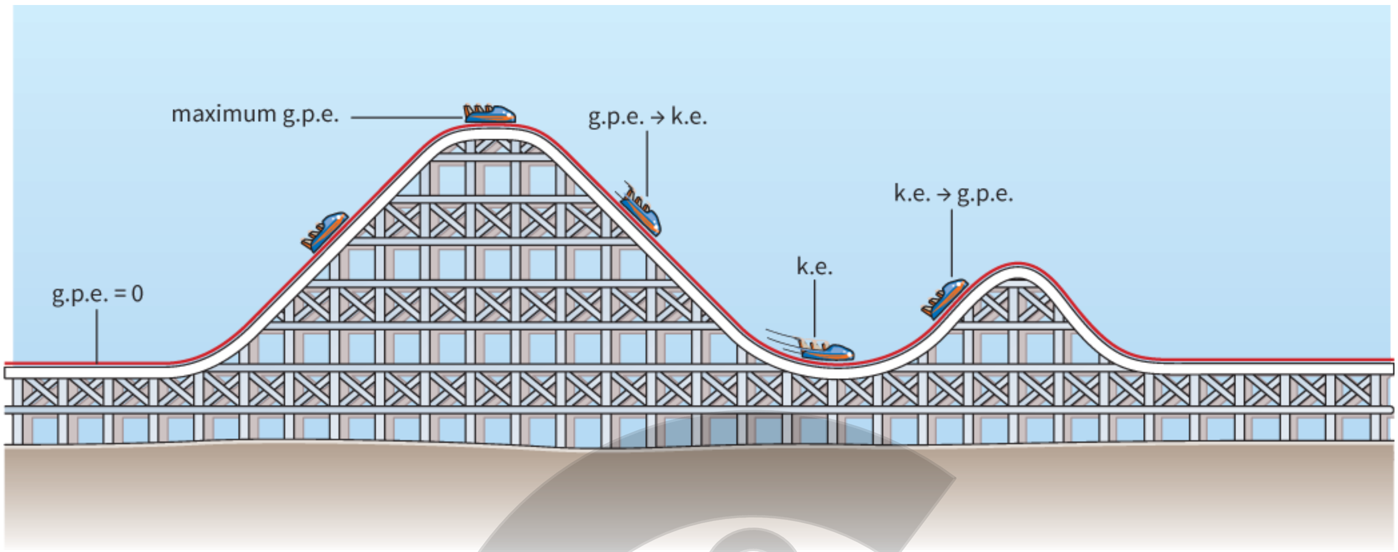


Figure 5.12: Energy changes along a roller-coaster.

As the car runs along the roller-coaster track (Figure 5.12), its energy changes.

1 At the top of the first hill, it has the most g.p.e.

2 As it runs downhill, its g.p.e. decreases and its k.e. increases.

3 At the bottom of the hill, all of its g.p.e. has been changed to k.e. and heat and sound energy.

4 As it runs back uphill, the force of gravity slows it down. k.e. is being changed to g.p.e.

Inevitably, some energy is lost by the car. There is friction with the track and air resistance. So, the car cannot return to its original height. That is why the second hill must be slightly lower than the first. It is fun if the car runs through a trough of water, but that takes even more energy, and the car cannot rise so high. There are many situations where an object's energy changes between gravitational potential energy and kinetic energy. For example:

- a high diver falling towards the water – g.p.e. changes to k.e.
- a ball is thrown upwards – k.e. changes to g.p.e.
- a child on a swing – energy changes back and forth between g.p.e. and k.e.

5.5 Down, up, down: energy changes

When an object falls, it speeds up. Its g.p.e. decreases and its k.e. increases. Energy is being transformed from gravitational potential energy to kinetic energy. Some energy is likely to be lost, usually as heat because of air resistance. However, if no energy is lost in the process, we have:

decrease in g.p.e. = gain in k.e.

We can use this idea to solve a variety of problems, as illustrated by Worked example 4.

WORKED EXAMPLE

4 A pendulum consists of a brass sphere of mass 5.0 kg hanging from a long string (see Figure 5.13).

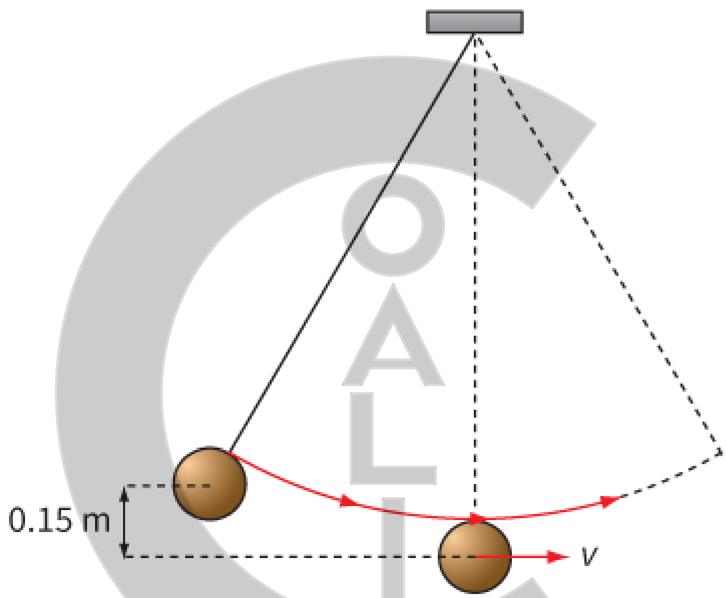


Figure 5.13: For Worked example 4.

The sphere is pulled to the side so that it is 0.15 m above its lowest position. It is then released. How fast will it be moving when it passes through the lowest point along its path?

Step 1 Calculate the loss in g.p.e. as the sphere falls from its highest position:

$E_p = mgh = 5.0 \times 9.81 \times 0.15 = 7.36$ J

Step 2 The gain in the sphere's k.e. is 7.36 J. We can use this to calculate the sphere's speed. First, calculate v^2, then v:

$$\frac{1}{2}mv^2 = 7.36$$

$$\frac{1}{2} \times 5.0 \times v^2 = 7.36$$

$$v^2 = 2 \times \frac{7.36}{5.0}$$

$$v^2 = 2.944$$

$$v = \sqrt{2.944} \approx 1.72 \text{ m s}^{-1} \approx 1.7 \text{ m s}^{-1}$$

Note that we would obtain the same result in Worked example 4 no matter what the mass of the sphere. This is because both k.e. and g.p.e. depend on mass *m*. If we write:

$$\text{change in g.p.e} = \text{change in k.e}$$

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

we can cancel *m* from both sides. Hence:

$$gh = \frac{v^2}{2}$$

$$v^2 = 2gh$$

$$v = \sqrt{2gh}$$

The final speed *v* only depends on *g* and *h*. The mass *m* of the object is irrelevant. This is not surprising; we could use the same equation to calculate the speed of an object falling from height *h*. An object of small mass gains the same speed as an object of large mass, provided air resistance has no effect.

Questions

11 Re-work Worked example 4 for a brass sphere of mass 10 kg, and show that you get the same result. Repeat with any other value of mass.

12 Calculate how much gravitational potential energy is lost by an aircraft of mass 80 000 kg if it descends from an altitude of 10 000 m to an altitude of 1000 m. What happens to this energy if the pilot keeps the aircraft's speed constant?

13 A high diver (see Figure 5.14) reaches the highest point in her jump with her centre of gravity 10 m above the water.

Figure 5.14: A high dive is an example of converting (transforming) gravitational potential energy to kinetic energy.

Assuming that all her gravitational potential energy becomes kinetic energy during the dive, calculate her speed just before she enters the water.

5.6 Energy transfers

Climbing bars

If you are going to climb a mountain, you will need a supply of energy. This is because your gravitational potential energy is greater at the top of the mountain than at the base. A good supply of energy would be some bars of chocolate. Each bar supplies 1200 kJ. Suppose your weight is 600 N and you climb a 2000 m high mountain. The work done by your muscles is:

$$\text{work done} = Fs = 600 \times 2000 = 1200 \text{ kJ}$$

So, one bar of chocolate should provide enough energy. Of course, in reality, it would not. Your body is inefficient. It cannot convert 100% of the energy from food into gravitational potential energy. A lot of energy is wasted as your muscles warm up, you perspire and your body rises and falls as you walk along the path. Your body is perhaps only 5% efficient as far as climbing is concerned, and you will need to eat 20 chocolate bars to get you to the top of the mountain. And you will need to eat more to get you back down again.

KEY EQUATION

$$\text{efficiency} = \frac{\text{useful output energy}}{\text{total input energy}} \times 100\%$$

Many energy transfers are inefficient. That is, only part of the energy is transferred to where it is wanted. The rest is wasted, and appears in some form that is not wanted (such as waste heat) or in the wrong place. You can determine the efficiency of any device or system using the following equation:

$$\text{efficiency} = \frac{\text{useful output energy}}{\text{total input energy}} \times 100\%$$

A car engine is more efficient than a human body, but not much more. Figure 5.15 shows how this can be represented by a Sankey diagram. The width of the arrow represents the fraction of the energy which is transformed to each new form. In the case of a car engine, we want it to provide kinetic energy to turn the wheels. In practice, 80% of the energy is transformed into heat: the engine gets hot, and heat escapes into the surroundings. So the car engine is only 20% efficient.

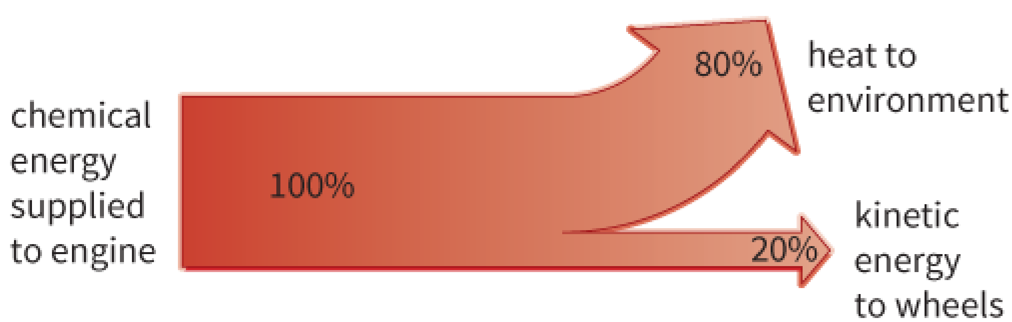


Figure 5.15: We want a car engine to supply kinetic energy. This Sankey diagram shows that only 20% of the energy supplied to the engine ends up as kinetic energy – it is 20% efficient.

We have previously considered situations where an object is falling, and all of its gravitational potential energy changes to kinetic energy.

In Worked example 5, we will look at a similar situation, but in this case the energy change is not 100% efficient.

WORKED EXAMPLE

5 Figure 5.16 shows a dam that stores water. The outlet of the dam is 20 m below the surface of the water in the reservoir. Water leaving the dam is moving at 16 m s^{-1}. Calculate the percentage of the gravitational potential energy that is lost when converted into kinetic energy.

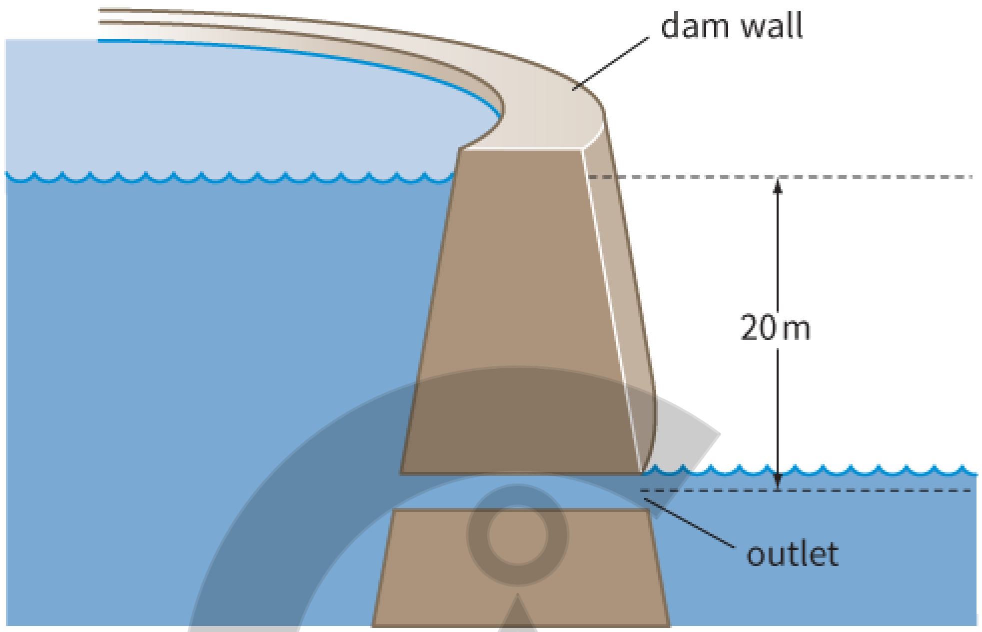


Figure 5.16: Water stored behind the dam has gravitational potential energy; the fast-flowing water leaving the foot of the dam has kinetic energy.

Step 1 We will picture 1 kg of water, starting at the surface of the lake (where it has g.p.e., but no k.e.) and flowing downwards and out at the foot (where it has k.e., but less g.p.e.). Then:

change in g.p.e. of water between surface and outflow = mgh = 1 × 9.81 × 20 = 196 J

Step 2 Calculate the k.e. of 1 kg of water as it leaves the dam:

$$\text{k.e. of water leaving dam} = \frac{1}{2}mv^2 = \frac{1}{2} \times 1 \times (16)^2 = 128\text{ J}$$

Step 3 For each kilogram of water flowing out of the dam, the loss of energy is:

loss = 196 − 128 = 68 J

$$\text{percentage loss} = \frac{68}{196} \times 100\% = 34.69\% \approx 35\%$$

If you wanted to use this moving water to generate electricity, you would have already lost more than a third of the energy that it stores when it is behind the dam.

Conservation of energy

Where does the lost energy from the water in the reservoir go? Most of it ends up warming the water, or warming the pipes that the water flows through. The outflow of water is probably noisy, so some sound is produced.

Here, we are assuming that all of the energy ends up somewhere. None of it disappears. We assume the same thing when we draw a Sankey diagram. The total thickness of the arrow remains constant. We could not have an arrow which got thinner (energy disappearing) or thicker (energy appearing out of nowhere).

We are assuming that energy is conserved. This is a principle, known as the **principle of conservation of energy**, which we expect to apply in all situations.

Energy cannot be created or destroyed. It can only be converted from one form to another.

We should always be able to add up the total amount of energy at the beginning, and be able to account for it all at the end. We cannot be sure that this is always the case, but we expect it to hold true.

We have to think about energy changes within a closed system; that is, we have to draw an imaginary boundary around all of the interacting objects that are involved in an energy transfer.

Sometimes, applying the principle of conservation of energy can seem like a scientific fiddle. When physicists were investigating radioactive decay involving beta particles, they found that the particles after the decay had less energy in total than the particles before. They guessed that there was another, invisible particle that was carrying away the missing energy. This particle, named the neutrino, was proposed by the theoretical physicist Wolfgang Pauli in 1931. The neutrino was not detected by experimenters until 25 years later.

Although we cannot prove that energy is always conserved, this example shows that the principle of conservation of energy can be a powerful tool in helping us to understand what is going on in nature, and that it can help us to make fruitful predictions about future experiments.

Question

14 A stone falls from the top of a cliff, 80 m high. When it reaches the foot of the cliff, its speed is 38 m s^{-1}.

- **a** Calculate the proportion of the stone's initial g.p.e. that is converted to k.e.
- **b** What happens to the rest of the stone's initial energy?

5.7 Power

The word **power** has several different meanings – such as political power, powers of ten or electrical power from power stations. In physics, it has a specific meaning related to these other meanings. Figure 5.17 illustrates what we mean by power in physics.

Figure 5.17: A lift needs a powerful motor to raise the car when it has a full load of people. The motor does many thousands of joules of work each second.

The lift shown in Figure 5.17 can lift a heavy load of people. The motor at the top of the building provides a force to raise the lift car, and this force does work against the force of gravity. The motor transfers energy to the lift car. The **power** *P* of the motor is the rate at which it does work over a unit of time.

Power is defined as the rate of work done per unit of time. As a word equation, power is given by:

$$\text{power} = \frac{\text{work done}}{\text{time taken}}$$
$$P = \frac{W}{t}$$

where *W* is the work done in a time *t*.

KEY EQUATION

$$\text{power} = \frac{\text{work done}}{\text{time taken}} \equiv P = \frac{W}{t}$$

Units of power: the watt

Power is measured in watts, named after James Watt, the Scottish engineer famous for his development of the steam engine in the second half of the 18th century. The **watt** is defined as a rate of working of 1 joule per second. Hence:

1 watt = 1 joule per second

or

$$1\ \text{W} = 1\ \text{J s}^{-1}$$

In practice, we also use kilowatts (kW) and megawatts (MW).

1000 watts = 1 kilowatt (1 kW)

1 000 000 watts = 1 megawatt (1 MW)

The labels on light bulbs display their power in watts; for example, 60 W or 10 W. The values of power on the labels tell you about the energy transferred by an electrical current, rather than by a force doing work.

WORKED EXAMPLE

6 The motor of the lift shown in Figure 5.18 provides a force of 20 kN; this force is enough to raise the lift by 18 m in 10 s. Calculate the output power of the motor.

Step 1 First, we must calculate the work done:

$$\text{work done} = \text{force} \times \text{distance moved}$$
$$= 20 \times 18 = 360\ \text{kJ}$$

Step 2 Now we can calculate the motor's output power:

$$\text{power} = \frac{\text{work done}}{\text{time taken}} = \frac{360 \times 10^3}{10} = 36 \text{ kW}$$

Hint: *Take care not to confuse the two uses of the letter 'W':*

W = *watt (a unit)*

W = work done (a quantity)

So the lift motor's power is 36 kW. Note that this is its mechanical power output. The motor cannot be 100% efficient since some energy is bound to be wasted as heat due to friction, so the electrical power input must be more than 36 kW.

Questions

15 Calculate how much work is done by a 50 kW car engine in a time of 1.0 minute.

16 A car engine does 4200 kJ of work in one minute. Calculate its output power, in kilowatts.

17 A particular car engine provides a force of 700 N when the car is moving at its top speed of 40 m s^{-1}.

a Calculate how much work is done by the car's engine in one second.

b State the output power of the engine.

Moving power

An aircraft is kept moving forwards by the force of its engines pushing air backwards. The greater the force and the faster the aircraft is moving, the greater the power supplied by its engines.

Suppose that an aircraft is moving with velocity *v*. Its engines provide the force *F* needed to overcome the drag of the air. In time *t*, the aircraft moves a distance *s* equal to *v* × *t*.

So, the work done by the engines is:

$$\text{work done} = \text{force} \times \text{distance}$$
$$W = F \times v \times t$$

We know that:

$$\text{power} = \frac{\text{work done}}{\text{time taken}}$$
$$P = \frac{W}{t}$$

Substituting *W* for gives:

$$P = \frac{F \times v \times t}{t}$$

Which can be simplified to:

$$P = F \times v$$
$$\text{power} = \text{force} \times \text{velocity}$$

KEY EQUATION

power = force × velocity ≡ $P = F \times v$

It may help to think of this equation in terms of units. The right-hand side is in N × m s^{-1}, and N m is the same as J. So the right-hand side has units of J s^{-1}, or W, the unit of power. If you look back to Question 17, you will see that, to find the power of the car engine, rather than considering the work done in 1 s, we could simply have multiplied the engine's force by the car's speed.

Human power

Our energy supply comes from our food. A typical diet supplies 2000–3000 kcal (kilocalories) per day. This is equivalent (in SI units) to about 10 MJ of energy. We need this energy for our daily requirements – keeping warm, moving about, brainwork and so on. We can determine the average power of all the activities of our body:

$$\begin{aligned}\text{average power} &= 10\ \text{MJ per day}\\ &= 10 \times \frac{10^6}{86\,400}\\ &= 116\ \text{W}\end{aligned}$$

So we dissipate energy at the rate of about 100 W. We supply roughly as much energy to our surroundings as a 100 W light bulb. Twenty people will keep a room as warm as a 2 kW electric heater.

Note that this is our average power. If you are doing some demanding physical task, your power will be greater. This is illustrated in Worked example 7.

Note also that the human body is not a perfectly efficient system; a lot of energy is wasted when, for example, we lift a heavy load. We might increase an object's g.p.e. by 1000 J when we lift it, but this might require five or ten times this amount of energy to be expended by our bodies.

WORKED EXAMPLE

7 A person who weighs 500 N runs up a flight of stairs in 5.0 s (Figure 5.18). Their gain in height is 3.0 m. Calculate the rate at which work is done against the force of gravity.

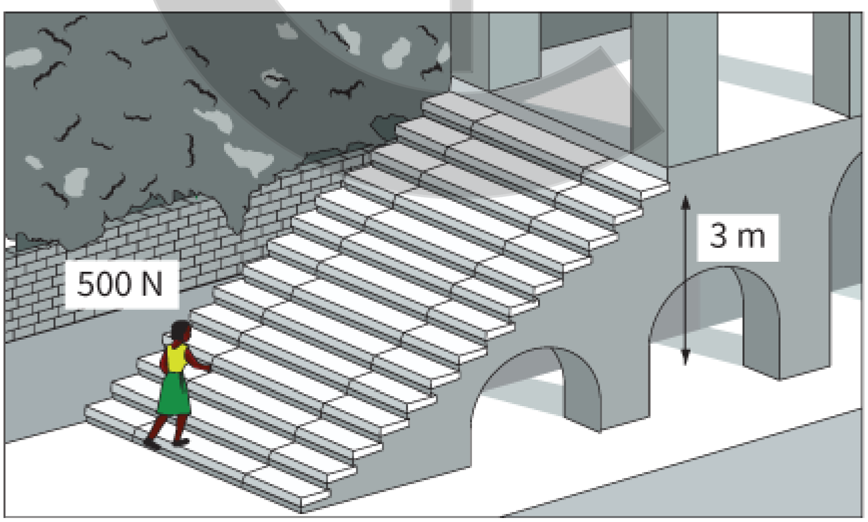


Figure 5.18: Running up stairs can require a high rate of doing work. You may have investigated your own power in this way.

Step 1 Calculate the work done against gravity:

$$\text{work done } W = F \times s = 500 \times 3.0 = 1500 \text{ J}$$

Step 2 Now calculate the power:

$$\text{power } P = \frac{W}{t} = \frac{1500}{5.0} = 300 \text{ W}$$

So, while the person is running up the stairs, they are doing work against gravity at a greater rate than their average power – perhaps three times as great. And, since our muscles are not very efficient, they need to be supplied with energy even faster, perhaps at a rate of 1 kW. This is why we cannot run up stairs all day long without greatly increasing the amount we eat. The inefficiency of our muscles also explains why we get hot when we exert ourselves.

Question

18 In an experiment to measure a student's power, she times herself running up a flight of steps. Use the data to work out her useful power.

number of steps = 28

height of each step = 20 cm

acceleration of free fall = 9.81 m s^{-2}

mass of student = 55 kg

time taken = 5.4 s

REFLECTION

How do you feel about this topic? What parts of it do you particularly like or dislike? And why?

Think about a number of important machines that you use in your house or school. Is it worthwhile increasing their efficiency and can you suggest how this might be done? Discuss this with others.

Make notes about the new things you have learnt from this chapter.

SUMMARY

The work done W when a force F moves through a displacement s in the direction of the force:

$$W = Fs \quad \text{or} \quad W = Fs\cos\theta$$

where θ is the angle between the force and the displacement.

A joule is defined as the work done (or energy transferred) when a force of 1 N moves a distance of 1 m in the direction of the force.

When an object of mass m rises through a height h, its gravitational potential energy E_p increases by an amount:

$$E_p = mgh$$

The kinetic energy E_k of a body of mass m moving at speed v is:

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

The principle of conservation of energy states that, for a closed system, energy can be transferred to other forms but the total amount of energy remains constant.

The efficiency of a device or system is determined using the equation:

$$\text{efficiency} = \frac{\text{useful output energy}}{\text{total input energy}} \times 100\%$$

Power is the rate at which work is done (or energy is transferred):

$$P = \frac{W}{t} \quad \text{and} \quad P = Fv$$

A watt is defined as a rate of transfer of energy of one joule per second.

EXAM-STYLE QUESTIONS

1 How is the joule related to the base units of m, kg and s? **[1]**

A kg m^{-1} s^2

B kg m^2 s^{-2}

C kg m^2 s^{-1}

D kg s^{-2}

2 An object falls at terminal velocity in air. What overall conversion of energy is occurring? **[1]**

A gravitational potential energy to kinetic energy

B gravitational potential energy to thermal energy

C kinetic energy to gravitational potential energy

D kinetic energy to thermal energy

3 In each case a–c, describe the energy changes taking place:

a An apple falling towards the ground **[1]**

b A car decelerating when the brakes are applied **[1]**

c A space probe falling towards the surface of a planet. **[1]**

[Total: 3]

4 A 120 kg crate is dragged along the horizontal ground by a 200 N force acting at an angle of 30° to the horizontal, as shown.

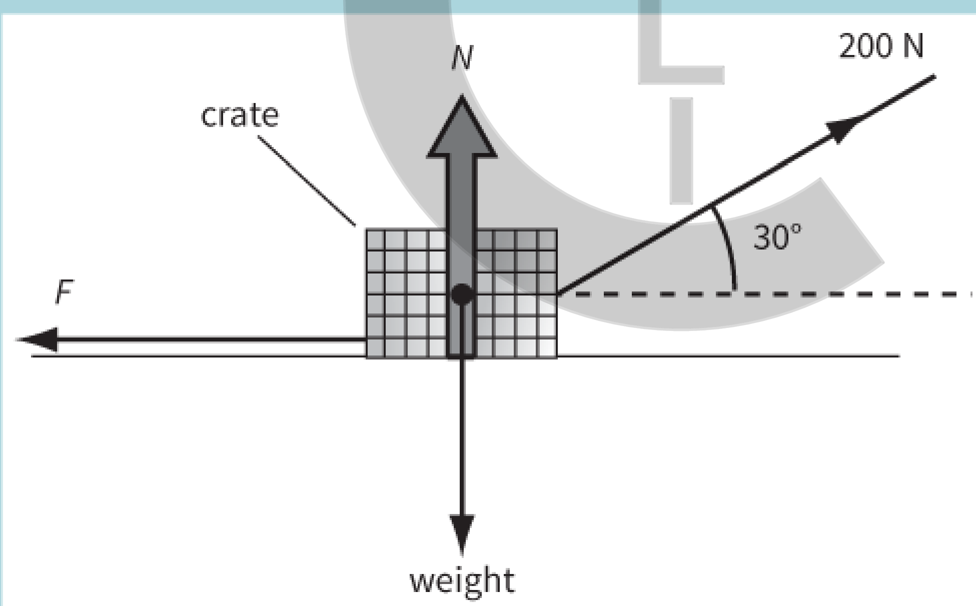


Figure 5.19

The crate moves along the surface with a constant velocity of 0.5 m s^{-1}. The 200 N force is applied for a time of 16 s.

a Calculate the work done on the crate by:

i the 200 N force **[3]**

ii the weight of the crate **[2]**

iii the normal contact force N. **[2]**

b Calculate the rate of work done against the frictional force F. **[1]**

[Total: 8]

5 Explain which of the following has greater kinetic energy?

- A 20-tonne truck travelling at a speed of 30 m s^{-1}
- A 1.2 g dust particle travelling at 150 km s^{-1} through space. **[3]**

6 A 950 kg sack of cement is lifted to the top of a building 50 m high by an electric motor.

a Calculate the increase in the gravitational potential energy of the sack of cement. **[2]**

b The output power of the motor is 4.0 kW. Calculate how long it took to raise the sack to the top of the building. **[2]**

c The electrical power transferred by the motor is 6.9 kW. In raising the sack to the top of the building, how much energy is wasted in the motor as heat? **[3]**

[Total: 7]

7 a Define power and state its unit. **[2]**

b Write a word equation for the kinetic energy of a moving object. **[1]**

c A car of mass 1100 kg starting from rest reaches a speed of 18 m s^{-1} in 25 s. Calculate the average power developed by the engine of the car. **[2]**

[Total: 5]

8 A cyclist pedals a long slope which is at 5.0° to the horizontal, as shown.

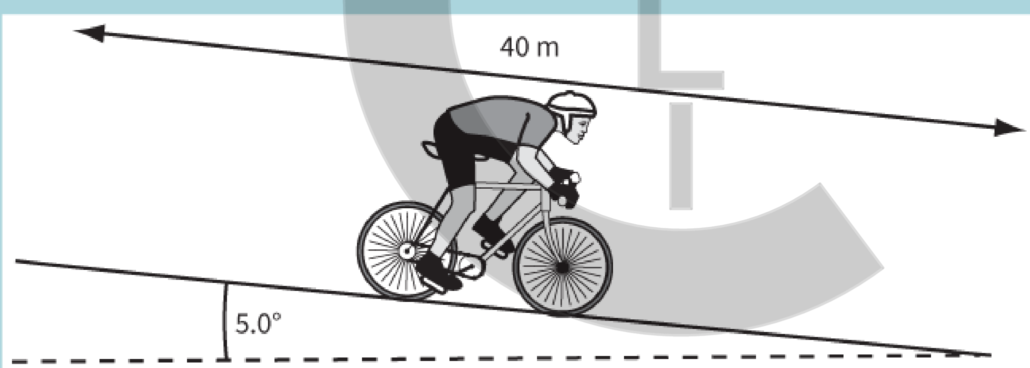


Figure 5.20

The cyclist starts from rest at the top of the slope and reaches a speed of 12 m s^{-1} after a time of 67 s, having travelled 40 m down the slope. The total mass of the cyclist and bicycle is 90 kg.

a Calculate:

i the loss in gravitational potential energy as he travels down the slope **[3]**

ii the increase in kinetic energy as he travels down the slope. **[2]**

b i Use your answers to a to determine the useful power output of the cyclist. **[3]**

ii Suggest one reason why the actual power output of the cyclist is larger than your value in **i**. **[2]**

[Total: 10]

9 a Explain what is meant by work. **[2]**

[2]

b **i** Explain how the principle of conservation of energy applies to a man sliding from rest down a vertical pole, if there is a constant force of friction acting on him.

ii The man slides down the pole and reaches the ground after falling a distance h = 15 m. His potential energy at the top of the pole is 1000 J. Sketch a graph to show how his gravitational potential energy E_p varies with h. Add to your graph a line to show the variation of his kinetic energy E_k with h. **[3]**

[Total: 7]

10 **a** Use the equations of motion to show that the kinetic energy of an object of mass m moving with velocity v is $\frac{1}{2}mv^2$ **[2]**

b A car of mass 800 kg accelerates from rest to a speed of 20 m s^{-1} in a time of 6.0 s.

i Calculate the average power used to accelerate the car in the first 6.0 s. **[2]**

ii The power passed by the engine of the car to the wheels is constant. Explain why the acceleration of the car decreases as the car accelerates. **[2]**

[Total: 6]

11 **a** **i** Define potential energy. **[1]**

ii **Identify** differences between gravitational potential energy and elastic potential energy. **[2]**

b Seawater is trapped behind a dam at high tide and then released through turbines. The level of the water trapped by the dam falls 10.0 m until it is all at the same height as the sea.

i Calculate the mass of seawater covering an area of 1.4×10^6 m^2 and with a depth of 10.0 m. (Density of seawater = 1030 kg m^{-3}.) **[1]**

ii Calculate the maximum loss of potential energy of the seawater in **i** when passed through the turbines. **[2]**

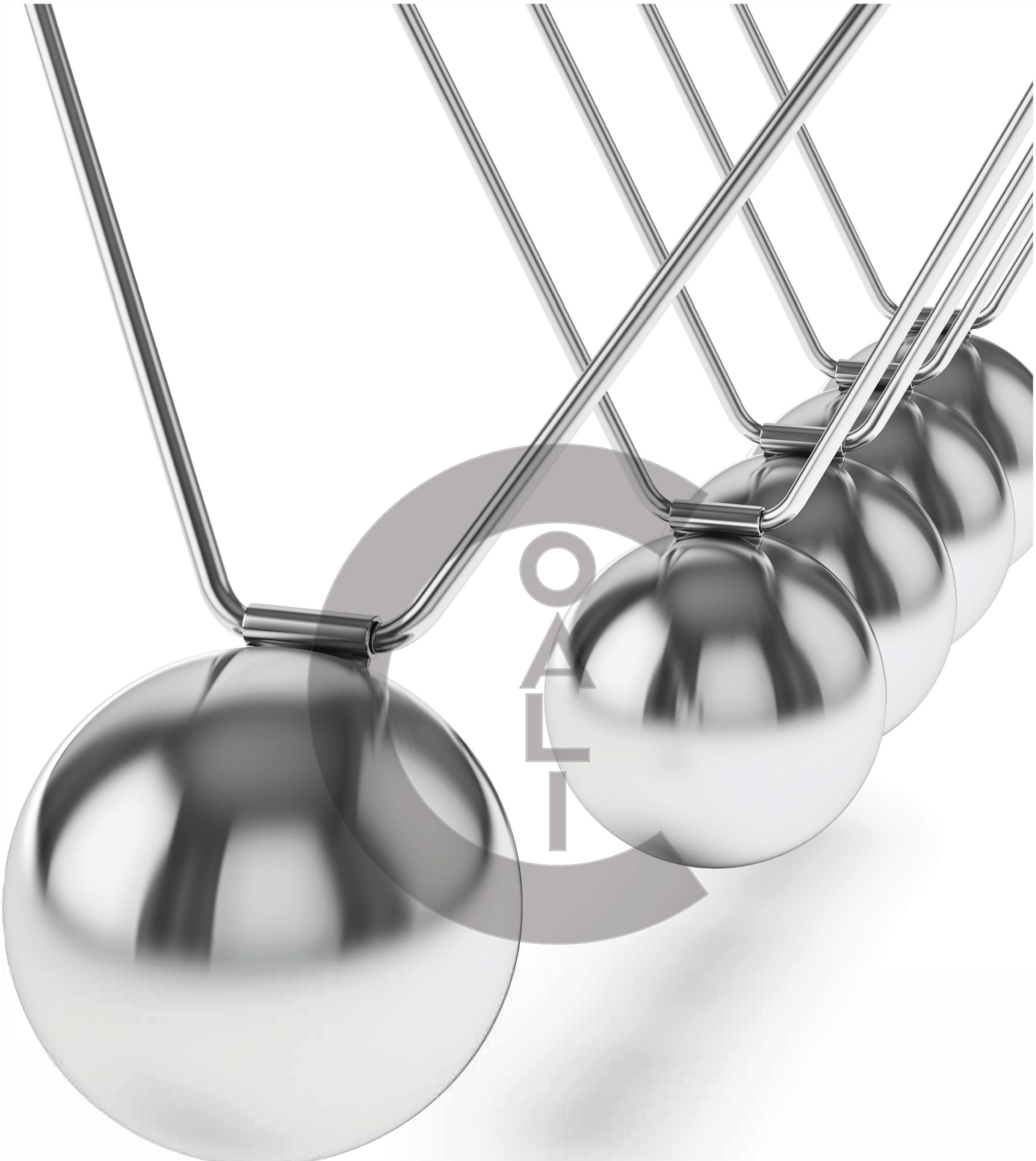
iii The potential energy of the seawater, calculated in **ii**, is lost over a period of 6.0 hours. Estimate the average power output of the power station over this time period, given that the efficiency of the power station is 50%. **[3]**

[Total: 9]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the concept of work, and recall and use work done = force × displacement in the direction of the force	5.1			
recall and apply the principle of conservation of energy	5.6			
recall and understand that the efficiency of a system is the ratio of useful energy output from the system to the total energy input	5.6			
use the concept of efficiency to solve problems	5.6			
define power as work done per unit time and solve problems using $P = \frac{W}{t}$	5.7			
derive $P = Fv$ and use it to solve problems	5.7			
derive, using $W = Fs$, the formula $\Delta E_p = mg\Delta h$	5.2			
recall and use the formula $\Delta E_p = mg\Delta h$	5.5			
derive, using the equations of motion, the formula $E_k = \frac{1}{2}mv^2$ and recall and use the formula.	5.3			



〉Chapter 6

Momentum

LEARNING INTENTIONS

In this chapter you will learn how to:

- define and use linear momentum
- state and apply the principle of conservation of momentum to collisions in one and two dimensions
- relate force to the rate of change of momentum and state Newton's second law of motion
- recall that, for a perfectly elastic collision, the relative speed of approach is equal to the relative speed of separation
- discuss energy changes in perfectly elastic and inelastic collisions.

BEFORE YOU START

- What do you understand about Newton's laws? Write down all three of them in your own words. Define any of the quantities mentioned in the laws.
- If you blow up a balloon and then let it go without tying the end, why does the balloon fly around in the air?

UNDERSTANDING COLLISIONS

To improve the safety of cars, the motion of a car during a crash must be understood and the forces on the driver minimised (Figure 6.1). In this way, safer cars have been developed and many lives have been saved. Find out about as many safety features of cars as you can and discuss with someone else why these features improve safety in a crash.

In this chapter, we will explore how the idea of momentum can allow us to predict how objects move after colliding (interacting) with each other. We will also see how Newton's laws of motion can be expressed in terms of momentum.

Figure 6.1: A high-speed photograph of a crash test. The cars collide head-on at 15 m s^{-1} with dummies as drivers.

6.1 The idea of momentum

Snooker players can perform some amazing moves on the table, without necessarily knowing Newton's laws of motion – see Figure 6.2.

Figure 6.2: If you play snooker often enough, you will be able to predict how the balls will move on the table. Alternatively, you can use the laws of physics to predict their motion.

However, the laws of physics can help us to understand what happens when two snooker balls collide or when one bounces off the side cushion of the table.

Here are some examples of situations involving collisions:

- Two cars collide head-on.
- A fast-moving car runs into the back of a slower car in front.
- A footballer runs into an opponent.
- A hockey stick strikes a ball.
- A comet or an asteroid collides with a planet as it orbits the Sun.
- The atoms of the air collide constantly with each other, and with the walls of their surroundings.
- Electrons that form an electric current collide with the vibrating ions that make up a metal wire.
- Two distant galaxies collide over millions of years.

From these examples, we can see that collisions are happening all around us, all the time. They happen on the microscopic scale of atoms and electrons, they happen in our everyday world, and they also happen on the cosmic scale of our Universe.

6.2 Modelling collisions

Springy collisions

Figure 6.3a shows what happens when one snooker ball collides head-on with a second, stationary ball. The result can seem surprising. The moving ball stops dead. The ball initially at rest moves off with the same velocity as that of the original ball. To achieve this, a snooker player must observe two conditions:

- The collision must be head-on. (If one ball strikes a glancing blow on the side of the other, they will both move off at different angles.)
- The moving ball must not be given any spin. (Spin is an added complication that we will ignore in our present study, although it plays a vital part in the games of pool and snooker.)

You can mimic the collision of two snooker balls in the laboratory using two identical trolleys, as shown in Figure 6.3b. The moving trolley has its spring-load released, so that the collision is springy. As one trolley runs into the other, the spring is at first compressed, and then it pushes out again to set the second trolley moving. The first trolley comes to a complete halt. The 'motion' of one trolley has transferred to the other.

You can see another interesting result if two moving identical trolleys collide head-on. If the collision is springy, both trolleys bounce backwards. If a fast-moving trolley collides with a slower one, the fast trolley bounces back at the speed of the slow one, and the slow one bounces back at the speed of the fast one. In this collision, it is as if the velocities of the trolleys have been swapped.

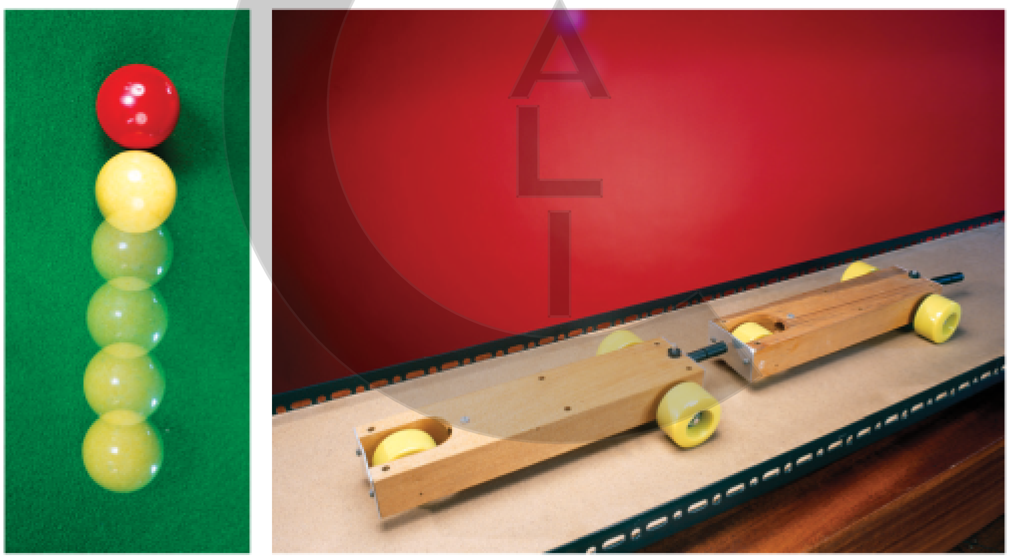


Figure 6.3: **a** The red snooker ball, coming from the left, has hit the yellow ball head-on. **b** You can do the same thing with two trolleys in the laboratory.

Sticky collisions

Figure 6.4 shows another type of collision. In this case, the trolleys have adhesive pads so that they stick together when they collide. A sticky collision like this is the opposite of a springy collision like the ones described previously.

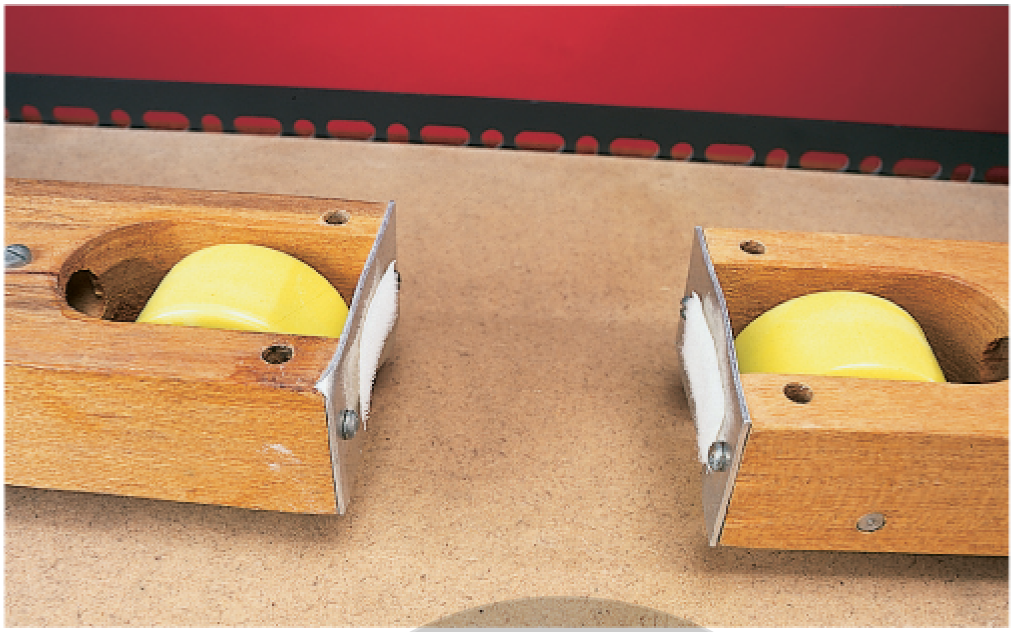


Figure 6.4: If a moving trolley sticks to a stationary trolley, they both move off together.

If a single moving trolley collides with an identical stationary one, they both move off together. After the collision, the speed of the combined trolleys is half that of the original trolley. It is as if the 'motion' of the original trolley has been shared between the two. If a single moving trolley collides with a stationary double trolley (twice the mass), they move off with one-third of the original velocity.

From these examples of sticky collisions, you can see that, when the mass of the trolley increases as a result of a collision, its velocity decreases. Doubling the mass halves the velocity, and so on.

Question

1 a Ball A, moving towards the right, collides with stationary ball B. Ball A bounces back; ball B moves off slowly to the right. Which has the greater mass, ball A or ball B?

b Trolley A, moving towards the right, collides with stationary trolley B. They stick together, and move off at less than half A's original speed. Which has the greater mass, trolley A or trolley B?

Defining linear momentum

From the examples discussed earlier, we can see that two quantities are important in understanding collisions:

- the mass m of the object
- the velocity v of the object.

These are combined to give a single quantity, called the **linear momentum** (or simply momentum) p of an object.

KEY EQUATION

$$\text{momentum} = \text{mass} \times \text{velocity}$$
$$p = mv$$

The momentum of an object is defined as the product of the mass of the object and its velocity. Hence:

$$\text{momentum} = \text{mass} \times \text{velocity}$$
$$p = mv$$

The SI unit of momentum is kg m s^{-1}. There is no special name for this unit in the SI system. The newton second (N s) can also be used as a unit of momentum (see topic 6.7).

Momentum is a vector quantity because it is a product of a vector quantity (velocity) and a scalar quantity (mass). Momentum has both magnitude and direction. Its direction is the same as the direction of the object's velocity.

In the earlier examples, we described how the 'motion' of one trolley appeared to be transferred to a second trolley, or shared with it. It is more correct to say that it is the trolley's momentum that is transferred or shared. (More precisely, we should refer to linear momentum, because there is another quantity called angular momentum that is possessed by spinning objects.)

As with energy, we find that momentum is also conserved. We have to consider objects that form a **closed system**–that is, no resultant external force acts on them. The principle of **conservation of momentum** states that, within a closed system, the total momentum in any direction is constant.

The principle of conservation of momentum can also be expressed as follows:

For a closed system where no resultant external force acts, in any direction:

total momentum of objects before collision = total momentum of objects after collision

A group of colliding objects always has as much momentum after the collision as it had before the collision. This principle is illustrated in Worked example 1.

WORKED EXAMPLE

1 In Figure 6.5, trolley A of mass 0.80 kg travelling at a velocity of 3.0 m s^{-1} collides head-on with a stationary trolley B. Trolley B has twice the mass of trolley A. The trolleys stick together and have a common velocity of 1.0 m s^{-1} after the collision. Show that momentum is conserved in this collision.

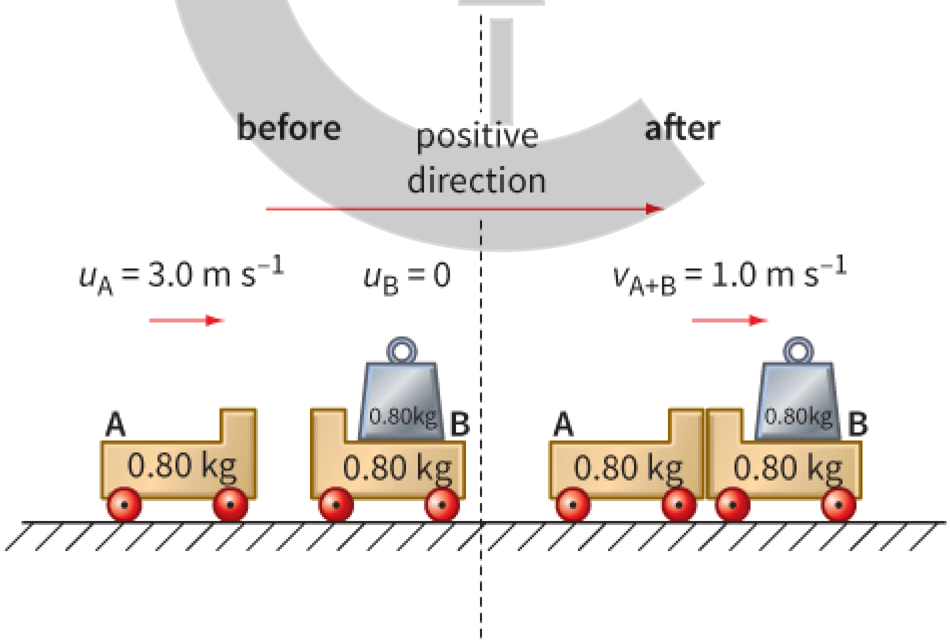


Figure 6.5: The state of trolleys A and B, before and after the collision.

Step 1 Make a sketch using the information given in the question. Notice that we need two diagrams to show the situations, one before and one after the collision. Similarly, we need two calculations –

one for the momentum of the trolleys before the collision and one for their momentum after the collision.

Step 2 Calculate the momentum before the collision:

momentum of trolleys before collision

$= m_A \times u_A + m_B \times u_B$

$= (0.80 \times 3.0) + 0$

$= 2.4 \text{ kg m s}^{-1}$

Trolley B has no momentum before the collision, because it is not moving.

Step 3 Calculate the momentum after the collision:

momentum of trolleys after collision

$= (m_A + m_B) \times v_{A+B}$

$= (0.80 + 1.60) \times 1.0$

$= 2.4 \text{ kg m s}^{-1}$

So, both before and after the collision, the trolleys have a combined momentum of 2.4 kg m s^{-1}. Momentum has been conserved.

Questions

2 Calculate the momentum of each of the following objects:

- **a** a 0.50 kg stone travelling at a velocity of 20 m s^{-1}
- **b** a 25 000 kg bus travelling at 20 m s^{-1} on a road
- **c** an electron travelling at $2.0 \times 10^7 \text{ m s}^{-1}$.

  (The mass of the electron is 9.1×10^{-31} kg.)

3 Two balls, each of mass 0.50 kg, collide as shown in Figure 6.6. Show that their total momentum before the collision is equal to their total momentum after the collision.

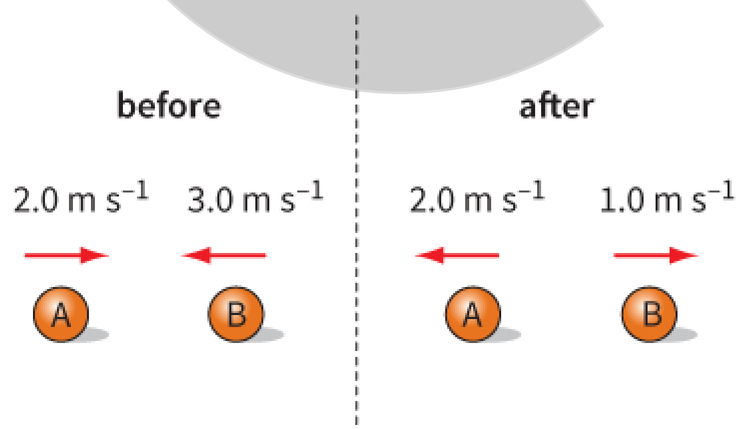


Figure 6.6: For Question 3.

6.3 Understanding collisions

The cars in Figure 6.7 have been badly damaged by a collision. The front of a car is designed to absorb the impact of the crash. It has a 'crumple zone', which collapses on impact. This absorbs most of the kinetic energy that the car had before the collision. It is better that the car's kinetic energy should be transferred to the crumple zone than to the driver and passengers.

Motor manufacturers make use of test labs to investigate how their cars respond to impacts. When a car is designed, the manufacturers combine soft, compressible materials that absorb energy with rigid structures that protect the people in the car. Old-fashioned cars had much more rigid structures. In a collision, they were more likely to bounce back and the violent forces involved were much more likely to prove fatal.

Figure 6.7: The front of each car has crumpled in, as a result of a head-on collision.

Two types of collision

When two objects collide, they may crumple and deform. Their kinetic energy may also disappear completely as they come to a halt. This is an example of an **inelastic collision**. Alternatively, they may spring apart, retaining all of their kinetic energy. This is a **perfectly elastic collision**. In practice, in most collisions, some kinetic energy is transformed into other forms (such as heat or sound) and the collision is inelastic. Previously we described the collisions as being 'springy' or 'sticky'. We should now use the correct scientific terms, perfectly elastic and inelastic.

We will look at examples of these two types of collision and consider what happens to linear momentum and kinetic energy in each.

A perfectly elastic collision

Two identical objects, A and B, moving at the same speed but in opposite directions, have a head-on collision, as shown in Figure 6.8. Each object bounces back with its velocity reversed. This is a perfectly elastic collision.

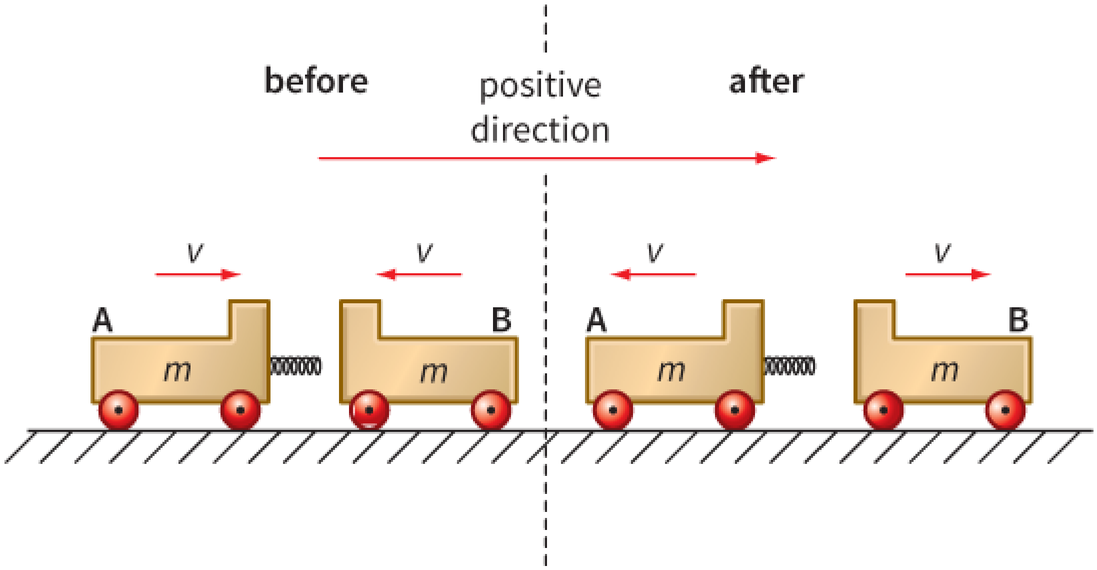


Figure 6.8: Two objects may collide in different ways: this is an elastic collision. An inelastic collision of the same two objects is shown in Figure 6.9.

You should be able to see that, in this collision, both momentum and kinetic energy are conserved. Before the collision, object A of mass m is moving to the right at speed v and object B of mass m is moving to the left at speed v. Afterwards, we still have two masses m moving with speed v, but now object A is moving to the left and object B is moving to the right. We can express this mathematically as follows.

Before the collision

Object	Mass	Velocity	Momentum
A	m	v	mv
B	m	$-v$	$-mv$

Object B has negative velocity and momentum because it is travelling in the opposite direction to object A. Therefore we have:

total momentum before collision

= momentum of A + momentum of B

$= mv + (-mv) = 0$

total kinetic energy before collision

= k.e. of A + k.e. of B

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

The magnitude of the momentum of each object is the same. Momentum is a vector quantity and we have to consider the directions in which the objects travel. The combined momentum is zero. On the other hand, kinetic energy is a scalar quantity and direction of travel is irrelevant. Both objects have the same kinetic energy and therefore the combined kinetic energy is twice the kinetic energy of a single object.

After the collision

Both objects have their velocities reversed, and we have:

$$\text{total momentum after collision} = (-mv) + mv = 0$$

$$\text{total kinetic energy after collision} = \frac{1}{2}mv^2 + \frac{1}{2}mv^2 = mv^2$$

So the total momentum and the total kinetic energy are unchanged. They are both conserved in a perfectly elastic collision such as this.

In this collision, the objects have a relative speed of $2v$ before the collision. After their collision, their velocities are reversed so their relative speed is $2v$ again. This is a feature of perfectly elastic collisions.

The relative speed of approach is the speed of one object measured relative to another. If two objects are travelling directly towards each other with speed v, as measured by someone stationary on the ground, then each object 'sees' the other one approaching with a speed of $2v$. Thus, if objects are travelling in opposite directions we add their speeds to find the relative speed. If the objects are travelling in the same direction then we subtract their speeds to find the relative speed.

To find the relative speed of two objects you subtract the velocity of one from the velocity of the other. This is the same as adding on a velocity in the opposite direction; so, if two objects approach each other in exactly opposite directions with velocities of v_1 and $-v_2$, their relative speed = $v_1 - (-v_2) = v_1 + v_2$.

KEY IDEA

In a perfectly elastic collision of two bodies, the relative speed of the body's approach is equal to the relative speed of their separation.

An inelastic collision

KEY IDEA

During an inelastic collision, the total kinetic energy of the bodies becomes smaller.

In Figure 6.9, the same two objects collide, but this time they stick together after the collision and come to a halt. Clearly, the total momentum and the total kinetic energy are both zero after the collision, since neither mass is moving. We have:

	Before collision	After collision
momentum	0	0
kinetic energy	$\frac{1}{2}mv^2$	0

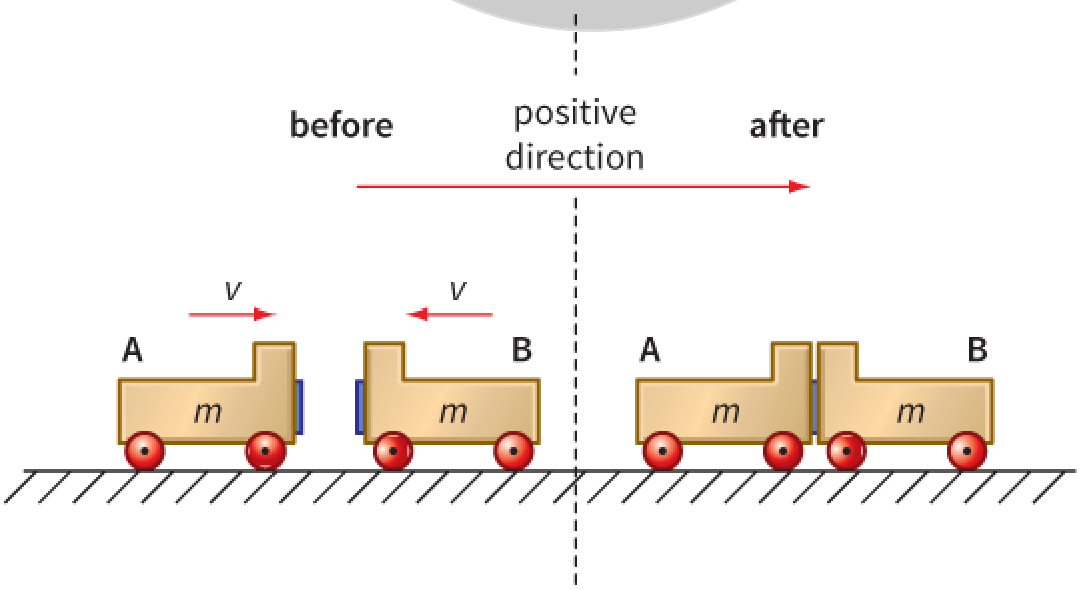


Figure 6.9: An inelastic collision between two identical objects. The trolleys are stationary after the collision.

Again we see that momentum is conserved. However, kinetic energy is not conserved. It is lost because work is done in deforming the two objects.

In fact, **momentum is always conserved in all collisions**. There is nothing else into which momentum can be converted. Kinetic energy is usually not conserved in a collision, because it can be transformed into other forms of energy – sound energy if the collision is noisy, and the energy involved in deforming the objects (which usually ends up as internal energy – they get warmer). Of course, the total amount of energy remains constant, as stated in the principle of conservation of energy.

Question

4 Copy this table, choosing the correct words from each pair.

Type of collision	perfectly elastic	inelastic
Momentum	conserved / not conserved	conserved / not conserved
Kinetic energy	conserved / not conserved	conserved / not conserved
Total energy	conserved / not conserved	conserved / not conserved

Solving collision problems

We can use the idea of conservation of momentum to solve numerical problems, as shown in Worked example 2.

WORKED EXAMPLE

2 In the game of bowls, a player rolls a large ball towards a smaller, stationary ball. A large ball of mass 5.0 kg moving at 10.0 m s^{-1} strikes a stationary ball of mass 1.0 kg. The smaller ball flies off at 10.0 m s^{-1}.

a Determine the final velocity of the large ball after the impact.

b Calculate the kinetic energy 'lost' in the impact.

Step 1 Draw two diagrams, showing the situations before and after the collision. Figure 6.10 shows the values of masses and velocities; since we don't know the velocity of the large ball after the collision, this is shown as *v*. The direction from left to right has been assigned the 'positive' direction.

Step 2 Using the principle of conservation of momentum, set up an equation and solve for the value of *v*:

$$\begin{aligned}
\text{total momentum before collison} &= \text{total momentum after collison} \\
(5.0 \times 10) + (1.0 \times 0) &= (5.0 \times v) + (1.0 \times 10) \\
50 + 0 &= 5.0v + 10 \\
v &= \frac{40}{5.0} \\
v &= 8.0 \text{ m s}^{-1}
\end{aligned}$$

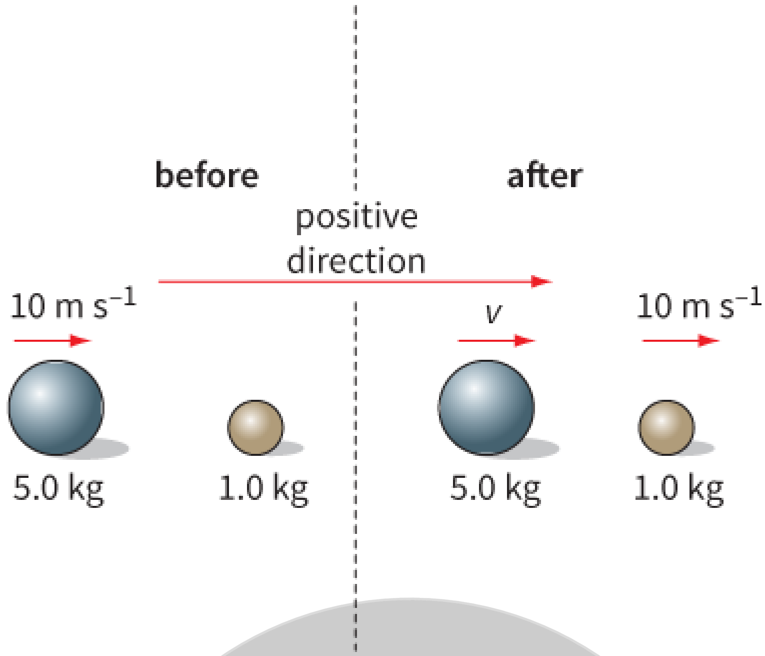


Figure 6.10: When solving problems involving collisions, it is useful to draw diagrams showing the situations before and after the collision. Include the values of all the quantities that you know.

So the speed of the large ball decreases to 8.0 m s^{-1} after the collision. Its direction of motion is unchanged – the velocity remains positive.

Step 3 Knowing the large ball's final velocity, calculate the change in kinetic energy during the collision:

total k.e. before collision $= \frac{1}{2} \times 5.0 \times 10^2 + 0$

$= 250$ J

total k.e. after collision $= \frac{1}{2} \times 5.0 \times 8.0^2 + \frac{1}{2} \times 1.0 \times 10^2$

$= 210$ J

total k.e. 'lost' in the collision $= 250 \text{ J} - 210 \text{ J}$

$= 40$ J

This 'lost' kinetic energy will appear as internal energy (the two balls get warmer) and as sound energy (we hear the collision between the balls).

Questions

5 Figure 6.11 shows two identical balls A and B about to make a head-on collision. After the collision, ball A rebounds at a speed of 1.5 m s^{-1} and ball B rebounds at a speed of 2.5 m s^{-1}. The mass of each ball is 4.0 kg.

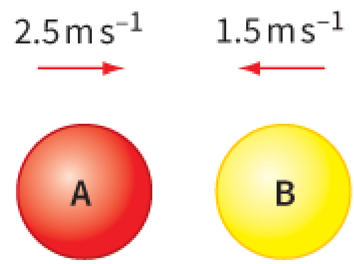


Figure 6.11: For Question 5.

a Calculate the momentum of each ball before the collision.

b Calculate the momentum of each ball after the collision.

c Is the momentum conserved in the collision?

d Show that the total kinetic energy of the two balls is conserved in the collision.

e Show that the relative speed of the balls is the same before and after the collision.

6 A trolley of mass 1.0 kg is moving at 2.0 m s^{-1}. It collides with a stationary trolley of mass 2.0 kg. This second trolley moves off at 1.2 m s^{-1}.

a Draw 'before' and 'after' diagrams to show the situation.

b Use the principle of conservation of momentum to calculate the speed of the first trolley after the collision. In what direction does it move?

6.4 Explosions and crash-landings

There are situations where it may appear that momentum is being created out of nothing, or that it is disappearing without trace. Do these contradict the principle of conservation of momentum?

The rockets shown in Figure 6.12 rise high into the sky. As they start to fall, they send out showers of chemical packages, each of which explodes to produce a brilliant sphere of burning chemicals. Material flies out in all directions to create a spectacular effect.

Does an explosion create momentum out of nothing? The important point to note here is that the burning material spreads out equally in all directions. Each tiny spark has momentum, but for every spark, there is another moving in the opposite direction, i.e., with opposite momentum. Since momentum is a vector quantity, the total amount of momentum created is zero.

Figure 6.12: These exploding rockets produce a spectacular display of bright sparks in the night sky.

At the same time, kinetic energy is created in an explosion. Burning material flies outwards; its kinetic energy has come from the chemical potential energy stored in the chemical materials before they burn.

More fireworks

Roman candles are a type of firework that fire a jet of burning material into the sky. This is another type of explosion, but it doesn't send material in all directions. The firework tube directs the material upwards. Has momentum been created out of nothing here?

Again, the answer is no. The chemicals have momentum upwards, but at the same time, the roman candle pushes downwards on the Earth. An equal amount of downwards momentum is given to the Earth. Of course, the Earth is massive, and we don't notice the tiny change in its velocity that results.

Down to Earth

If you push a large rock over a cliff, its speed increases as it falls. Where does its momentum come from? And when it lands, where does its momentum disappear to?

The rock falls because of the pull of the Earth's gravity on it. This force is its weight and it makes the rock accelerate towards the Earth. Its weight does work and the rock gains kinetic energy. It gains momentum downwards. Something must be gaining an equal amount of momentum in the opposite (upward) direction. It is the Earth, which starts to move upwards as the rock falls downwards. The mass of the Earth is so great that its change in velocity – far too small to be noticeable.

When the rock hits the ground, its momentum becomes zero. At the same instant, the Earth also stops moving upwards. The rock's momentum cancels out the Earth's momentum. At all times during the rock's fall and crash-landing, momentum has been conserved.

If a rock of mass 60 kg is falling towards the Earth at a speed of 20 m s^{-1}, how fast is the Earth moving towards it? Figure 6.13 shows the situation. The mass of the Earth is 6.0×10^{24} kg. We have:

$$\text{total momentum of Earth and rock} = 0$$

Therefore:

$$(60 \times 20) + (6.0 \times 10^{24} \times v) = 0$$

$$v = -2.0 \times 10^{-22} \text{ m s}^{-1}$$

The minus sign shows that the Earth's velocity is in the opposite direction to that of the rock. The Earth moves very slowly indeed. In the time of the rock's fall, it will move much less than the diameter of the nucleus of an atom!

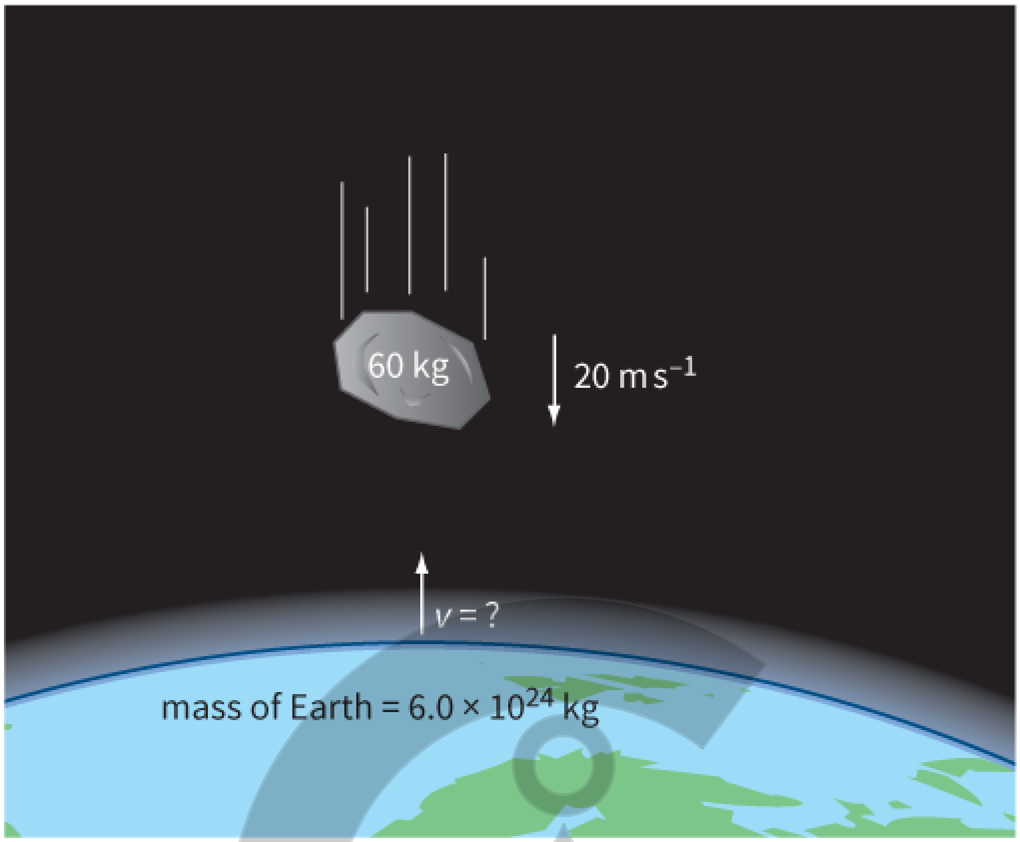


Figure 6.13: The rock and Earth gain momentum in opposite directions.

Questions

7 Discuss whether momentum is conserved in each of the following situations.

a A star explodes in all directions – a supernova.

b You jump up from a trampoline. As you go up, your speed decreases; as you come down again, your speed increases.

8 A ball of mass 0.40 kg is thrown at a wall. It strikes the wall with a speed of 1.5 m s^{-1} perpendicular to the wall and bounces off the wall with a speed of 1.2 m s^{-1}. Explain the changes in momentum and energy that happen in the collision between the ball and the wall. Give numerical values where possible.

6.5 Collisions in two dimensions

It is rare that collisions happen in a straight line–in one dimension. Figure 6.14 shows a two-dimensional collision between two snooker balls. From the multiple images, we can see how the velocities of the two balls change:

- At first, the white ball is moving straight forwards. When it hits the red ball, it moves off to the right. Its speed decreases; we can see this because the images get closer together.
- The red ball moves off to the left. It moves off at a bigger angle than the white ball, but more slowly – the images are even closer together.

How can we understand what happens in this collision, using the ideas of momentum and kinetic energy?

At first, only the white ball has momentum, and this is in the forward direction. During the collision, this momentum is shared between the two balls. We can see this because each has a component of velocity in the forward direction.

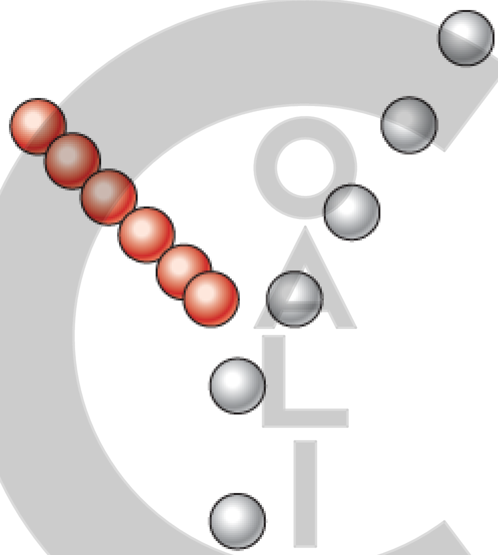


Figure 6.14: The white ball strikes the red ball a glancing blow. The two balls move off in different directions.

At the same time, each ball gains momentum in the sideways direction, because each has a sideways component of velocity – the white ball to the right, and the red ball to the left. These must be equal in magnitude and opposite in direction, otherwise we would conclude that momentum had been created out of nothing. The red ball moves at a greater angle, but its velocity is less than that of the white ball, so that the component of its velocity at right angles to the original track is the same as the white ball's.

Figure 6.15a shows the momentum of each ball before and after the collision. We can draw a vector triangle to represent the changes of momentum in this collision (Figure 6.15b). The two momentum vectors after the collision add up to equal the momentum of the white ball before the collision. The vectors form a closed triangle because momentum is conserved in this two-dimensional collision.

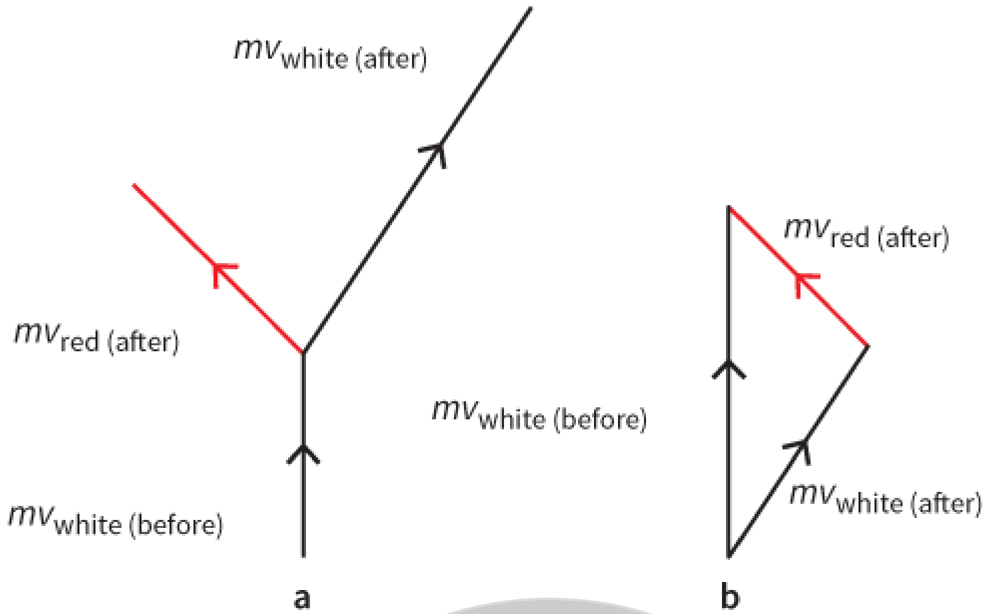


Figure 6.15: a These vectors represent the momenta of the colliding balls shown in Figure 6.14. **b** The closed vector triangle shows that momentum is conserved in the collision.

Components of momentum

Momentum is a vector quantity and so we can split it into components in order to solve problems.

Worked example 3 shows how to find an unknown velocity.

Worked example 4 shows how to demonstrate that momentum has been conserved in a two-dimensional collision.

WORKED EXAMPLES

3 A white ball of mass m = 1.0 kg and moving with initial speed u = 0.5 m s^{-1} collides with a stationary red ball of the same mass. They move off so that each has the same speed and the angle between their paths is 90°. What is their speed?

Step 1 Draw a diagram to show the velocity vectors of the two balls, before and after the collision (Figure 6.16). We will show the white ball initially travelling along the y-direction.

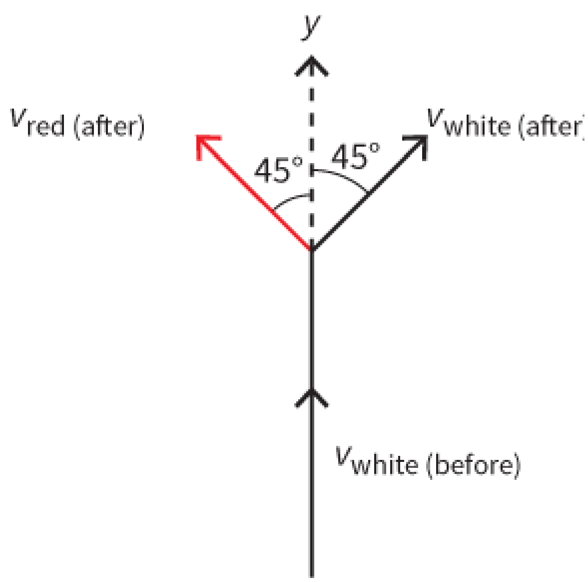


Figure 6.16: Velocity vectors for the white and red balls.

Because we know that the two balls have the same final speed *v*, their paths must be symmetrical about the *y*-direction. Since their paths are at 90° to one other, each must be at 45° to the *y*-direction.

Step 2 We know that momentum is conserved in the *y*-direction. Hence we can say:

initial momentum of white ball in *y*-direction

= final component of momentum of white ball in *y*-direction

+ final component of momentum of red ball in *y*-direction

This is easier to understand using symbols:

$$mu = mv_y + mv_y$$

where v_y is the component of *v* in the *y*-direction. The right-hand side of this equation has two identical terms, one for the white ball and one for the red. We can simplify the equation to give:

$$mu = 2mv_y$$

Step 3 The component of *v* in the *y*-direction is *v* cos 45°. Substituting this, and including values of *m* and *u*, gives

$$0.5 = 2v \cos 45°$$

and hence

$$v = \frac{0.5}{2 \cos 45°} \approx 0.354 \text{ m s}^{-1}$$

So each ball moves off at 0.354 m s^{-1} at an angle of 45° to the initial direction of the white ball.

4 Figure 6.17 shows the momentum vectors for particles 1 and 2, before and after a collision. Show that momentum is conserved in this collision.

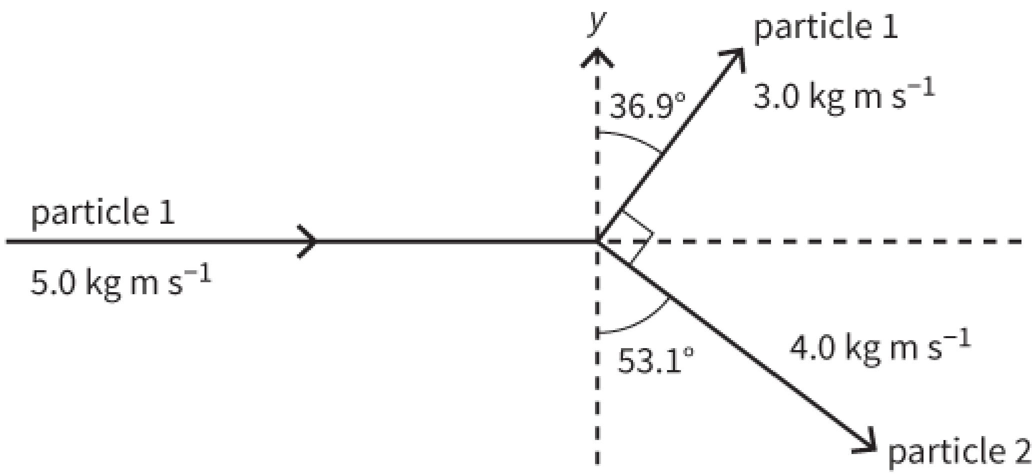


Figure 6.17: Momentum vectors: particle 1 has come from the left and collided with particle 2.

Step 1 Consider momentum changes in the *y*-direction.

Before collision:

$$\text{momentum} = 0$$

(because particle 1 is moving in the *x*-direction and particle 2 is stationary).

After collision:

component of momentum of particle 1

$= 3.0 \cos 36.9° \approx 2.40$ kg m s^{-1} upwards

component of momentum of particle 2

$= 4.0 \cos 53.1° \approx 2.40$ kg m s^{-1} downwards

These components are equal and opposite, and hence their sum is zero. Hence, momentum is conserved in the *y*-direction.

Step 2 Consider momentum changes in the *x*-direction.

Before collision:
momentum = 5.0 kg m s^{-1} to the right

After collision:

component of momentum of particle 1

$= 3.0 \cos 53.1° \approx 1.80$ kg m s^{-1} to the right

component of momentum of particle 2

$= 4.0 \cos 36.9° \approx 3.20$ kg m s^{-1} to the right

total momentum to the right = 5.0 kg m s^{-1}

Hence, momentum is conserved in the *x*-direction.

Step 3 An alternative approach would be to draw a vector triangle similar to Figure 6.15b. In this case, the numbers have been chosen to make this easy; the vectors form a 3–4–5 right-angled triangle.

Because the vectors form a closed triangle, we can conclude that:

momentum before collision = momentum after collision

(in other words, momentum is conserved)

Questions

9 A snooker ball strikes a stationary ball. The second ball moves off sideways at 60° to the initial path of the first ball.

Use the idea of conservation of momentum to explain why the first ball cannot travel in its initial direction after the collision. Illustrate your answer with a diagram.

10 Look back to Worked example 4. Draw the vector triangle that shows that momentum is conserved in the collision described in the question. Show the value of each angle in the triangle.

11 Figure 6.18 shows the momentum vectors for two identical particles, 1 and 2, before and after a collision. Particle 2 was at rest before the collision. Show that momentum is conserved in this collision.

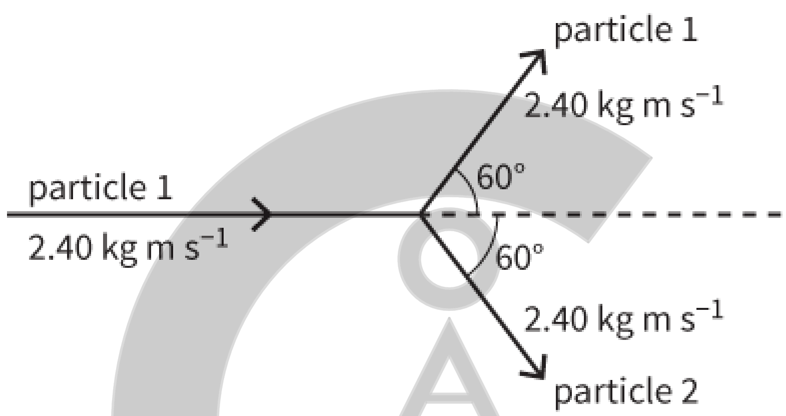


Figure 6.18: For Question 11.

12 A snooker ball collides with a second identical ball as shown in Figure 6.19.

a Determine the components of the velocity of the first ball in the *x*- and *y*-directions.

b Hence, determine the components of the velocity of the second ball in the *x*- and *y*-directions.

c Hence, determine the velocity (magnitude and direction) of the second ball.

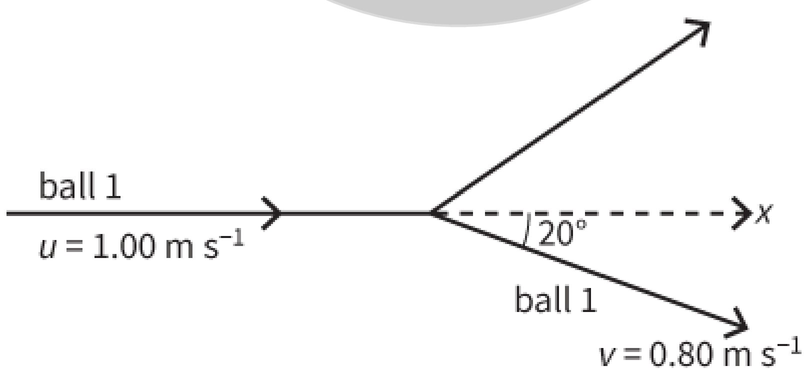
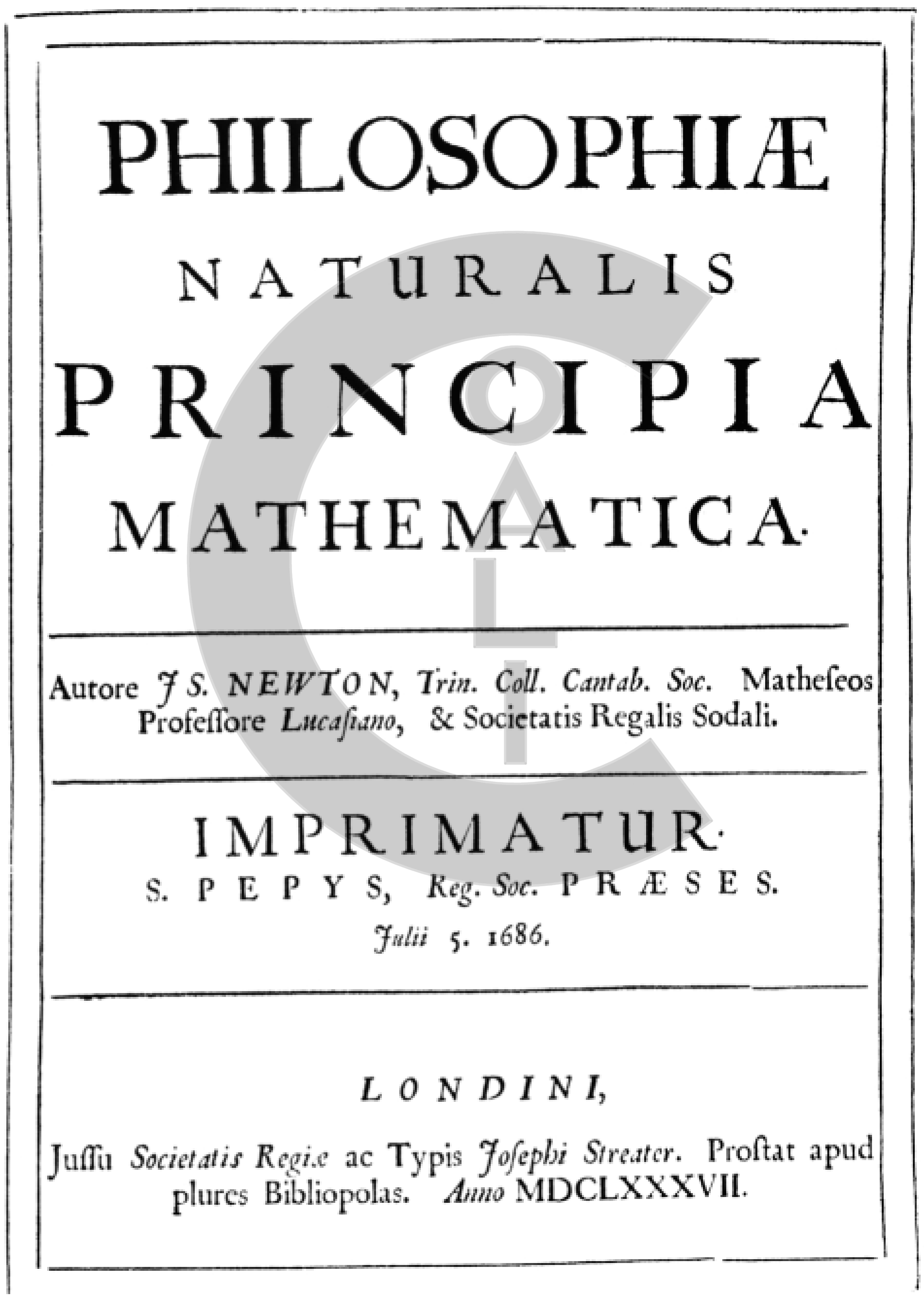


Figure 6.19: For Question 12.

6.6 Momentum and Newton's laws

The main concepts in physics are often very simple; it takes only a few words to express them and they can be applied to lots of situations. However, 'simple' does not mean 'easy'. Some concepts are quite abstract – such as force, energy and voltage. Scientists had to use their imagination to conceive such concepts. Other scientists then spent years working, experimenting, testing and refining the concepts until they finally reached the established concepts that we use today.



PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.

Autore *J S. NEWTON*, *Trin. Coll. Cantab. Soc.* Matheſeos Profeſſore *Lucaſiano*, & Societatis Regalis Sodali.

IMPRIMATUR.
S. PEPYS, *Reg. Soc.* PRÆSES.
Julii 5. 1686.

LONDINI,

Juſſu *Societatis Regiæ* ac Typis *Joſephi Streater*. Proſtat apud plures Bibliopolas. *Anno* MDCLXXXVII.

Figure 6.20: The title page of Newton's *Principia*, in which he outlined his theories of the laws that govern the motion of objects.

Isaac Newton’s work on motion is a good example. Newton published his ideas in a book; the book’s title translates as *Mathematical Principles of Natural Philosophy*.

Newton wanted to develop an understanding of the idea of ‘force’. You may have been told in your early studies of science that ‘a force is a push or a pull’. Newton’s idea was that forces are interactions between bodies and that they change the motion of the body that they act on. Forces acting on an object can produce acceleration. For an object of constant mass, this acceleration is directly proportional to the resultant force acting on the object. That is much more like a scientific definition of force.

6.7 Understanding motion

In Chapter 3, we looked at Newton's laws of motion. We can get further insight into these laws by thinking about them in terms of momentum.

Newton's first law of motion

In everyday speech, we sometimes say that something has momentum when we mean that it keeps on moving on its own. An oil tanker is difficult to stop at sea, because of its momentum. We use the same word even when we're not talking about an object: 'The election campaign is gaining momentum', for example. This idea of keeping on moving is just what we discussed in connection with **Newton's first law of motion**:

An object will remain at rest or keep travelling at constant velocity unless it is acted on by a resultant force.

An object travelling at constant velocity has constant momentum. Hence, the first law is really saying that the momentum of an object remains the same unless the object experiences an external force.

Newton's second law of motion

Newton's second law of motion links the idea of the resultant force acting on an object and its momentum. A statement of Newton's second law is:

The resultant force acting on an object is directly proportional to the rate of change of the linear momentum of that object. The resultant force and the change in momentum are in the same direction.

Hence:

$$\text{resultant force} \propto \text{rate of change of momentum}$$

This can be written as:

$$F \propto \frac{\Delta p}{\Delta t}$$

where F is the resultant force and Δp is the change in momentum taking place in a time interval of Δt. (Remember that the Greek letter delta, Δ, is a shorthand for 'change in', so Δp means 'change in momentum'.) The changes in momentum and force are both vector quantities, so these two quantities must be in the same direction.

The unit of force (the newton, N) is defined to make the constant of proportionality equal to one, so we can write the second law of motion mathematically as:

$$F = \frac{\Delta p}{\Delta t}$$

Worked example 5 shows how to use this equation. This equation also shows the newton second (N s) can be used as a unit of momentum.

If the forces acting on an object are balanced, there is no resultant force and the object's momentum will remain constant. If a resultant force acts on an object, its momentum (velocity and/or direction) will change. The equation gives us another way of stating Newton's second law of motion:

The resultant force acting on an object is equal to the rate of change of its momentum. The resultant force and the change in momentum are in the same direction.

This statement effectively defines what we mean by a force; it is an interaction that causes an object's momentum to change. So, if an object's momentum is changing, there must be a force acting on it. We can find the size and direction of the force by measuring the rate of change of the object's momentum.

KEY EQUATION

Resultant force ∝ rate of change of momentum:

$$F = \frac{\Delta p}{\Delta t}$$

WORKED EXAMPLE

5 Calculate the average force acting on a 900 kg car when its velocity changes from 5.0 m s^{-1} to 30 m s^{-1} in a time of 12 s.

Step 1 Write down the quantities given:

m = 900 kg

initial velocity *u* = 5.0 m s^{-1}

Δt = 12 s

Step 2 Calculate the initial momentum and the final momentum of the car:

momentum = mass × velocity

$$\begin{aligned} \text{initial momentum} &= mu = 900 \times 5.0 \\ &= 4500\text{k g m s}^{-1} \\ \text{final momentum} &= mv = 900 \times 30 \\ &= 27\,000\text{ kgm s}^{-1} \end{aligned}$$

Step 3 Use Newton's second law of motion to calculate the average force on the car:

$$\begin{aligned} F &= \frac{\Delta p}{\Delta t} \\ &= \frac{27000-4500}{12} \\ &= 1875\text{ N} \approx 1900\text{ N} \end{aligned}$$

The average force acting on the car is about 1.9 kN.

A special case of Newton's second law of motion

Imagine an object of constant mass *m* acted upon by a resultant force *F*. The force will change the momentum of the object. According to Newton's second law of motion, we have:

$$F = \frac{\Delta p}{\Delta t} = \frac{mv-mu}{t}$$

where *u* is the initial velocity of the object, *v* is the final velocity of the object and *t* is the time taken for the change in velocity. The mass *m* of the object is a constant; hence the equation can be rewritten as:

$$\begin{aligned} F &= \frac{m(v-u)}{\Delta t} \\ &= m\left(\frac{v-u}{t}\right) \end{aligned}$$

The term in brackets on the right-hand side is the acceleration *a* of the object. Therefore, a special case of Newton's second law is:

$$F = ma$$

We have already met this equation in Chapter 3. In Worked example 5, you could have determined the average force acting on the car using this simplified equation for Newton's second law of motion. Remember that the equation $F = ma$ is a special case of $F = \frac{\Delta p}{\Delta t}$ that only applies when the mass of the object is constant. There are situations where the mass of an object changes as it moves, for example, a rocket that burns a phenomenal amount of chemical fuel as it accelerates upwards.

Questions

13 A car of mass 1000 kg is travelling at a velocity of +10 m s^{-1}. It accelerates for 15 s, reaching a velocity of +24 m s^{-1}. Calculate:

- **a** the change in the momentum of the car in the 15 s period
- **b** the average resultant force acting on the car as it accelerates.

14 A ball is kicked by a footballer. The average force on the ball is 240 N and the impact lasts for a time interval of 0.25 s.

- **a** Calculate the change in the ball's momentum.
- **b** State the direction of the change in momentum.

15 Water pouring from a broken pipe lands on a flat roof. The water is moving at 5.0 m s^{-1} when it strikes the roof. The water hits the roof at a rate of 10 kg s^{-1}. Calculate the force of the water hitting the roof. (Assume that the water does not bounce as it hits the roof. If it did bounce, would your answer be greater or smaller?)

16 A golf ball has a mass of 0.046 kg. The final velocity of the ball after being struck by a golf club is 50 m s^{-1}. The golf club is in contact with the ball for a time of 1.3 ms. Calculate the average force exerted by the golf club on the ball.

Newton's third law of motion

Newton's third law of motion is about interacting objects. These could be two magnets attracting or repelling each other, two electrons repelling each other, etc. Newton's third law states:

When two bodies interact, the forces they exert on each other are equal and opposite.

How can we relate this to the idea of momentum? Imagine holding two magnets, one in each hand. You gradually bring them towards each other (Figure 6.21) so that they start to attract each other. Each feels a force pulling it towards the other. The two forces are the same size, even if one magnet is stronger than the other. One magnet could even be replaced by an unmagnetised piece of steel and they would still attract each other equally.

If you release the magnets, they will gain momentum as they are pulled towards each other. One gains momentum to the left while the other gains equal momentum to the right.

Each is acted on by the same force, and for the same time. So, momentum is conserved. In fact, the law of conservation of momentum can be proved using Newton's second and third laws of motion. Consider an object of mass m_x and velocity v_x colliding with a mass m_y and velocity v_y. If the system is closed, then the force F_x and the force F_y on the twomasses are equal and opposite.

$$F_x = -F_y$$

$$\frac{\Delta m_x v_x}{\Delta t} = -\frac{\Delta m_y v_y}{\Delta t}$$

$$\frac{\Delta(m_x v_x + m_y v_y)}{\Delta t} = 0$$

So, $\Delta(m_x v_x + m_y v_y) = 0$ and there has been no change in the total momentum.

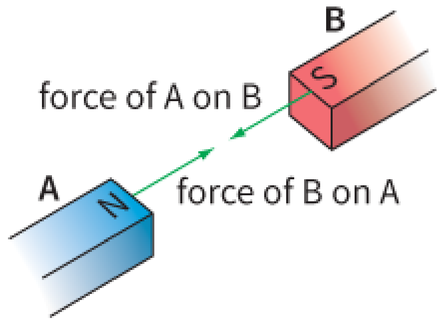


Figure 6.21: Newton's third law states that the forces these two magnets exert one each other must be equal and opposite.

REFLECTION

What did you learn about yourself as you worked through this chapter?

Which principle do you think is the most important, conservation of momentum or conservation of energy?

SUMMARY

Linear momentum is the product of mass and velocity: $p = mv$

The principle of conservation of momentum: For a closed system, the total momentum before an interaction (e.g., collision) is equal to the total momentum after the interaction.

In all interactions or collisions, momentum and total energy are conserved.

Kinetic energy is conserved in a perfectly elastic collision; relative speed is unchanged in a perfectly elastic collision.

In an inelastic collision, kinetic energy is not conserved. It is transferred into other forms of energy (such as heat or sound). Most collisions are inelastic.

Newton's first law of motion: An object will remain at rest or keep travelling at constant velocity unless it is acted on by a resultant force.

Newton's second law of motion: The resultant force acting on a body is equal to the rate of change of its momentum:

resultant force = rate of change of momentum or $F = \frac{\Delta p}{\Delta t} = ma$ when mass m remains constant.

Newton's third law of motion: When two bodies interact, the forces they exert on each other are equal and opposite.

EXAM-STYLE QUESTIONS

1 Which quantity has the same unit as the rate of change of momentum? **[1]**

A acceleration

B energy

C weight

D work

2 A railway truck of mass 8000 kg travels along a level track at a velocity of 2.5 m s^{-1} and collides with a stationary truck of mass 12 000 kg. The collision takes 4.0 s and the two trucks move together at the same velocity after the collision.

What is the average force that acts on the 8000 kg truck during the collision? **[1]**

A 2000 N

B 3000 N

C 5000 N

D 12 000 N

3 An object has mass 2.0 ± 0.2 kg and a velocity of 10 ± 1 m s^{-1}.

What is the percentage uncertainty in the momentum of the object? **[1]**

A 1%

B 6%

C 10%

D 20%

4 An object is dropped and its momentum increases as it falls toward the ground. Explain how the law of conservation of momentum and Newton's third law of motion can be applied to this situation. **[2]**

5 A ball of mass 2.0 kg, moving at 3.0 m s^{-1}, strikes a wall and rebounds with almost exactly the same speed. State and explain whether there is a change in:

a the momentum of the ball **[3]**

b the kinetic energy of the ball. **[1]**

[Total: 4]

6 **a** Define linear momentum. **[1]**

b Determine the base units of linear momentum in the SI system. **[1]**

c A car of mass 900 kg starting from rest has a constant acceleration of 3.5 m s^{-2}. Calculate its momentum after it has travelled a distance of 40 m. **[2]**

d This diagram shows two identical objects about to make a head-on collision. The objects stick together during the collision. Determine the final speed of the objects. State the direction in which they move. **[3]**

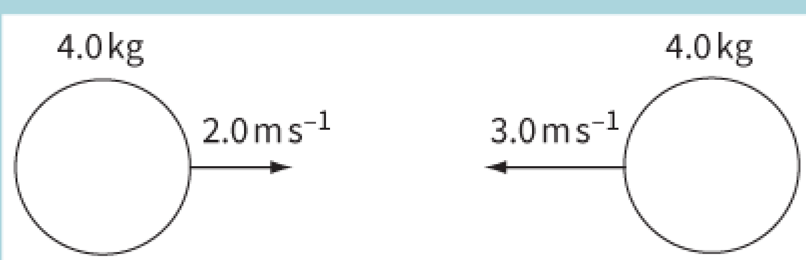


Figure 6.22

[Total: 7]

7 a Explain what is meant by an:

i elastic collision **[1]**

ii inelastic collision. **[1]**

b A snooker ball of mass 0.35 kg hits the side of a snooker table at right angles and bounces off also at right angles. Its speed before collision is 2.8 m s^{-1} and its speed after is 2.5 m s^{-1}. Calculate the change in the momentum of the ball. **[2]**

c Explain whether or not momentum is conserved in the situation described in part **b**. **[1]**

[Total: 5]

8 A car of mass 1100 kg is travelling at 24 m s^{-1}. The driver applies the brakes and the car decelerates uniformly and comes to rest in 20 s.

a Calculate the change in momentum of the car. **[2]**

b Calculate the braking force on the car. **[2]**

c Determine the braking distance of the car. **[2]**

[Total: 6]

9 A marble of mass 100 g is moving at a speed of 0.40 m s^{-1} in the *x*-direction.

a Calculate the marble's momentum. **[2]**

The marble strikes a second, identical marble. Each moves off at an angle of 45° to the *x*-direction.

b Use the principle of conservation of momentum to determine the speed of each marble after the collision. **[3]**

c Show that kinetic energy is conserved in this collision. **[2]**

[Total: 7]

10 A cricket bat strikes a ball of mass 0.16 kg travelling towards it. The ball initially hits the bat at a speed of 25 m s^{-1} and returns along the same path with the same speed. The time of impact is 0.0030 s. You may assume no force is exerted on the bat by the cricketer during the actual collision.

a Determine the change in momentum of the cricket ball. **[2]**

b Determine the force exerted by the bat on the ball. **[2]**

c Describe how the laws of conservation of energy and momentum apply to this impact and state whether the impact is elastic or inelastic. **[4]**

[Total: 8]

11 a State the principle of conservation of momentum and state the condition under which it is valid. **[2]**

b An arrow of mass 0.25 kg is fired horizontally towards an apple of mass 0.10 kg that is hanging on a string, as shown in Figure 6.23.

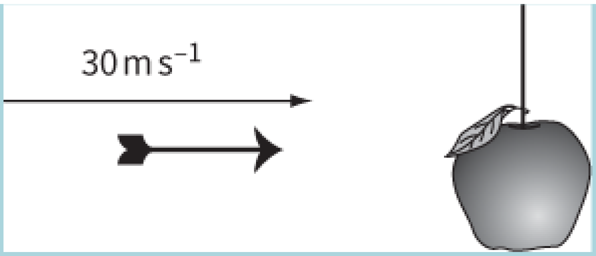


Figure 6.23

The horizontal velocity of the arrow as it enters the apple is 30 m s^{-1}. The apple was initially at rest and the arrow sticks in the apple.

i Calculate the horizontal velocity of the apple and arrow immediately after the impact. **[2]**

ii Calculate the change in momentum of the arrow during the impact. **[2]**

iii Calculate the change in total kinetic energy of the arrow and apple during the impact. **[2]**

iv A rubber-tipped arrow of mass 0.25 kg is fired at the centre of a stationary ball of mass 0.25 kg. The collision is perfectly elastic. Describe what happens and state the relative speed of separation of the arrow and the ball. **[2]**

[Total: 10]

12 a State what is meant by:

i a perfectly elastic collision **[1]**

ii a completely inelastic collision. **[1]**

b A stationary uranium nucleus disintegrates, emitting an alpha-particle of mass 6.65×10^{-27} kg and another nucleus X of mass 3.89×10^{-25} kg.

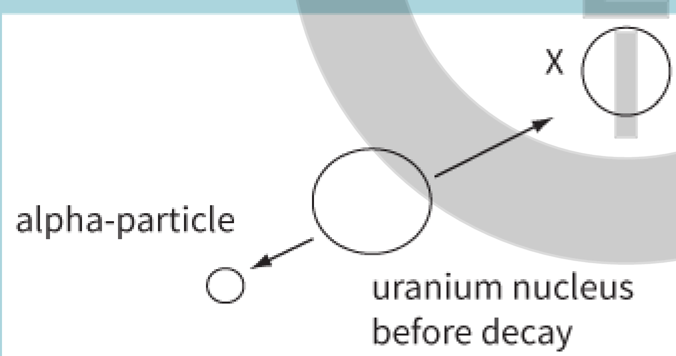


Figure 6.24

i Explain why the alpha-particle and nucleus X must be emitted in exactly opposite directions. **[2]**

ii Using the symbols v_α and v_x for velocities, write an equation for the conservation of momentum in this disintegration. **[1]**

iii Using your answer to part **b ii**, calculate the ratio $v_\alpha : v_x$ after the disintegration. **[1]**

[Total: 6]

13 a State **two** quantities that are conserved in an elastic collision. **[1]**

b A machine gun fires bullets of mass 0.014 kg at a speed of 640 m s^{-1}.

i Calculate the momentum of each bullet as it leaves the gun. **[1]**

ii Explain why a soldier holding the machine gun experiences a force when the gun is firing. **[2]**

iii The maximum steady horizontal force that a soldier can exert on the gun is 140 N. Calculate the maximum number of bullets that the gun can fire in one second. **[2]**

[Total: 6]

14 Two railway trucks are travelling in the same direction and collide. The mass of truck X is 2.0×10^4 kg and the mass of truck Y is 3.0×10^4 kg. This graph shows how the velocity of each truck varies with time.

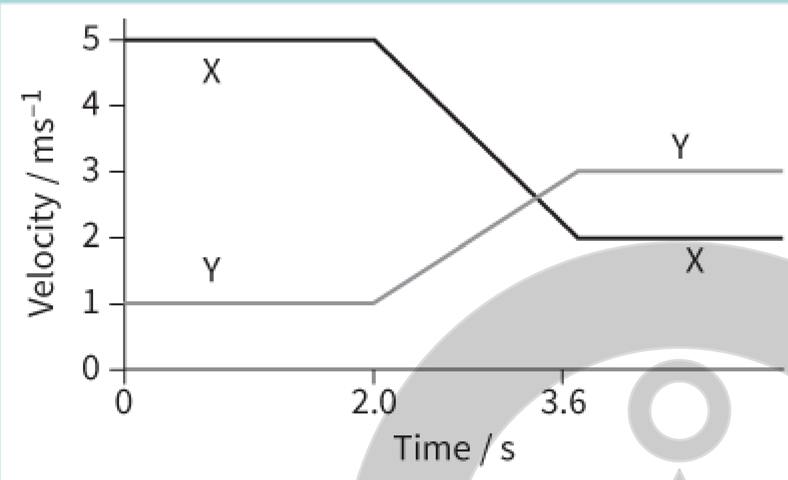


Figure 6.25

a Copy and complete the table. **[6]**

	Change in momentum / kg m s^{-1}	Initial kinetic energy / J	Final kinetic energy / J
truck X			
truck Y			

Table 6.1: For Question 14.

b State and explain whether the collision of the two trucks is an example of an elastic collision. **[2]**

c Determine the force that acts on each truck during the collision. **[2]**

[Total: 10]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define and use linear momentum	6.2			
state and apply the principle of conservation of momentum to collisions in one and two dimensions	6.3, 6.5			
relate force to the rate of change of momentum	6.2			
state all three of Newton’s laws of motion	6.7			
recall that, for a perfectly elastic collision, the relative speed of approach is equal to the relative speed of separation	6.3			
discuss energy changes in perfectly elastic and inelastic collisions.	6.3			

〉Chapter 7

Matter and materials

LEARNING INTENTIONS

In this chapter you will learn how to:

- define and use density
- define and use pressure and calculate the pressure in a fluid
- derive and use the equation $\Delta p = \rho g \Delta h$
- use a difference in hydrostatic pressure to explain and calculate upthrust
- explain how tensile and compressive forces cause deformation
- describe the behaviour of springs and use Hooke's law
- distinguish between elastic and plastic deformation, limit of proportionality and the elastic limit
- define and use stress, strain and the Young modulus
- describe an experiment to measure the Young modulus
- calculate the energy stored in a deformed material.

BEFORE YOU START

- Write down some notes to answer these questions: What are physical properties of materials? What properties make some materials really useful?
- Have you ever stretched a spring, rubber band or a small strip of plastic? Try to describe what you notice when these materials are stretched.

SPRINGY STUFF

In everyday life, we make great use of elastic materials. The term 'elastic' means springy; that is, the material deforms when a force is applied and returns to its original shape when the force is removed. Rubber is an elastic material. This is obviously important for a bungee jumper (Figure 7.1). The bungee rope must have the correct degree of elasticity. The jumper must be brought gently to a halt. What happens if the rope is too stiff or too springy? Discuss these problems with others – particularly if you have had experience of a bungee jump.

In this chapter, we will look at how forces can change the shape of an object. Before that, we will look at two important quantities, density and pressure.

Figure 7.1: The stiffness and elasticity of rubber are crucial factors in bungee jumping.

7.1 Density

Density is a property of matter. It tells us about how concentrated the matter is in a particular material. Density is a constant for a given material under specific conditions.

Density is defined as the mass per unit volume of a substance:

$$\text{density} = \frac{\text{mass}}{\text{volume}}$$
$$\rho = \frac{m}{V}$$

The symbol used here for density, ρ, is the Greek letter rho.

The standard unit for density in the SI system is kg m^{-3}, but you may also find values quoted in g cm^{-3}. It is useful to remember that these units are related by:

$$1000 \text{ kg m}^{-3} = 1 \text{ g cm}^{-3}$$

and that the density of water is approximately 1000 kg m^{-3}.

KEY EQUATION

$$\text{density} = \frac{\text{mass}}{\text{volume}}$$
$$\rho = \frac{m}{V}$$

Questions

1 A cube of copper has a mass of 240 g. Each side of the cube is 3.0 cm long. Calculate the density of copper in g cm^{-3} and in kg m^{-3}.

2 The density of steel is 7850 kg m^{-3}. Calculate the mass of a steel sphere of radius 0.15 m. (First, calculate the volume of the sphere using the formula $V = \frac{4}{3}\pi r^3$ and then use the density equation.)

7.2 Pressure

A fluid (liquid or gas) exerts **pressure** on the walls of its container, or on any surface with which it is in contact. Solids can also exert pressure on a surface with which it is in contact.

The pressure in a gas or liquid produces a force perpendicular to any surface.

The force the fluid pressure produces on the walls of a container can be in any direction, because the walls of the container may be horizontal, vertical or at any angle. A big force on a small area produces a high pressure.

Pressure is defined as the normal force acting per unit cross-sectional area.

We can write this as a word equation:

$$\text{pressure} = \frac{\text{normal force}}{\text{cross-sectional area}}$$

$$p = \frac{F}{A}$$

The word 'normal' in this context means at right angles to the surface.

KEY EQUATION

$$\text{pressure} = \frac{\text{normal force}}{\text{cross-sectional area}}$$

$$p = \frac{F}{A}$$

Force is measured in newtons and area is measured in square metres. The units of pressure are thus newtons per square metre ($N\ m^{-2}$), which are given the special name of pascals (Pa).

$$1\ \text{Pa} = 1\ \text{N m}^{-2}$$

Questions

3 A chair stands on four feet, each of area 10 cm^2. The chair weighs 80 N. Calculate the pressure it exerts on the floor.

4 Estimate the pressure you exert on the floor when you stand on both feet. (You could draw a rough rectangle around both your feet placed together to find the area in contact with the floor. You will also need to calculate your weight from your mass.)

Pressure in a fluid

The pressure in a fluid (a liquid or gas) increases with depth. Divers know this – the further they dive down, the greater the water pressure acting on them. The pressure acts at right angles to every part of their body and acts to crush them. Pilots know this – the higher they fly, the lower is the pressure of the atmosphere. The atmospheric pressure we experience on the surface of the Earth is due to the weight of the atmosphere above us, pressing downwards on the surface of the Earth or at right angles to every surface of our bodies.

The pressure in a fluid depends on three factors:

- the depth h below the surface
- the density ρ of the fluid
- the acceleration due to gravity, g.

In fact, change in pressure p is proportional to each of these and we have:

$$\text{change in pressure} = \text{density} \times \text{acceleration due to gravity} \times \text{depth}$$
$$\Delta p = \rho g h$$

KEY EQUATION

$$\text{change in pressure} = \text{density} \times \text{acceleration due to gravity} \times \text{depth}$$
$$\Delta p = \rho g h$$

You must learn how to derive this equation.

We can derive this relationship using Figure 7.2.

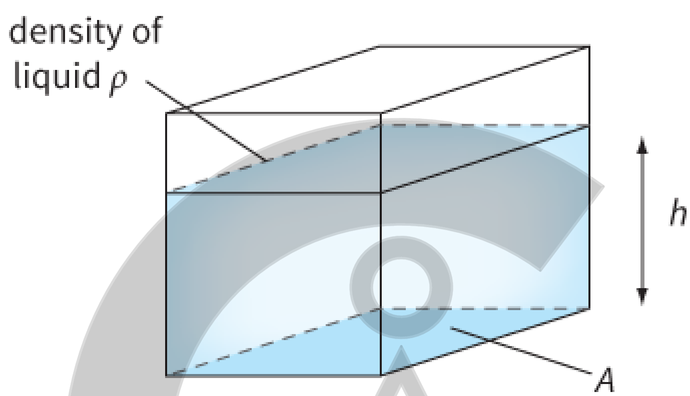


Figure 7.2: The weight of water in a tank exerts pressure on its base.

The force acting on the shaded area *A* on the bottom of the tank is caused by the weight of water above it, pressing downwards. We can calculate this force and hence the pressure as follows:

$$\text{volume of water} = A \times h$$
$$\text{mass of water} = \text{density} \times \text{volume} = \rho \times A \times h$$
$$\text{weight of water} = \text{mass} \times g = \rho \times A \times h \times g$$
$$\text{change in pressure} = \frac{\text{force}}{\text{area}}$$
$$= \rho \times A \times h \times \frac{g}{A}$$
$$= \rho \times g \times h$$

The equation is written as $\Delta p = \rho g h$ because this formula calculates the *difference* in pressure between the top and bottom of the water in the tank. There is, of course, atmospheric pressure acting on the water at the top of the tank. The total pressure at the bottom of the tank is atmospheric pressure + Δp.

WORKED EXAMPLE

1 Figure 7.3 shows a manometer used to measure the pressure of a gas supply. Calculate the pressure difference between the gas inside the pipe and atmospheric pressure.

Step 1 Determine the difference in height *h* of the water on the two sides of the manometer.

$$h = 60 - 30 = 30 \text{ cm}$$

Step 2 Because the level of water on the side of the tube next to the gas pipe is lower than on the side open to the atmosphere, the pressure in the gas pipe is above atmospheric pressure.

pressure difference = $\rho \times g \times h$

$= 1000 \times 9.81 \times 0.30 = 2940$ Pa

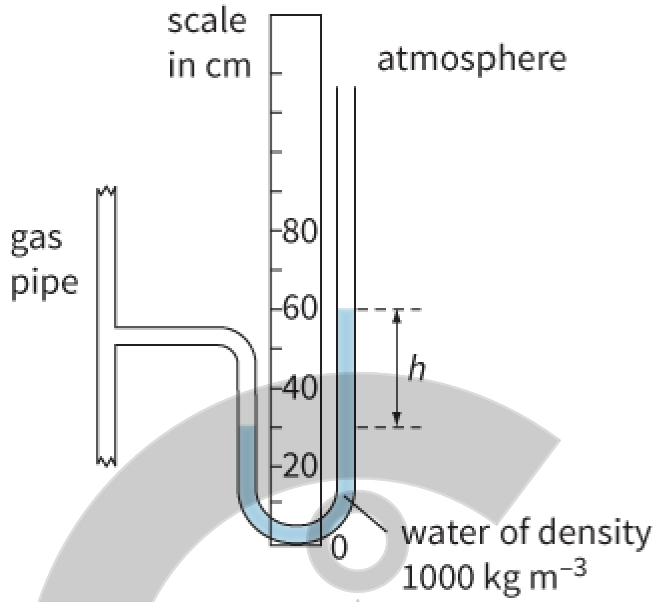


Figure 7.3: For Worked example 1.

Questions

5 Calculate the pressure due to the water on the bottom of a swimming pool if the depth of water in the pool varies between 0.8 m and 2.4 m. (Density of water = 1000 kg m^{-3}.) If atmospheric pressure is 1.01×10^5 Pa, calculate the maximum total pressure at the bottom of the swimming pool.

6 Estimate the height of the atmosphere if atmospheric density at the Earth's surface is 1.29 kg m^{-3}. (Atmospheric pressure = 101 kPa.)

7.3 Archimedes' principle

The variation of pressure with depth can be used to explain **Archimedes' principle**.

Archimedes' principle states that the upthrust acting on a body is equal to the weight of the liquid or gas that it displaces.

KEY EQUATION

$$\text{upthrust} = \rho g V$$
$$= \text{Weight of liquid displaced}$$

When the object is placed in a liquid, it **displaces** *some* of the liquid. In other words, it takes up some of the space of the liquid. The volume of the liquid displaced is equal to the volume of the liquid taken up by the object. If the object floats, the volume displaced is equal to the volume of the part of the object that is *under* the surface of the liquid.

Consider a rectangular shaped object immersed in a liquid (Figure 7.4). There is a larger pressure on the bottom surface than there is on the top surface because the bottom surface is deeper in the liquid.

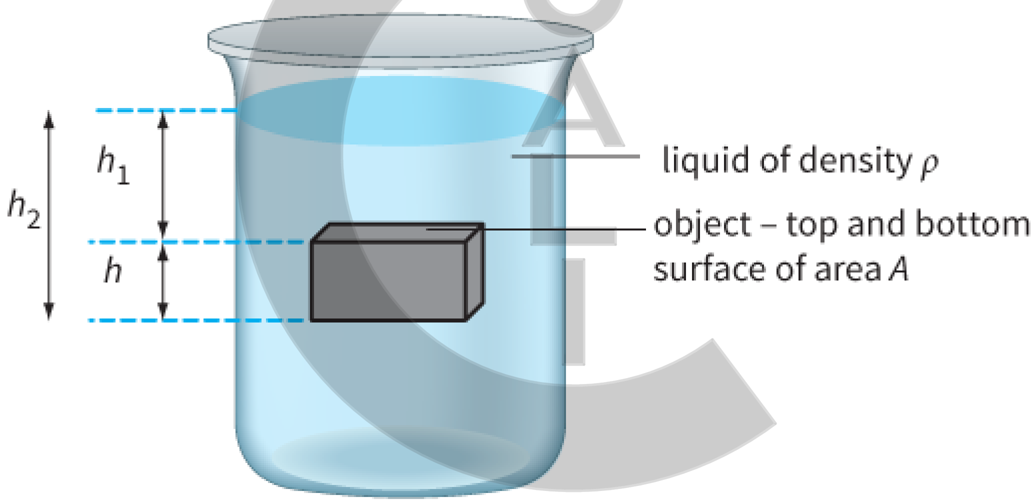


Figure 7.4: To explain Archimedes' principle.

The pressure on the top surface produces a force downwards on the top. It may seem surprising, but the pressure on the bottom surface actually produces a force upwards on the object. This is because pressure can act in any direction and always acts at right angles to a surface in a liquid. You may also be surprised to know that pressure is a scalar quantity even though it is defined in terms of force (which is a vector). Since pressure acts in all directions at a point it is not possible to define a single direction for it!

Because the pressure is larger on the bottom surface, the force acting upwards on the bottom surface is larger than the force acting downwards on the top surface. This is the cause of the upthrust, which you experience when you swim. Because your density is less than that of water, when you are underwater, the weight of water you displace is greater than your own weight. The upthrust is, therefore, greater than your own weight and there is a resultant force upwards to bring you to the surface.

To calculate this upthrust:

The force due to water on the top surface $F_1 = \rho \times g \times h_1 \times A$

Similarly, the force due to the water on the bottom surface is $F_2 = \rho \times g \times h_2 \times A$

$$\text{upthrust} = F_2 - F_1 = \rho \times g \times (h_2 - h_1) \times A = \rho \times g \times h \times A$$
$$= \rho \times g \times V$$

where the volume of the object $V = h \times A$

= the weight of the liquid displaced

WORKED EXAMPLES

2 A cube of side 0.20 m floats in water with 0.15 m below the surface of the water. The density of water is 1000 kg m^{-3}. Calculate the pressure due to the water that acts upwards on the bottom surface of the cube and the force upwards on the cube caused by this pressure. (This force is the upthrust on the cube.)

Step 1 Use the equation for pressure:

$$p = \rho \times g \times h = 1000 \times 9.81 \times 0.15$$
$$= 1470 \text{ Pa}$$

Step 2 Calculate the area of the base of the cube, and use this area in the equation for force.

area of base of cube = 0.2 × 0.2 = 0.04 m^2

$$\text{force} = \text{pressure} \times \text{area}$$
$$= 1470 \times 0.04 = 58.8 \text{ N}$$

3 A metal block of mass 0.60 kg has dimensions 0.050 m × 0.040 m × 0.030 m.
The block is hung from a newton-meter. What is the reading on the newton-meter when the block is fully submerged in liquid of density 1200 kg m^{-3}?

Step 1 Calculate the weight of the block. This is the reading on the meter when the block is in the air, before it is placed in the liquid.

weight = mg = 0.60 × 9.81 = 5.886 = 5.9 N

Step 2 Calculate the upthrust.

The volume of liquid displaced = 0.05 × 0.04 × 0.03 = 6.0 × 10^{-5} m^3

mass of liquid displaced = density × volume = 1200 × 6.0 × 10^{-5} = 7.2 × 10^{-2} kg

upthrust = weight of liquid displaced = 7.2 × 10^{-2} × 9.81 = 0.71 N

Step 3 Calculate the final reading

The upthrust must be subtracted from the weight of the object, so the newton-meter reads 5.89 − 0.71 = 5.2 N.

Questions

7 **a** Why is it difficult to hold an inflated plastic ball underwater?

b A submarine floats at rest under the water. To rise to the surface compressed air is used to push water out of its ‘ballast’ tanks into the sea. Why does this cause the submarine to rise?

8 A boat has a uniform cross-sectional area at the water line of 750 m^2. Fifteen cars of average mass 1200 kg are driven on board. Calculate the extra depth that the boat sinks in water of density 1000 kg m^{-3}.

9 Describe how to use a newton-meter, a micrometer screw gauge, a metal cube of side approximately 1.0 cm and a beaker of water to show *experimentally* that Archimedes’ principle is correct. The density of water is known to be 1000 kg m^{-3}.

10 A balloon of volume 3000 m^{-3} is filled with hydrogen of density 0.090 kg m^{-3}. The mass of the fabric of the balloon is 100 kg. Calculate the greatest mass that the balloon can lift in air of density 1.2 kg m^{-3}.

7.4 Compressive and tensile forces

A pair of forces is needed to change the shape of a spring. If the spring is being squashed and shortened, we say that the forces are **compressive**. More usually, we are concerned with stretching a spring, in which case the forces are described as **tensile** (Figure 7.5).

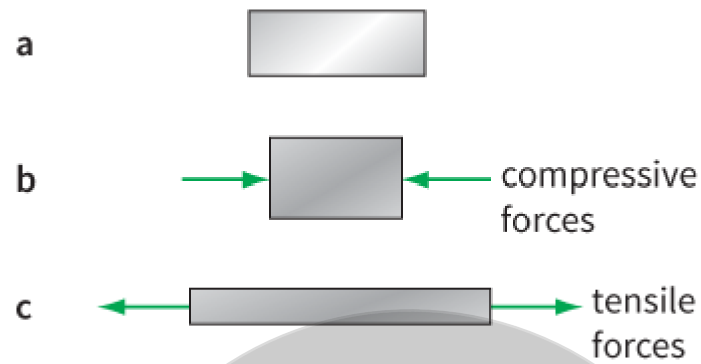


Figure 7.5: The effects of compressive and tensile forces.

When a wire is bent, some parts become longer and are in tension while other parts become shorter and are in compression. Figure 7.6 shows that the line AA becomes longer when the wire is bent and the line BB becomes shorter. The thicker the wire, the greater the compression and tension forces along its edges.

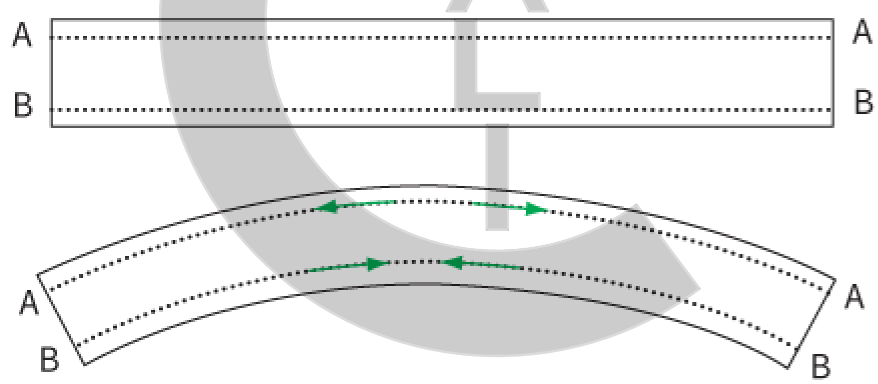


Figure 7.6: Bending a straight wire or beam results in tensile forces along the upper surface (the outside of the bend) and compressive forces on the inside of the bend.

It is simple to investigate how the length of a helical spring is affected by the applied force or load. The spring hangs freely with the top end clamped firmly (Figure 7.7). A load is added and gradually increased. For each value of the load, the extension of the spring is measured. Note that it is important to determine the increase in length of the spring, which we call the **extension**.

We can plot a graph of *force* against *extension* to find the stiffness of the spring, as shown in Figure 7.8.

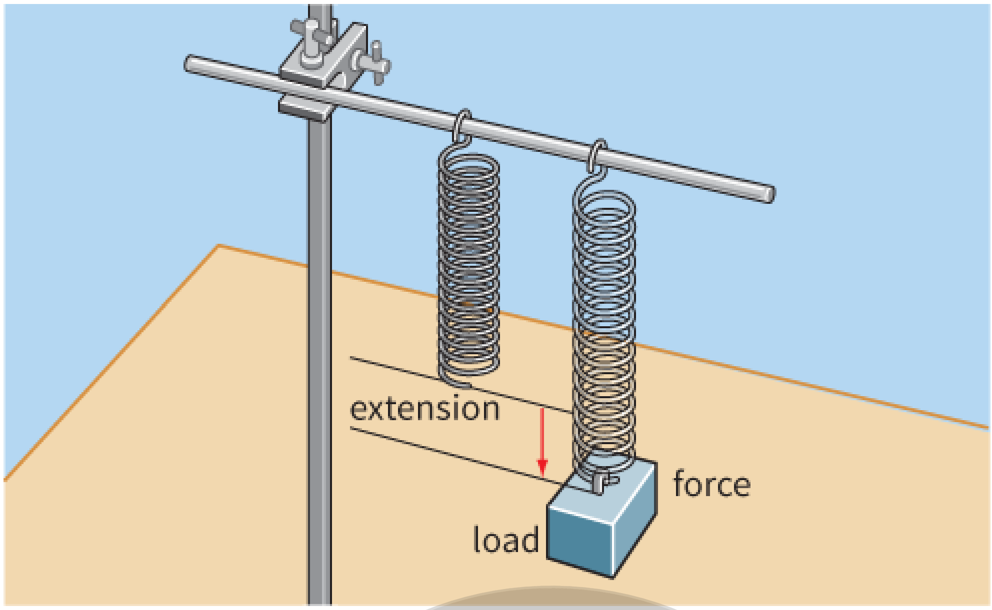


Figure 7.7: Stretching a spring.

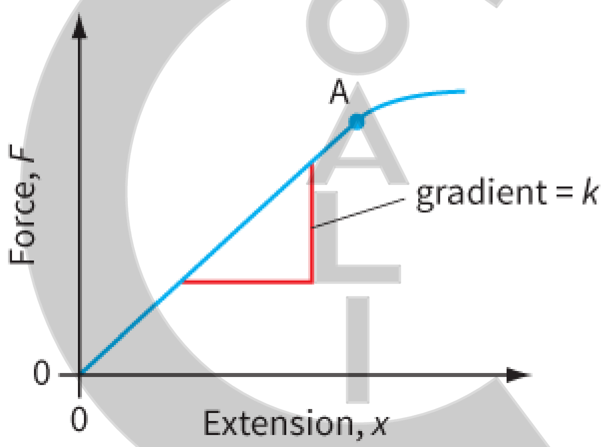


Figure 7.8: Force–extension graph for a spring.

Hooke's law

The usual way of plotting the results would be to have the force along the horizontal axis and the extension along the vertical axis. This is because we are changing the force (the independent variable) and this results in a change in the extension (the dependent variable). The graph shown in Figure 7.7 has extension on the horizontal axis and force on the vertical axis. This is a departure from the convention because the gradient of the straight section of this graph turns out to be an important quantity, known as the **spring constant**.

For a typical spring, the first section of this graph OA is a straight line passing through the origin. The extension x is directly proportional to the applied force (load) F. The behaviour of the spring in the linear region OA of the graph can be expressed by the following equation:

$$x \propto F \text{ or } F = kx$$

where k is the spring constant (sometimes called the stiffness or force constant of the spring). The spring constant is the force per unit extension, given by:

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

KEY EQUATION

$$\text{spring constant} = \frac{\text{force}}{\text{extension}}$$

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

The SI unit for the force constant is newtons per metre or N m^{-1}. We can find the force constant *k* from the gradient of section OA of the graph:

$$k = \text{gradient}$$

A stiffer spring will have a larger value for the force constant *k*. Beyond point A, the graph is no longer a straight line; its gradient changes and we can no longer use the equation $F = kx$.

If a spring or anything else responds to a pair of tensile forces in the way shown in section OA of Figure 7.7, we say that it obeys **Hooke's law**. A material obeys Hooke's law if the extension produced in it is proportional to the applied force (load).

The point A is known as the **limit of proportionality**. This is the point beyond which the extension is no longer proportional to the force.

If you apply a small force to a spring and then release it, it will return to its original length (this is **elastic deformation**.). However, if you apply a large force, the spring may not return to its original length; the spring has become permanently deformed (this is **plastic deformation**.). The force beyond which the spring becomes permanently deformed is known as the **elastic limit**.

The elastic limit is not necessarily the same point as the limit of proportionality, although they are likely to be close to each other.

This use of the word 'elastic' in elastic limit is slightly different from the idea of an elastic collision covered in Chapter 6. But the two ideas are related.

Question

11 Figure 7.9 shows the force–extension graph for a wire that is stretched and then released.

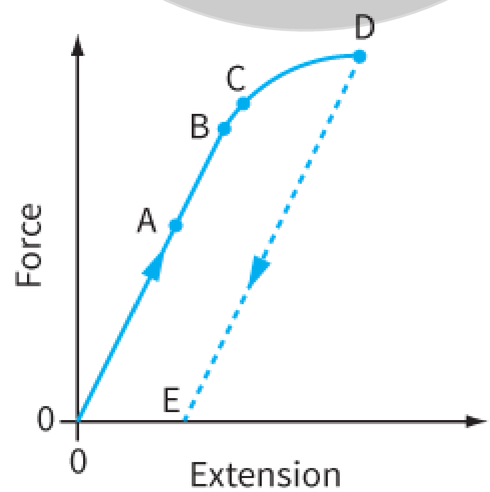


Figure 7.9: Force–extension graph for a wire.

a Which point shows the limit of proportionality?

b Which point shows the elastic limit?

PRACTICAL ACTIVITY 7.1

Investigating springs

Springs can be combined in different ways (Figure 7.10): end-to-end (in series) and side-by-side (in parallel). Using identical springs, you can measure the force constant of a single spring, and of springs in series and in parallel. Before you do this, predict the outcome of such an experiment. If the force constant of a single spring is *k*, what will be the equivalent force constant of:

- two springs in series?
- two springs in parallel?

This approach can be applied to combinations of three or more springs.

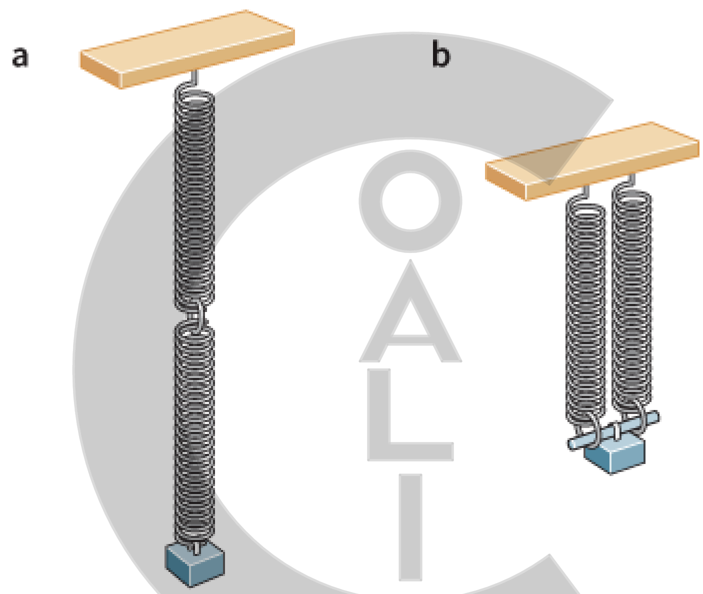


Figure 7.10: Two ways to combine a pair of springs: *a* in series; *b* in parallel.

12 Figure 7.11 shows the force–extension graphs for four springs, A, B, C and D.

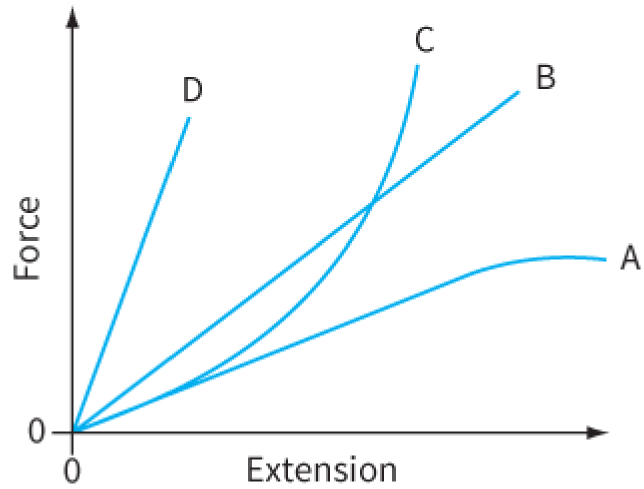


Figure 7.11: Force–extension graphs for four different springs.

a State which spring has the greatest value of force constant.

b State which is the least stiff.

c State which of the four springs does not obey Hooke's law.

7.5 Stretching materials

When we determine the force constant of a spring, we are only finding out about the stiffness of that particular spring. However, we can compare the stiffness of different materials. For example, steel is stiffer than copper, but copper is stiffer than lead.

Stress and strain

Figure 7.12 shows a simple way of assessing the stiffness of a wire in the laboratory. As the long wire is stretched, the position of the sticky tape pointer can be read from the scale on the bench.

Why do we use a long wire? Obviously, this is because a short wire would not stretch as much as a long one.

We need to take account of this in our calculations, and we do this by calculating the strain produced by the load. The **strain** is defined as the increase in length of a wire (its extension) divided by its the original length.

That is:

$$\text{strain} = \frac{\text{extension}}{\text{original length}}$$

$$\varepsilon = \frac{x}{L}$$

where ε is the strain, x is the extension of the wire and L is its original length.

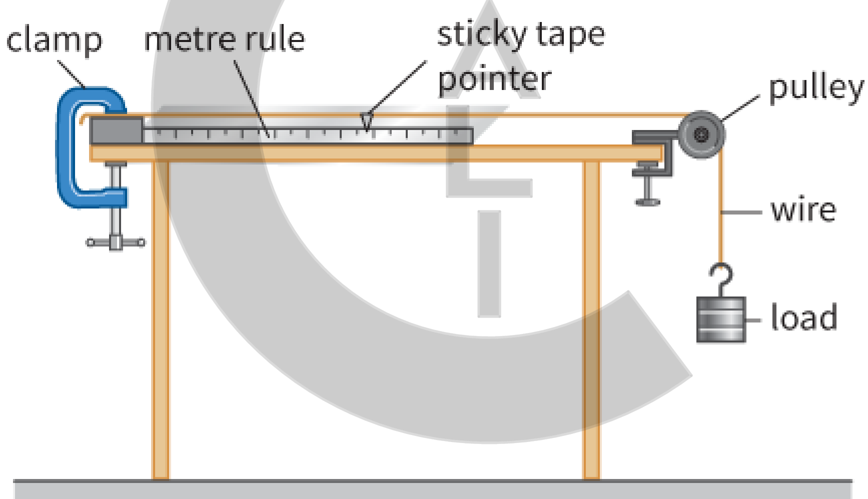


Figure 7.12: Stretching a wire in the laboratory. Wear eye protection and be careful not to overload the wire.

KEY EQUATION

$$\text{strain} = \frac{\text{extension}}{\text{original length}}$$

$$\varepsilon = \frac{x}{L}$$

For example, if a wire of length 1.500 m is stretched and the length becomes 1.518 m, the extension is 0.018 m and the strain $= \frac{0.018}{1.5000} = 0.012$

Note that both extension and original length must be in the same units, and so strain is a ratio, without units. Sometimes, strain is given as a percentage. For example, a strain of 0.012 is equivalent to 1.2%.

Why do we use a thin wire? This is because a thick wire would not stretch as much for the same force. Again, we need to take account of this in our calculations, and we do this by calculating the **stress** produced by the load.

The stress is defined as the force applied per unit cross-sectional area of the wire. That is:

$$\text{stress} = \frac{\text{normal force}}{\text{cross-sectional area}}$$

$$\sigma = \frac{F}{A}$$

where σ is the stress, F is the applied force that acts normally (at right angles) on a wire of cross-sectional area A.

KEY EQUATION

$$\text{stress} = \frac{\text{normal force}}{\text{cross-sectional area}}$$

$$\sigma = \frac{F}{A}$$

The units of stress are newtons per square metre (N m^{-2}) or pascals (Pa), the same as the units of pressure:

$$1\ \text{Pa} = 1\ \text{N m}^{-2}$$

The Young modulus

We can now find the stiffness of the material we are stretching. Rather than calculating the ratio of force to extension as we would for a spring or a wire, we calculate the ratio of stress to strain. This ratio is a constant for a particular material and does not depend on its shape or size. The ratio of stress to strain is called the **Young modulus** of the material. That is:

$$\text{Young modulus} = \frac{\text{stress}}{\text{strain}}$$

$$E = \frac{\sigma}{\varepsilon}$$

where E is the Young modulus of the material, σ is the stress and ε is the strain.

The unit of the Young modulus is the same as that for stress, N m^{-2} or Pa. In practice, values may be quoted in MPa or GPa. These units are related as:

$$1\ \text{MPa} = 10^6\ \text{Pa}$$

$$1\ \text{GPa} = 10^9\ \text{Pa}$$

Usually, we plot a graph with stress on the vertical axis and strain on the horizontal axis (Figure 7.13).

It is drawn like this so that the gradient is the Young modulus of the material. It is important to consider only the first, linear section of the graph. In the linear section stress is proportional to strain and the wire under test obeys Hooke's law.

Table 7.1 gives some values of the Young modulus for different materials.

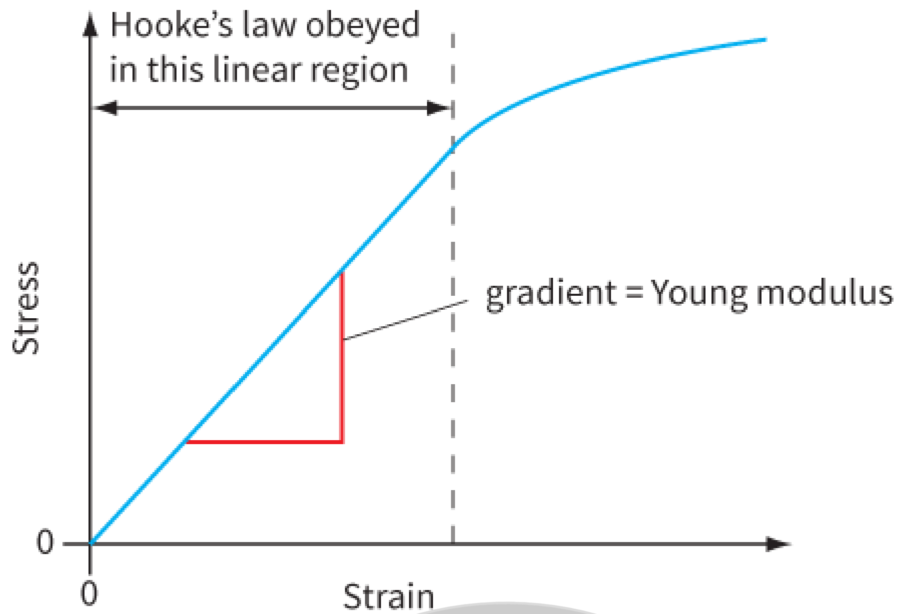


Figure 7.13: Stress–strain graph, and how to deduce the Young modulus. Note that we can only use the first, straight-line section of the graph.

Material	Young modulus / GPa
aluminium	70
brass	90–110
brick	7–20
concrete	40
copper	130
glass	70–80
iron (wrought)	200
lead	18
Perspex®	3
polystyrene	2.7–4.2
rubber	0.01
steel	210
tin	50
wood	10 approx.

Table 7.1: The Young modulus of various materials. Many of these values depend on the precise composition of the material concerned. (Remember, 1 GPa = 10^9 Pa.)

Questions

13 List the metals in Table 7.1 from stiffest to least stiff.

14 Which of the non-metals in Table 7.1 is the stiffest?

15 Figure 7.14 shows stress–strain graphs for two materials, A and B. Use the graphs to determine the Young modulus of each material.

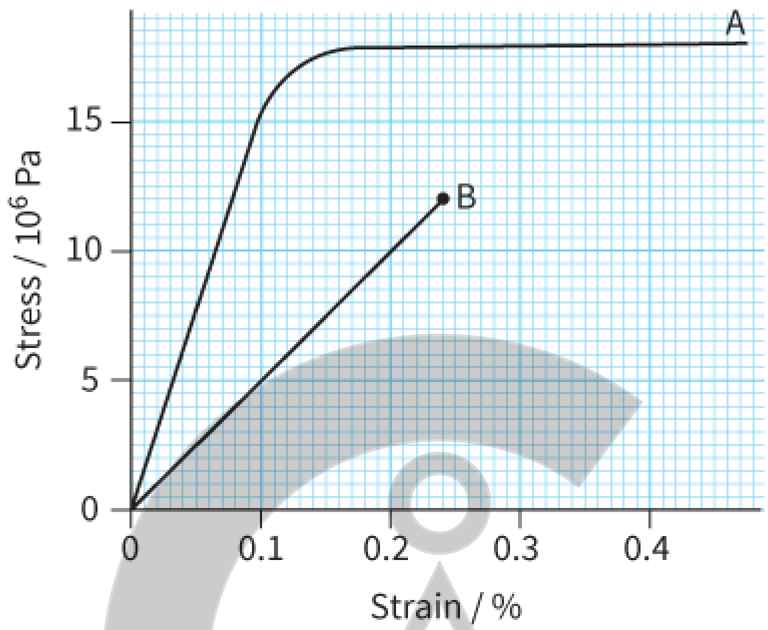


Figure 7.14: Stress–strain graphs for two different materials.

16 A piece of steel wire, 200.0 cm long and having cross-sectional area of 0.50 mm^2, is stretched by a force of 50 N. Its new length is found to be 200.1 cm. Calculate the stress and strain, and the Young modulus of steel.

17 Calculate the extension of a copper wire of length 1.00 m and diameter 1.00 mm when a tensile force of 10 N is applied to the end of the wire. (Young modulus of copper = 130 GPa.)

18 In an experiment to measure the Young modulus of glass, a student draws out a glass rod to form a fibre 0.800 m in length. Using a travelling microscope, she estimates its diameter to be 0.40 mm. Unfortunately, it proves impossible to obtain a series of readings for load and extension. The fibre snaps when a load of 1.00 N is hung on the end. The student judges that the fibre extended by no more than 1 mm before it snapped.

Use these values to obtain an estimate for the Young modulus of the glass used. Explain how the actual or accepted value for the Young modulus might differ from this estimate.

19 For each of the materials whose stress–strain graphs are shown in Figure 7.15, deduce the values of the Young modulus.

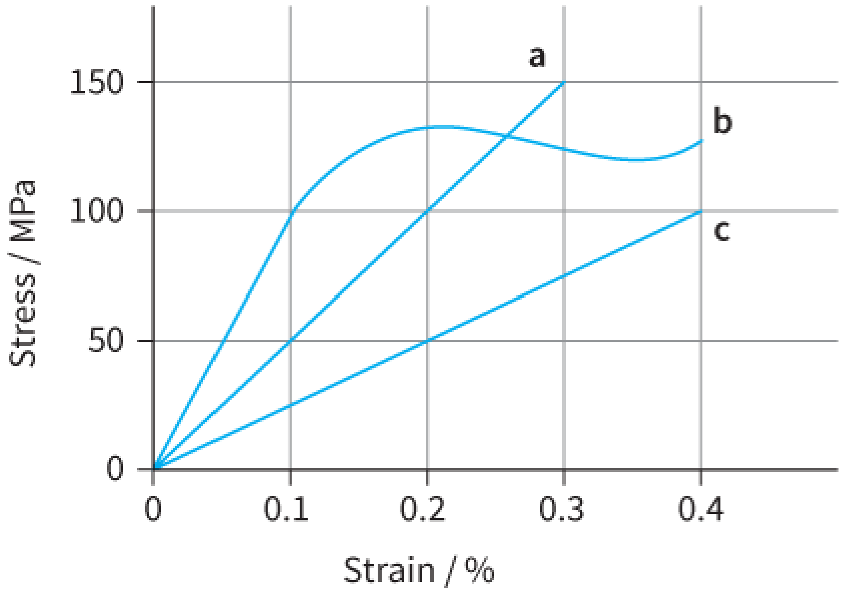


Figure 7.15: Stress–strain graphs for three materials.

PRACTICAL ACTIVITY 7.2

Determining the Young modulus

You must be able to describe this experiment in detail. Learn how to draw the diagram in Figure 7.11 and how to measure each of the quantities in:

$$E = \frac{FL}{Ax} = \left(\frac{F}{x}\right) \times \left(\frac{4L}{\pi d^2}\right)$$

Metals are not very elastic. Normally, they can only be stretched by about 0.1% of their original length. Beyond this, they become permanently or plastically deformed. As a result, some careful thought must be given to getting results that are good enough to give an accurate value of the Young modulus.

First, the wire used must be long. The increase in length is proportional to the original length, and so a longer wire gives larger and more measurable extensions. Typically, extensions up to 1 mm must be measured for a wire of length 1 m. To get suitable measurements of extension there are two possibilities: use a very long wire, or use a method that allows measurement of extensions that are a fraction of a millimetre.

The apparatus shown in Figure 7.12 can be used with a travelling microscope placed above the wire and focused on the sticky tape pointer. When the pointer moves, the microscope is adjusted to keep the pointer at the middle of the cross-hairs on the microscope. The distance that the pointer has moved can then be measured accurately from the scale on the microscope.

Second, the cross-sectional area of the wire must be known accurately. The diameter of the wire is measured using a micrometer screw gauge. This is reliable to within ±0.01 mm. Once the wire has been loaded in increasing steps, the load must be gradually decreased to ensure that there has been no permanent deformation of the wire.

A graph of *F* against *x* can be drawn and the gradient used to find an average value of $\frac{F}{x}$, where *F* is the weight of the load and *x* is the extension shown by the distance moved by the pointer.

The area *A* is found from $A = \frac{\pi d^2}{4}$, where *d* is the diameter of the wire.

The diameter should be measured at several points along the wire and the average value found. The length *L* is measured from the sticky pointer to the point where the wire is clamped.

Other materials, such as glass and many plastics, are also quite stiff and so it is difficult to measure their Young modulus. Rubber is not as stiff, and strains of several hundred per cent can be achieved. However, the stress–strain graph for rubber is not a straight line. This means the value of the Young modulus found is not very precise, because it only has a very small linear region on a stress–strain graph.

7.6 Elastic potential energy

Whenever you stretch a material, you are doing work. This is because you have to apply a force and the material extends in the direction of the force. You will know this if you have ever used exercise equipment with springs to develop your muscles (such as in Figure 7.16). Similarly, when you push down on the end of a springboard before diving, you are doing work. You transfer energy to the springboard, and you recover the energy when it pushes you up into the air.

We call the energy in a deformed solid the **elastic potential energy** or **strain energy**. If the material has been strained elastically (the elastic limit has not been exceeded), the energy can be recovered. If the material has been plastically deformed, some of the work done has gone into moving atoms past one another and the energy cannot be recovered.

The material warms up slightly. We can determine how much elastic potential energy is involved from a force–extension graph: see Figure 7.17. We need to use the equation that defines the amount of work done by a force. That is:

$$\text{work done} = \text{force} \times \text{distance moved in the direction of the force}$$

Figure 7.16: Using springs to help you exercise is hard work.

This equation only holds when the force is constant. When you stretch a spring the force varies; so how can you find the work done?

There are two approaches. You can:

- use the average force in the equation for work; this works well where the force–extension graph is a straight line
- add together many small extensions, in each of which the force hardly changes; adding together lots of very small extensions shows us that the work done is the area under the force–extension graph.

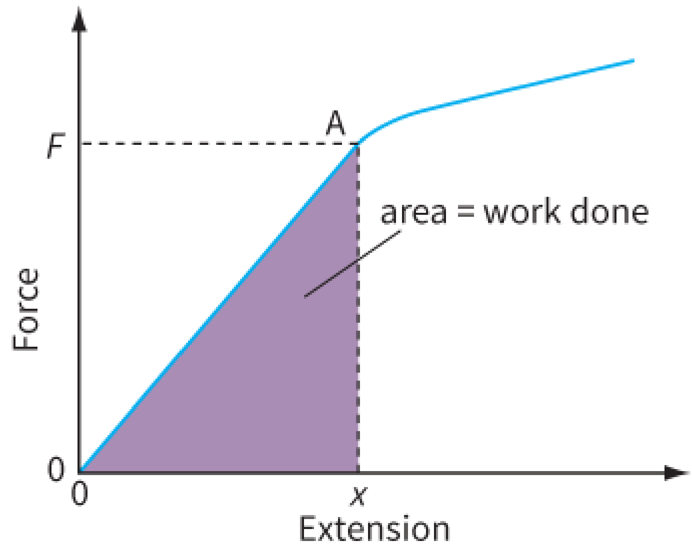


Figure 7.17: Elastic potential energy is equal to the area under the force–extension graph.

First, consider the linear region of the graph where Hooke's law is obeyed, OA. The graph in this region is a straight line through the origin. The extension *x* is directly proportional to the applied force *F*. There are two ways to find the work done.

Method 1

We can think about the average force needed to produce an extension *x*. The average force is half the final force *F*, and so we can write:

$$\begin{aligned} \text{elastic potential energy} &= \text{work done} \\ \text{elastic potential energy} &= \frac{\text{final force}}{2} \times \text{extension} \\ E &= \frac{1}{2} F x \end{aligned}$$

Method 2

The other way to find the elastic potential energy is to recognise that we can get the same answer by finding the area under the graph. The area shaded in Figure 7.17 is a triangle whose area is given by:

$$\text{area} = \frac{1}{2} \times \text{base} \times \text{height}$$

This again gives:

$$E = \frac{1}{2} F x$$

The **work done** in **stretching** or **compressing** a material is always equal to the area under the graph of force against extension.

This is true whatever the shape of the graph, provided we draw the graph with extension on the horizontal axis. If the graph is not a straight line, we cannot use the *Fx* relationship, so we have to resort to counting squares or some other technique to find the answer.

Note that the elastic potential energy relates to the elastic part of the graph (i.e. up to the elastic limit), so we can only consider the force–extension graph up to the elastic limit.

There is an alternative equation for elastic potential energy. We know that, according to Hooke's law, applied force *F* and extension *x* are related by $F = kx$, where *k* is the force constant. Substituting for *F* gives:

$$\text{elastic potential energy} = \tfrac{1}{2}Fx = \tfrac{1}{2} \times kx \times x = \tfrac{1}{2}kx^2$$

KEY EQUATION

$$\text{elastic potential energy} = \tfrac{1}{2}Fx$$

$$E = \tfrac{1}{2}Fx = \tfrac{1}{2}kx^2$$

Questions

20 A force of 12 N extends a length of rubber band by 18 cm. Estimate the energy stored in this rubber band. Explain why your answer can only be an estimate.

21 A spring has a force constant of 4800 N m^{-1}. Calculate the elastic potential energy when it is compressed by 2.0 mm.

WORKED EXAMPLE

4 Figure 7.18 shows a simplified version of a force–extension graph for a piece of metal. Find the elastic potential energy when the metal is stretched to its limit of proportionality, and the total work that must be done to break the metal.

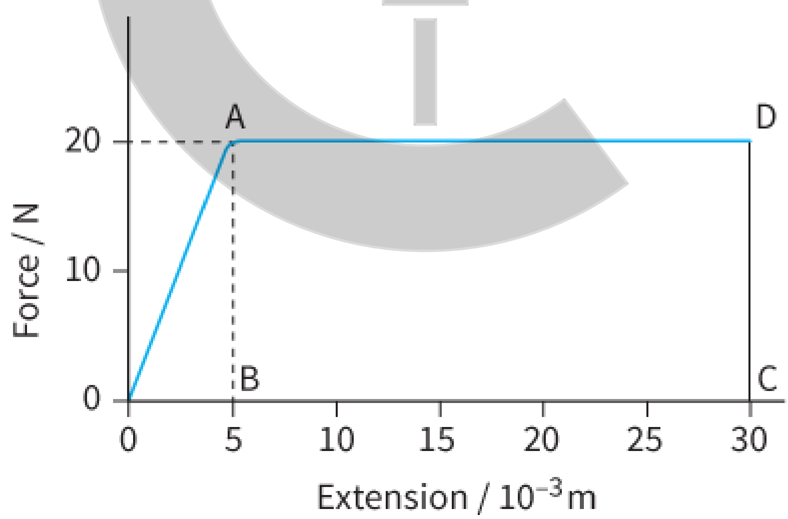


Figure 7.18: For Worked example 4.

Step 1 The elastic potential energy when the metal is stretched to its elastic limit is given by the area under the graph up to the elastic limit. We will have to take the limit of proportionality, the point A, as the elastic limit. The graph is a straight line up to $x = 5.0$ mm, $F = 20$ N, so the elastic potential energy is the area of triangle 0AB:

$$\text{elastic potential energy} = \frac{1}{2}Fx = \frac{1}{2} \times 20 \times 5.0 \times 10^{-3} = 0.05\text{ J}$$

Step 2 To find the work done to break the metal, we need to add on the area of the rectangle ABCD:

$$\text{work done} = \text{total area under graph} = 0.05 + (20 \times 25 \times 10^{-3}) = 0.05 + 0.50 = 0.55\text{ J}$$

22 Figure 7.19 shows force–extension graphs for two materials. For each of the following questions, make the statement required. Also explain how you deduce your answer from the graphs.

a State which polymer has the greater stiffness.

b State which polymer requires the greater force to break it.

c State which polymer requires the greater amount of work to be done in order to break it.

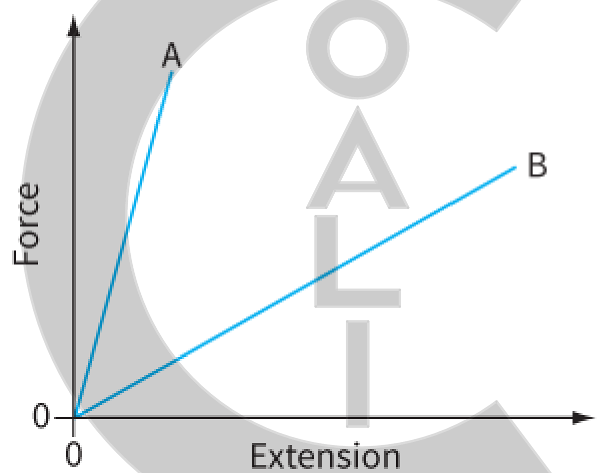


Figure 7.19: Force–extension graph for two polymers.

REFLECTION

What is the most important thing that you learned personally in this chapter?

Think of examples where having materials with high or low Young modulus is useful.

Make sure you know the formulae for stress, strain and Young modulus and can write them down in terms of *F, A, x* and *L*.

SUMMARY

Density is defined as the mass per unit volume of a substance:

$$\text{density} = \frac{\text{mass}}{\text{volume}}$$

Pressure is defined as the normal force acting per unit cross-sectional area:

$$\text{pressure} = \frac{\text{normal force}}{\text{cross-sectional area}}$$

Pressure in a fluid increases with depth: $p = \rho g h$

Upthrust on an object in a fluid is given by $F = \rho g V$ (Archimedes' principle).

Hooke's law states that the extension of a material is directly proportional to the applied force, provided the limit of proportionality is not exceeded. For a spring or a wire, $F = kx$, where k is the force constant. The force constant has units of N m^{-1}.

Stress is defined as:

$$\text{stress} = \frac{\text{force}}{\text{cross-sectional area}} \text{ or } \sigma = \frac{F}{A}$$

Strain is defined as:

$$\text{strain} = \frac{\text{extension}}{\text{original length}} \text{ or } \varepsilon = \frac{x}{L}$$

To describe the behaviour of a material under tensile and compressive forces, we have to draw a graph of stress against strain. The gradient of the initial linear section of the graph is equal to the Young modulus. The Young modulus is an indication of the stiffness of the material.

The Young modulus E is given by:

$$E = \frac{\text{stress}}{\text{strain}} = \frac{\sigma}{\varepsilon} (\text{unit: pascal (Pa) or N m}^{-2})$$

The area under a force–extension graph is equal to the work done by the force.

For a spring or a wire obeying Hooke's law, the elastic potential energy E is given by:

$$E = \frac{1}{2}Fx \text{ or } E = \frac{1}{2}kx^2$$

EXAM-STYLE QUESTIONS

1 Which force is caused by a difference in pressure? **[1]**

A drag

B friction

C upthrust

D weight

2 Two wires P and Q both obey Hooke's law. They are both stretched and have the same strain. The Young modulus of P is four times larger than that of Q. The diameter of P is twice that of Q.

What is the ratio of the tension in P to the tension in Q? **[1]**

A $\frac{1}{2}$

B 1

C 2

D 16

3 **a** **i** Define density. **[1]**

ii State the SI base units in which density is measured. **[1]**

b **i** Define pressure. **[1]**

ii State the SI base units in which pressure is measured. **[1]**

[Total: 4]

4 Sketch a force–extension graph for a spring that has a spring constant of 20 N m^{-1} and that obeys Hooke's law for forces up to 5.0 N. Your graph should cover forces between 0 and 6 N and show values on both axes. **[3]**

5 Two springs, each with a spring constant 20 N m^{-1}, are connected in series. Draw a diagram of the two springs in series and determine the total extension if a mass, with weight 2.0 N, is hung on the combined springs. **[5]**

6 A sample of fishing line is 1.0 m long and is of circular cross-section of radius 0.25 mm. When a weight is hung on the line, the extension is 50 mm and the stress is 2.0×10^8 Pa. Calculate:

a the cross-sectional area of the line **[1]**

b the weight added **[2]**

c the strain in the line **[2]**

d the Young modulus **[2]**

e the percentage and absolute uncertainties in the Young modulus if the uncertainty in the extension is ±1 mm. **[2]**

[Total: 9]

7 This is the force–extension graph for a metal wire of length 2.0 m and cross-sectional area 1.5×10^{-7} m^2.

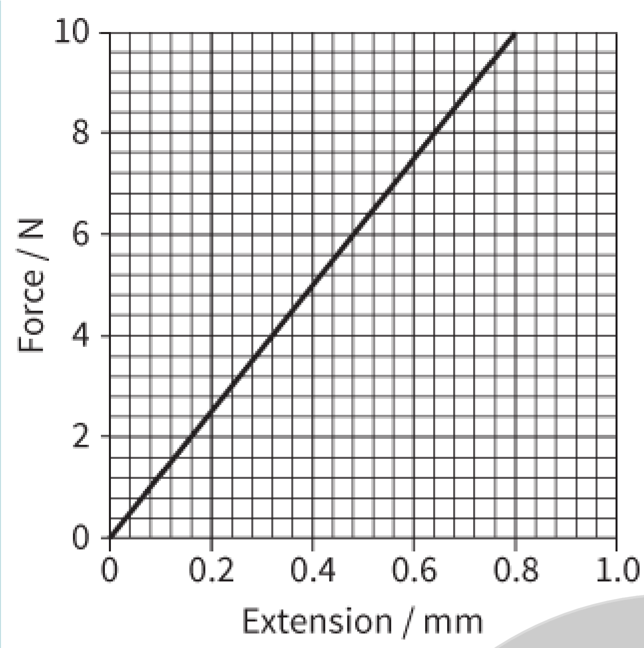


Figure 7.20

a Calculate the Young modulus. **[3]**

b Determine the energy stored in the wire when the extension is 0.8 mm. **[2]**

c Calculate the work done in stretching the wire between 0.4 mm and 0.8 mm. **[2]**

[Total: 7]

8 A piece of wax is attached to a newton-meter. In air, the reading on the newton-meter is 0.27 N and when submerged in water of density 1000 kg m^{-3} the reading is 0.16 N. Calculate:

a the upthrust on the wax when in water **[1]**

b the volume of the wax **[2]**

c the reading on the newton-meter when the wax is submerged in a liquid of density 800 kg m^{-3}. **[2]**

[Total: 5]

9 a These are stress–strain curves for three different materials, P, Q and R. State and explain which material has the greatest Young modulus. **[2]**

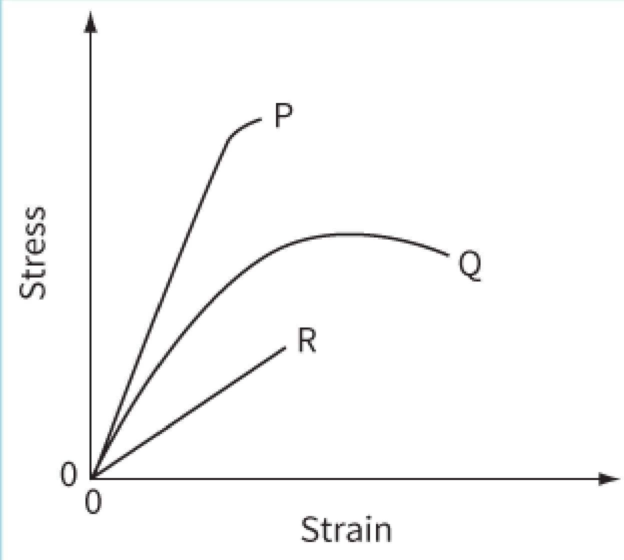


Figure 7.21

b Describe an experiment to determine the Young modulus for a material in the form of a wire. Include a labelled diagram and explain how you would make the necessary measurements. Show how you would use your measurements to calculate the Young modulus. **[7]**

[Total: 9]

10 a State the meaning of tensile stress and tensile strain. **[2]**

b A vertical steel wire of length 1.6 m and cross-sectional area 1.3×10^{-6} m^2 carries a weight of 60 N. The Young modulus for steel is 2.1×10^{11} Pa. Calculate:

i the stress in the wire **[2]**

ii the strain in the wire **[2]**

iii the extension produced in the wire by the weight. **[2]**

[Total: 6]

11 To allow for expansion in the summer when temperatures rise, a steel railway line laid in cold weather is pre-stressed by applying a force of 2.6×10^5 N to the rail of cross-sectional area 5.0×10^{-3} m^2.

If the railway line is not pre-stressed, a strain of 1.4×10^{-5} is caused by each degree Celsius rise in temperature. The Young modulus of the steel is 2.1×10^{11} Pa.

a State and explain whether the force applied to the rail when it is laid should be tensile or compressive. **[2]**

b Calculate:

i the strain produced when the rail is laid **[3]**

ii the temperature rise when the rail becomes unstressed. **[2]**

[Total: 7]

12 This is a stress–strain graph for a metal wire.

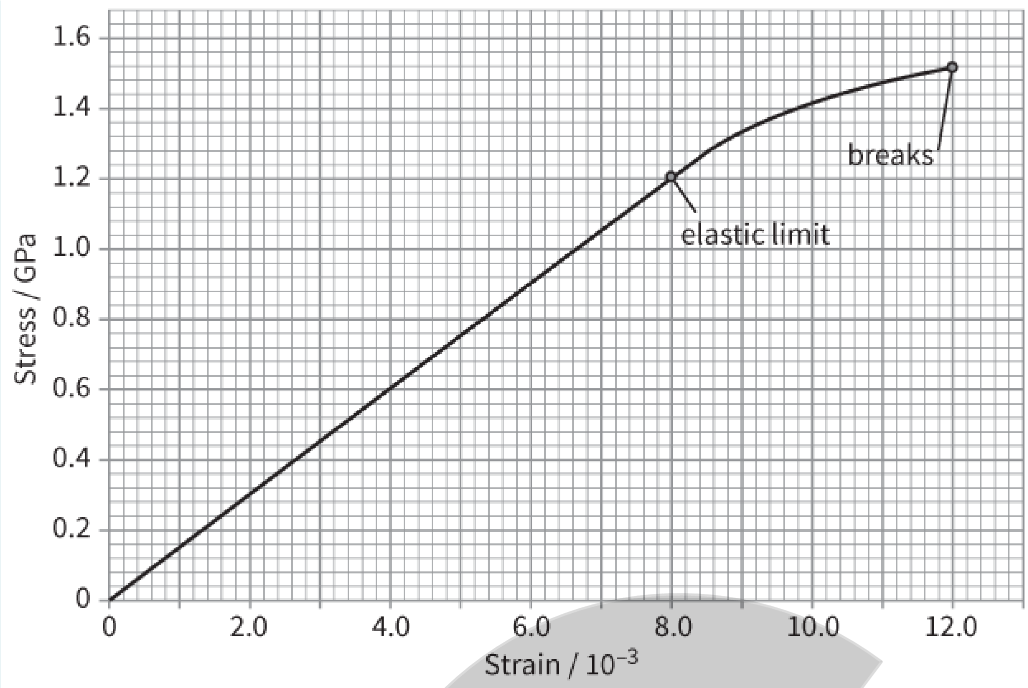


Figure 7.22

The wire has a diameter of 0.84 mm and a natural length of 3.5 m.

Use the graph to determine:

a the Young modulus of the wire **[3]**

b the extension of the wire when the stress is 0.60 GPa **[2]**

c the force that causes the wire to break, assuming that the cross-sectional area of the wire remains constant **[3]**

d the energy stored when the wire has a stress of 0.60 GPa. **[3]**

[Total: 11]

13 This is a force–extension graph for a spring.

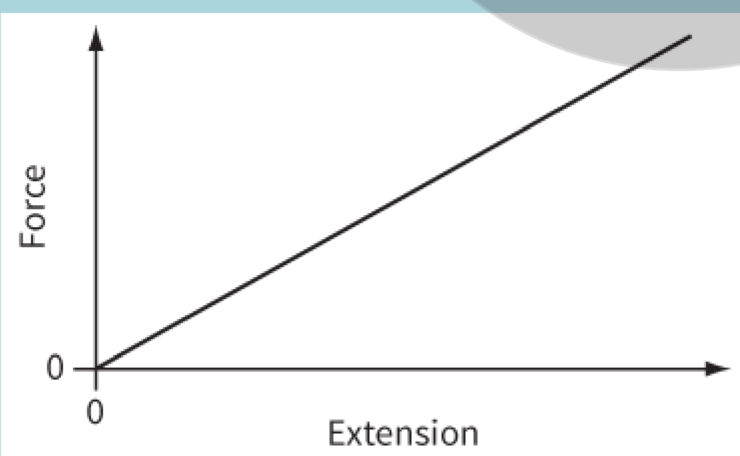


Figure 7.23

a State what is represented by:

i the gradient of the graph **[1]**

ii the area under the graph. **[1]**

b The spring has force constant $k = 80\ \text{N m}^{-1}$. The spring is compressed by 0.060 m, within the limit of proportionality, and placed between two trolleys that run on a friction-free, horizontal track. Each trolley has a mass of 0.40 kg. When the spring is released the trolleys fly apart with equal speeds but in opposite directions.

i How much energy is stored in the spring when it is compressed by 0.060 m? **[2]**

ii Explain why the two trolleys must fly apart with equal speeds. **[2]**

iii Calculate the speed of each trolley. **[2]**

[Total: 8]

14 a Liquid of density ρ fills a cylinder of base area A and height h.

i Using the symbols provided, state the mass of liquid in the container. **[1]**

ii Using your answer to **i**, derive a formula for the pressure exerted on the base of the cylinder. **[2]**

b A boy stands on a platform of area $0.050\ \text{m}^2$ and a manometer measures the pressure created in a flexible plastic container by the weight W of the boy, as shown.

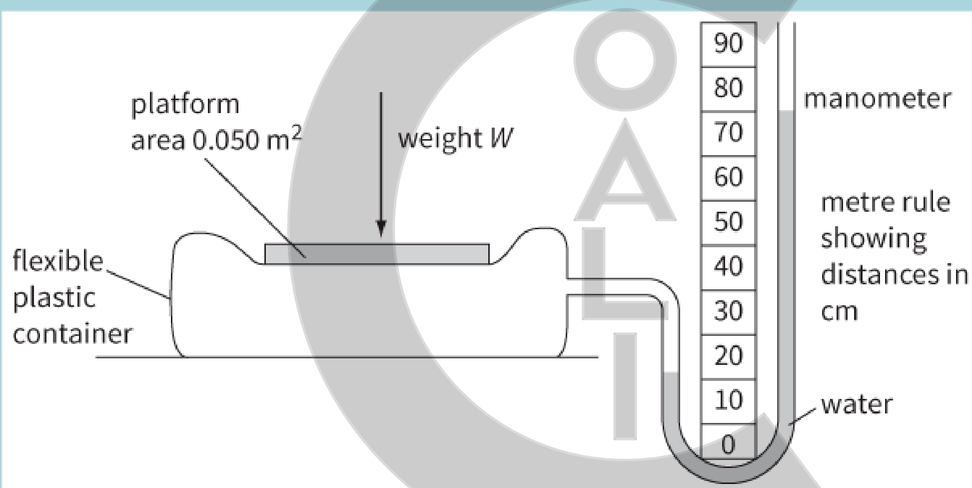


Figure 7.24

The density of water is $1000\ \text{kg m}^{-3}$. Determine:

i the pressure difference between the inside of the plastic container and the atmosphere outside **[2]**

ii the weight W of the boy. **[2]**

[Total: 7]

15 This diagram shows water in a container filled to a depth of 0.50 m. The density of water is $1000\ \text{kg m}^{-3}$.

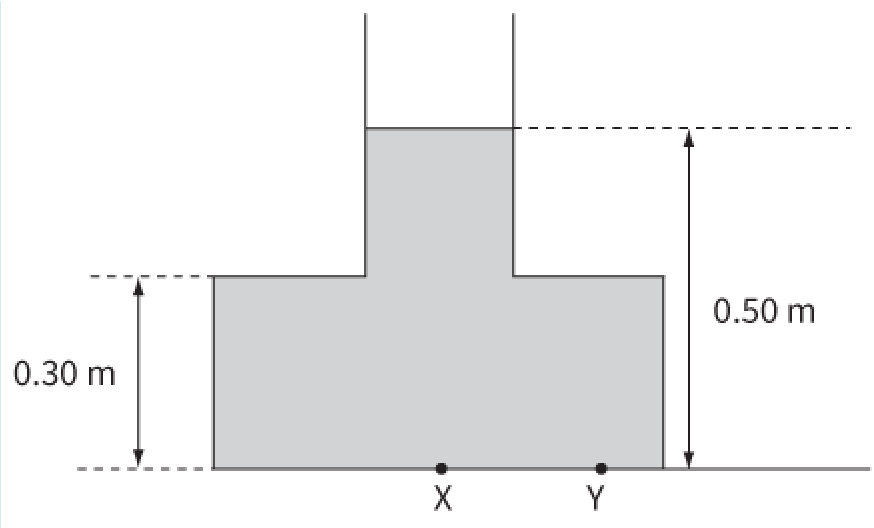


Figure 7.25

a Calculate the pressure at X on the base of the container. **[2]**

b Explain why the pressure at X must be equal to the pressure at Y. **[1]**

c Explain why the force downwards on the base of the container is larger than the weight of the liquid in the container. **[2]**

[Total: 5]

16 A light spring that obeys Hooke's law has an unstretched length of 0.250 m. When an object of mass 2.0 kg is hung from the spring the length of the spring becomes 0.280 m. When the object is fully submerged in a liquid of density 1200 kg m^{-3}, the length of the spring becomes 0.260 m.

Calculate:

a the spring constant of the spring. **[2]**

b the upthrust on the object. **[2]**

c the volume of the object. **[2]**

d the density of the material from which the object is made. **[2]**

[Total: 8]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define and use density and pressure	7.1, 7.2			
derive and use the equation $\Delta p = \rho g \Delta h$	7.2			
understand that upthrust is caused by a difference in pressure and use the equation $F = \rho g V$	7.3			
understand tensile and compressive forces, load, extension, limit of proportionality, elastic deformation, plastic deformation and elastic limit	7.4			
recall and use Hooke's law and the formula for the spring constant: $k = \frac{F}{x}$	7.4			
define and use stress, strain and the Young modulus	7.5			
describe an experiment to measure the Young modulus	7.5			
determine elastic potential or strain energy.	7.6			

〉Chapter 8

Electric current, potential difference and resistance

LEARNING INTENTIONS

In this chapter you will learn how to:

- understand of the nature of electric current
- understand the term charge and recognise its unit, the coulomb
- understand that charge is quantised
- solve problems using the equation $Q = It$
- solve problems using the formula $I = nAve$
- solve problems involving the mean drift velocity of charge carriers
- understand the terms potential difference, e.m.f. and the volt
- use energy considerations to distinguish between p.d. and e.m.f.
- define resistance and recognise its unit, the ohm
- solve problems using the formula $V = IR$
- solve problems concerning energy and power in electric circuits.

BEFORE YOU START

- Write down what you understand by the terms current, charge, potential difference, e.m.f. and resistance.
- Can you set up a simple circuit to measure the current in a lamp and the potential difference across it? Sketch the circuit and swap it with a classmate to check.

DEVELOPING IDEAS

Electricity plays a vital part in our lives. We use electricity as a way of transferring energy from place to place – for heating, lighting and making things move. For people in a developing nation, the arrival of a reliable electricity supply marks a great leap forward. In Kenya, a micro-hydroelectric scheme has been built on Kabiri Falls, on the slopes of Mount Kenya. Although this produces just 14 kW of power, it has given work to a number of people, as shown in Figures 8.1, 8.2 and 8.3.

Figure 8.1: An operator controls the water inlet at the Kabiri Falls power plant. The generator is on the right.

Figure 8.2: A metal workshop uses electrical welding equipment. This allows rapid repairs to farmers' machinery.

Figure 8.3: A hairdresser can now work in the evenings, thanks to electrical lighting.

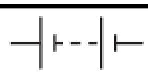
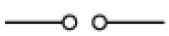
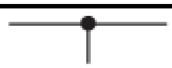
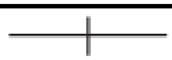
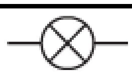
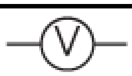
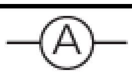
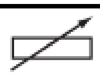
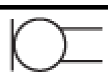
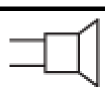
With an increasing need for small generating facilities at a town or even village level, can you suggest what types of generator would be suitable in your neighbourhood? What are the advantages and disadvantages of each type of generator?

8.1 Circuit symbols and diagrams

Before we go on to study electricity we need to introduce the concept of circuit diagrams. It is impossible to draw anything but the simplest circuits as a detailed drawing. To make it possible to draw complex circuits, a shorthand method using standard circuit symbols is used. You will have seen many circuit components and their symbols in your previous studies. Some are shown in Table 8.1 and Figure 8.4.

The symbols in Table 8.1 are a small part of a set of internationally agreed conventional symbols for electrical components. It is essential that scientists, engineers, manufacturers and others around the world use the same symbol for a particular component. In addition, many circuits are now designed by computers and these need a universal language in which to work and to present their results.

The International Electrotechnical Commission (IEC) is the body that establishes agreements on such things as electrical symbols, as well as safety standards, working practices and so on. The circuit symbols used here form part of an international standard known as IEC 60617. Because this is a shared ‘language’, there is less chance that misunderstandings will arise between people working in different organisations and different countries.

Symbol	Component name
	connecting lead
	cell
	battery of cells
	fixed resistor
	power supply
	junction of conductors
	crossing conductors (no connection)
	filament lamp
	voltmeter
	ammeter
	switch
	variable resistor
	microphone
	loudspeaker
	fuse
	earth

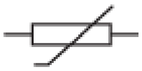
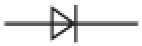
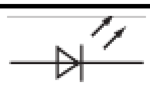
Symbol	Component name
	alternating signal
	capacitor
	thermistor
	light-dependent resistor (LDR)
	semiconductor diode
	light-emitting diode (LED)

Table 8.1: Electrical components and their circuit symbols.

What's in a word?

Electricity is a rather tricky word. In everyday life, its meaning may be rather vague – sometimes we use it to mean electric current; at other times, it may mean electrical energy or electrical power. In this chapter and the ones that follow, we will avoid using the word electricity and try to develop the correct usage of these more precise scientific terms.

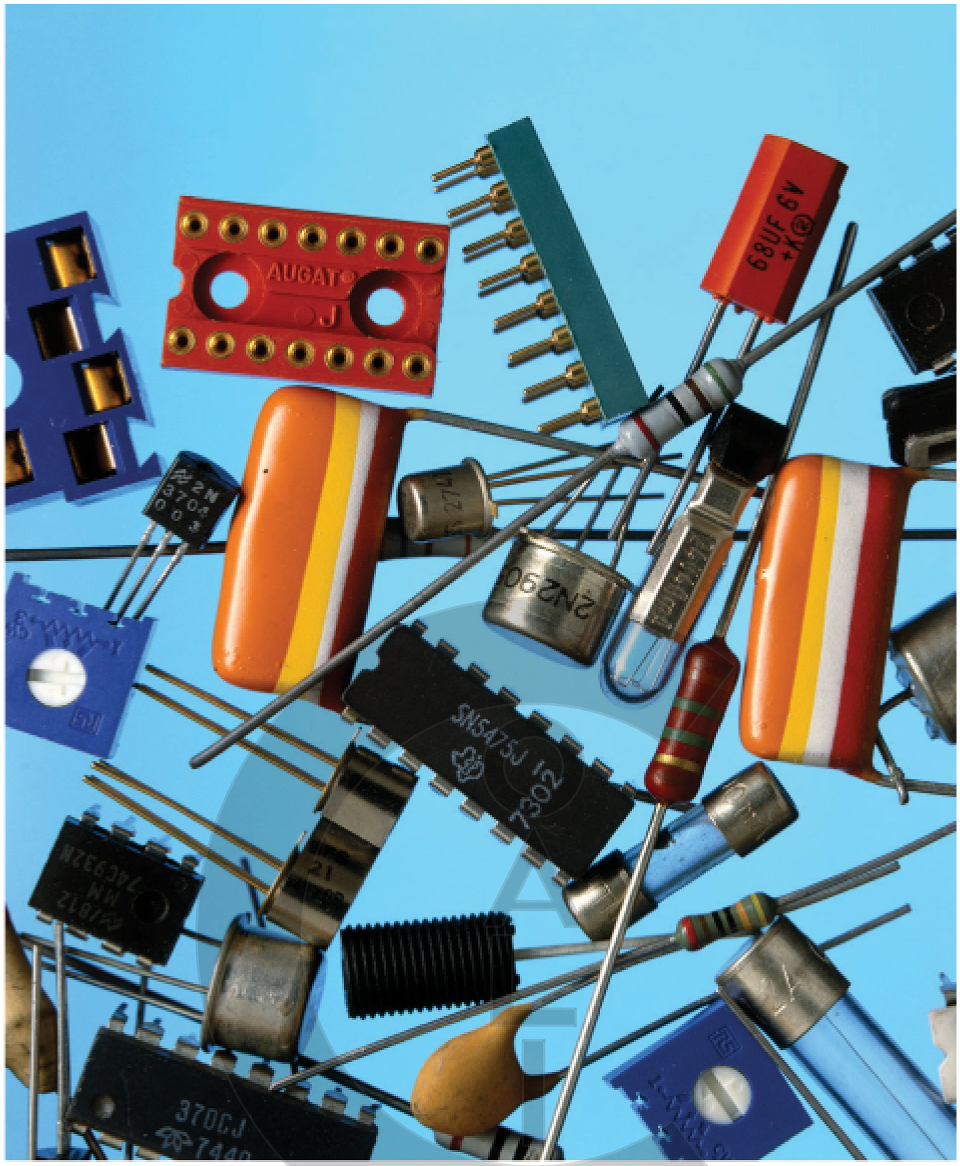


Figure 8.4: A selection of electrical components, including resistors, fuses, capacitors and microchips.

8.2 Electric current

You will have carried out many practical activities involving electric current. For example, if you connect a wire to a cell (Figure 8.5), there will be current in the wire. And of course you make use of electric currents every day of your life – when you switch on a lamp or a computer, for example.

In the circuit of Figure 8.5, the direction of the current is from the positive terminal of the cell, around the circuit to the negative terminal. This is a scientific convention: the direction of current is from positive to negative, and hence the current may be referred to as **conventional current**. But what is going on inside the wire?

A wire is made of metal. Inside a metal, there are negatively charged electrons that are free to move about. We call these **conduction** or **free** electrons, because they are the particles that allow a metal to conduct an electric current. The atoms of a metal bind tightly together; they usually form a regular array, as shown in Figure 8.6. In a typical metal, such as copper or silver, one or more electrons from each atom breaks free to become conduction electrons. The atom remains as a positively charged ion. Since there are equal numbers of free electrons (negative) and ions (positive), the metal has no overall charge – it is neutral.

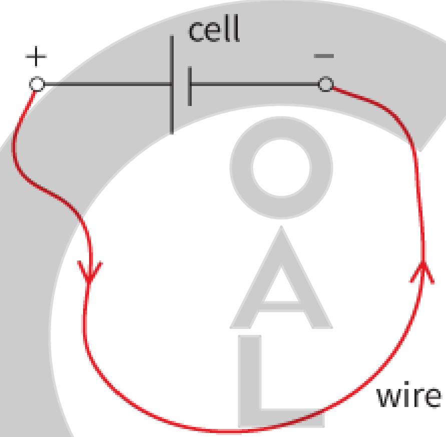


Figure 8.5: There is current in the wire when it is connected to a cell.

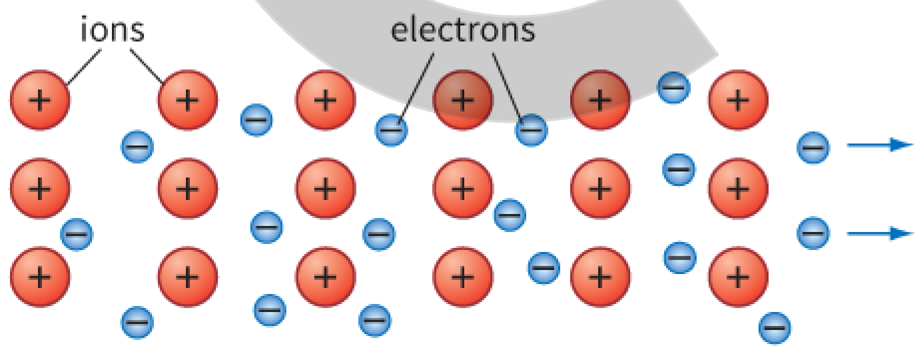


Figure 8.6: In a metal, conduction electrons are free to move among the fixed positive ions. A cell connected across the ends of the metal causes the electrons to drift towards its positive terminal.

When the cell is connected to the wire, it exerts an electrical force on the conduction electrons that makes them travel along the length of the wire. Since electrons are negatively charged, they flow away from the negative terminal of the cell and towards the positive terminal. This is in the opposite direction to conventional current. This may seem a bit strange; it happens because the direction of conventional current was chosen long before anyone had any idea what was going on inside a piece of metal carrying a current. If the names positive and negative had originally been allocated the other way round, we would now label electrons as positively charged, and conventional current and electron flow would be in the same direction.

Note that there is a current at all points in the circuit as soon as the circuit is completed. We do not have to wait for charge to travel around from the cell. This is because the charged electrons are already present throughout the metal before the cell is connected.

We can use the idea of an electric field to explain why charge flows almost instantly. Connect the terminals of a cell to the two ends of a wire and we have a complete circuit. The cell produces an electric field in the wire; the field lines are along the wire, from the positive terminal to the negative. This means that there is a force on each electron in the wire, so each electron starts to move and the current exists almost instantly.

Charge carriers

Sometimes a current is a flow of positive charges–for example, a beam of protons produced in a particle accelerator. The current is in the same direction as the particles. Sometimes a current is due to both positive and negative charges – for example, when charged particles flow through a solution. A solution that conducts is called an electrolyte and it contains both positive and negative ions. These move in opposite directions when the solution is connected to a cell (Figure 8.7). These charged particles are known as **charge carriers**. If you consider the structure of charged particles you will appreciate that charge comes in definite sized 'bits'; the smallest bit being the charge on an electron or on a single proton. This 'bittiness' is what is meant when charge is described as being **quantised**.

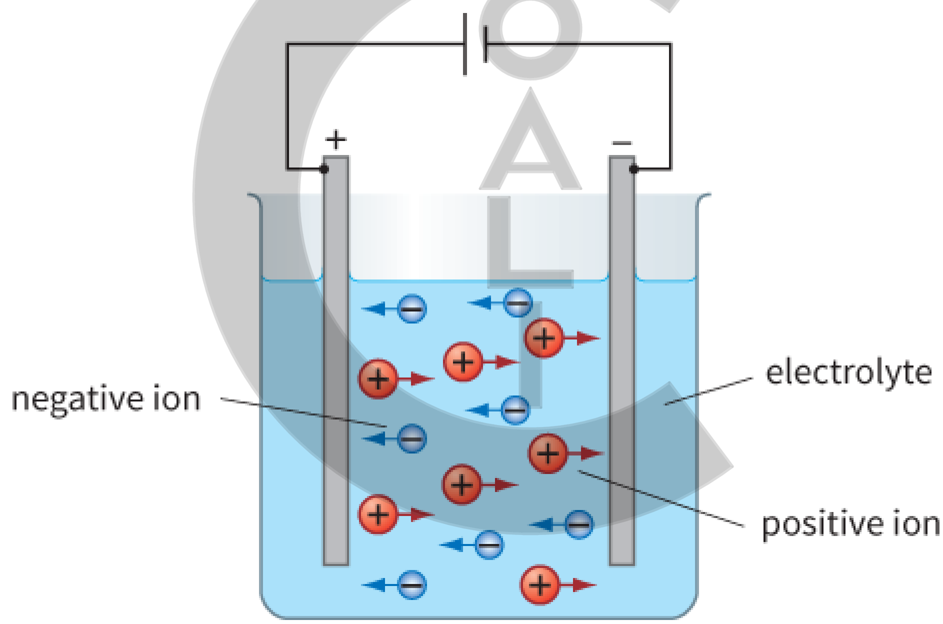


Figure 8.7: Both positive and negative charges are free to move in a solution. Both contribute to the electric current.

Questions

1 Look at Figure 8.7 and state the direction of the conventional current in the electrolyte (towards the left, towards the right or in both directions at the same time?).

2 Figure 8.8 shows a circuit with a conducting solution having both positive and negative ions.

   a Copy the diagram and draw in a cell between points A and B. Clearly indicate the positive and negative terminals of the cell.

   b Add an arrow to show the direction of the conventional current in the solution.

   c Add arrows to show the direction of the conventional current in the connecting wires.

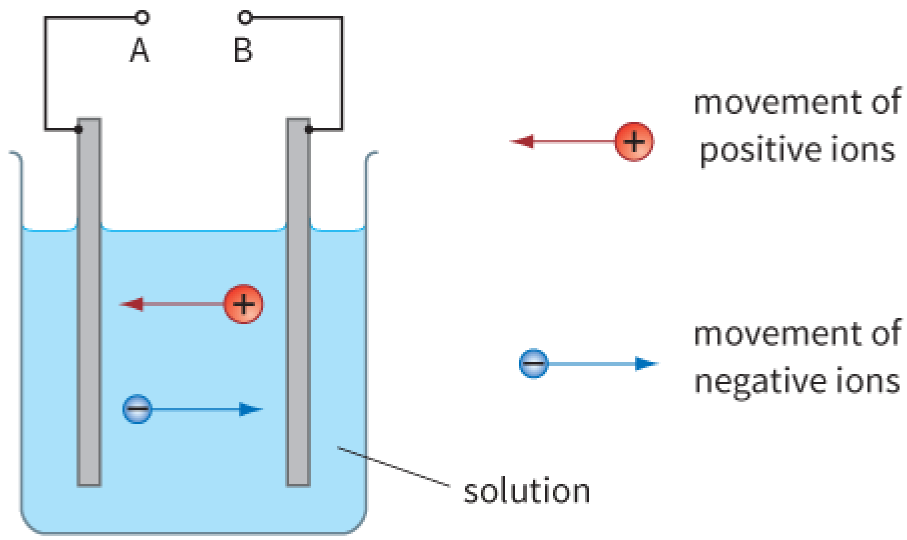


Figure 8.8: For Question 2.

Current and charge

When charged particles flow past a point in a circuit, we say that there is a **current** in the circuit. Electrical current is measured in **amperes** (A). So how much charge is moving when there is a current of 1 A? Charge is measured in **coulombs** (C). For a current of 1 A, the rate at which charge passes a point in a circuit is 1 C in a time of 1 s. Similarly, a current of 2 A gives a charge of 2 C in a time of 1 s. A current of 3 A gives a charge of 6 C in a time of 2 s, and so on. The relationship between charge, current and time may be written as the following word equation:

$$\text{current} = \frac{\text{charge}}{\text{time}}$$

This equation explains what we mean by electric current.

The equation for current can be rearranged to give an equation for charge:

$$\text{charge} = \text{current} \times \text{time}$$

KEY EQUATION

$$\text{charge} = \text{current} \times \text{time}$$
$$\Delta Q = I\Delta t$$

The unit of charge is the coulomb.

In symbols, the charge flowing past a point is given by the relationship:

$$\Delta Q = I\Delta t$$

where ΔQ is the charge that flows during a time Δt, and I is the current.

Note that the ampere and the coulomb are both SI units; the ampere is a base unit while the coulomb is a derived unit (see Chapter 3).

Questions

3 The current in a circuit is 0.40 A. Calculate the charge that passes a point in the circuit in a period of 15 s.

4 Calculate the current that gives a charge flow of 150 C in a time of 30 s.

5 In a circuit, a charge of 50 C passes a point in 20 s. Calculate the current in the circuit.

6 A car battery is labelled '50 A h'. This means that it can supply a current of 50 A for one hour.

   a For how long could the battery supply a continuous current of 200 A needed to start the car?

   b Calculate the charge that flows past a point in the circuit in this time.

Charged particles

As we have seen, current is the flow of charged particles called charge carriers. But how much charge does each particle carry?

Electrons each carry a tiny negative charge of approximately -1.6×10^{-19} C. This charge is represented by $-e$. The magnitude of the charge is known as the **elementary charge**. This charge is so tiny that you would need about six million million million electrons – that's 6 000 000 000 000 000 000 of them – to have a charge equivalent to one coulomb.

Protons are positively charged, with a charge $+e$. This is equal and opposite to that of an electron. Ions carry charges that are multiples of $+e$ and $-e$.

WORKED EXAMPLES

1 There is a current of 10 A through a lamp for 1.0 hour. Calculate how much charge flows through the lamp in this time.

Step 1 We need to find the time t in seconds:

$\Delta t = 60 \times 60 = 3600$ s

Step 2 We know the current $I = 10$ A, so the charge that flows is:

$\Delta Q = I\Delta t = 10 \times 3600 = 36\,000 \text{ C} = 3.6 \times 10^4 \text{ C}$

2 Calculate the current in a circuit when a charge of 180 C passes a point in a circuit in 2.0 minutes.

Step 1 Rearranging $\Delta Q = I\Delta t$ gives:

$$I = \frac{\Delta Q}{\Delta t} = \frac{\text{charge}}{\text{time}}$$

Step 2 With time in seconds, we then have:

$$\text{current } I = \frac{180}{120} = 1.5 \text{ A}$$

Because electric charge is carried by particles, it must come in amounts that are multiples of e. So, for example, 3.2×10^{-19} C is possible, because this is $+2e$, but 2.5×10^{-19} C is impossible, because this is not an integer multiple of e.

This reinforces the idea that charge is quantised; it means that it can only come in amounts that are integer multiples of the elementary charge. If you are studying chemistry, you will know that ions have charges of $\pm e$, $\pm 2e$, etc. The only exception is in the case of the fundamental particles called quarks, which are the building blocks from which particles such as protons and neutrons are made. These have charges of $\pm \frac{1}{3}e$ or $\pm \frac{2}{3}e$. However, quarks always appear in twos or threes in such a way that their combined charge is zero or a multiple of e.

Questions

7 Calculate the number of protons that would have a charge of one coulomb. (Proton charge = $+1.6 \times 10^{-19}$ C.)

8 Which of the following quantities of electric charge is possible? Explain how you know.

6.0×10^{-19} C, 8.0×10^{-19} C, 10.0×10^{-19} C

8.3 An equation for current

Copper, silver and gold are good conductors of electric current. There are large numbers of conduction electrons in a copper wire – as many conduction electrons as there are atoms. The number of conduction electrons per unit volume (for example, in 1 m^3 of the metal) is called the **number density** and has the symbol *n*. For copper, the value of *n* is about 10^{29} m^{-3}.

Figure 8.9 shows a length of wire, with cross-sectional area *A*, along which there is a current *I*.

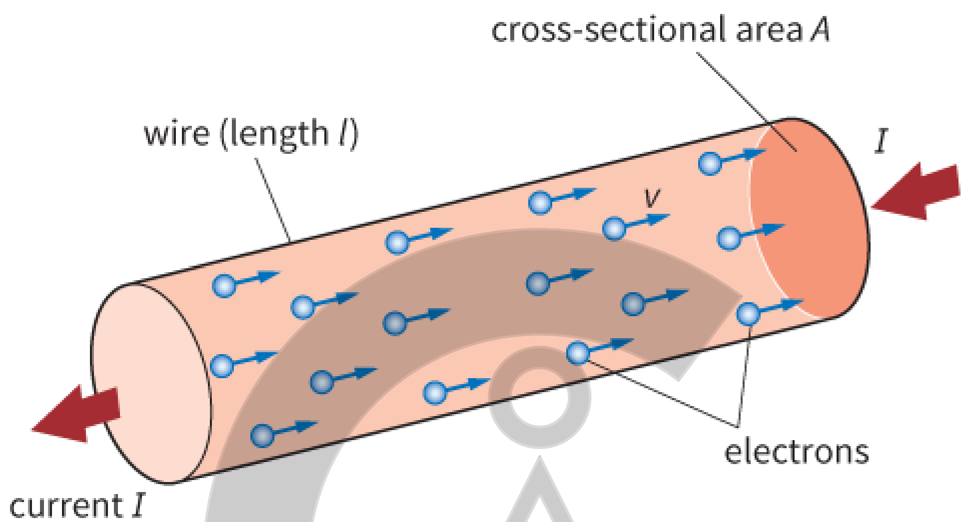


Figure 8.9: A current *I* in a wire of cross-sectional area *A*. The charge carriers are mobile conduction electrons with **mean drift velocity** *v*.

How fast do the electrons in Figure 8.9 have to travel? The following equation allows us to answer this question:

$$I = nAvq$$

where *n* = the number density, *A* = cross sectional area of the conductor, *v* = mean drift velocity of the charge carriers, *q* = the charge on each charge carrier.

KEY EQUATION

Electric current:

$$I = nAvq$$

The length of the wire in Figure 8.9 is *l*. We imagine that all of the electrons shown travel at the same speed *v* along the wire.

Now imagine that you are timing the electrons to determine their speed. You start timing when the first electron emerges from the right-hand end of the wire. You stop timing when the last of the electrons shown in the diagram emerges. (This is the electron shown at the left-hand end of the wire in the diagram.) Your timer shows that this electron has taken time *t* to travel the distance *l*.

In the time *t*, all of the electrons in the length *l* of wire have emerged from the wire. We can calculate how many electrons this is, and hence the charge that has flowed in time *t*:

$$\text{number of electrons} = \text{number density} \times \text{volume of wire}$$
$$= n \times A \times l$$

$$\text{charge of electrons} = \text{number} \times \text{electron charge}$$
$$= n \times A \times l \times e$$

We can find the current *I* because we know that this is the charge that flows in time *t*, and $\text{current} = \frac{\text{charge}}{\text{time}}$:

$$I = n \times A \times l \times \frac{e}{t}$$

Substituting *v* for $\frac{l}{t}$ gives:

$$I = nAve$$

The moving charge carriers that make up a current are not always electrons. They might, for example, be ions (positive or negative) whose charge *q* is a multiple of *e*. Hence we can write a more general version of the equation as:

$$I = nAvq$$

Worked example 3 shows how to use this equation to calculate a typical value of *v*.

WORKED EXAMPLE

3 Calculate the mean drift velocity of the electrons in a copper wire of cross-sectional area $5.0 \times 10^{-6}\ \text{m}^2$ carrying a current of 1.0 A. The electron number density for copper is $8.5 \times 10^{28}\ \text{m}^{-3}$.

Step 1 Rearrange the equation *I* = *nAve* to make *v* the subject:

$$v = \frac{I}{nAe}$$

Step 2 Substitute values and calculate *v*:

$$v = \frac{1.0}{8.5 \times 10^{28} \times 5.0 \times 10^{-6} \times 1.6 \times 10^{-19}}$$
$$= 1.47 \times 10^{-5}\ \text{m s}^{-1}$$
$$= 0.015\ \text{mm s}^{-1}$$

You do not need to know how to derive *I* = *nAvq* but it is interesting to recognise that the units are **homogeneous**.

The unit of current (*I*) is the ampere (A).

The unit of the number of charge carriers per unit volume (*n*) is m^{-3}.

The unit of area (*A*) is m^2.

The unit of the drift velocity *v* is m s^{-1}.

The unit of charge (*q*) is the coulomb (C).

All these are in base units except the coulomb and 1 coulomb is 1 ampere second (A s).

Putting the units into the right-hand side of the equation:

$$\text{m}^{-3} \times \text{m}^2 \times \text{m s}^{-1} \times \text{A s} = \text{A}$$

This is the same as the left-hand side of the equation. Although this does not prove the equation to be correct, it does give strong evidence for it.

This technique is often used for checking the validity of an expression and also to predict a possible formula.

WORKED EXAMPLE

4 A student knows that the power transfer in a resistor depends on two variables: the current and the resistance of the resistor, but is unsure of the precise nature of the relationships.

Suggest the form of the equation.

Step 1 Identify the units of the terms:

power–watt (W)

current–ampere (A)

resistance–ohms (Ω)

Step 2 Break these down into base units:

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

hence $1\ W = 1\ kg\ m\ s^{-2}\ m\ s^{-1} = 1\ kg\ m^2\ s^{-3}$

$1\ \Omega = 1\ V\ A^{-1}$, $1\ V = 1\ J\ C^{-1}$, $1\ J = 1\ kg\ m^2\ s^{-2}$, $1\ C = 1\ A\ s$

thus $1\ \Omega = 1\ kg\ m^2\ s^{-2}\ [A\ s]^{-1}\ A^{-1} = 1\ kg\ m^2\ s^{-2}\ A^{-2}\ s^{-1} = 1\ kg\ m^2\ s^{-3}\ A^{-2}$

The unit for current, the ampere is already a base unit.

Step 3 Write a possible equation linking power, current and resistance

$P = K\ I^p R^q$ where p and q are pure numbers and K is a dimensionless constant.

The units on the right-hand side of this equation are $A^p \times [1\ kg\ m^2\ s^{-3}\ A^{-2}]^q$

The units on the left-hand side of the equation are $1\ kg\ m^2\ s^{-3}$

By inspection, we can see that the amperes do not figure in the left-hand side of the equation, thus p must equal 2 to cancel the amperes on the right-hand side and q must equal 1. This would leave both sides of the equation as $1\ kg\ m^2\ s^{-3}$.

Therefore we know that the equation is of the form:

$P = K\ I^2 R$

Note that this method does not give any information of the value of the constant K, although in this case, from our choice of units, K = 1 and the equation becomes the familiar $P = I^2R$.

Slow flow

It may surprise you to find that, as suggested by the result of Worked example 3, electrons in a copper wire drift at a fraction of a millimetre per second. To understand this result fully, we need to closely examine how electrons behave in a metal. The conduction electrons are free to move around inside the metal. When the wire is connected to a battery or an external power supply, each electron within the metal experiences an electrical force that causes it to move towards the positive end of the battery. The electrons randomly collide with the fixed but vibrating metal ions. Their journey along the metal is very haphazard. The actual velocity of an electron between collisions is of the order of magnitude $10^5\ m\ s^{-1}$, but its haphazard journey causes it to have a drift velocity towards the positive end of the battery. Since there are billions of electrons, we use the term mean drift velocity v of the electrons.

Figure 8.10 shows how the mean drift velocity of electrons varies in different situations.

We can understand this using the equation:

$$v = \frac{I}{nAe}$$

- If the current increases, the drift velocity *v* must increase. That is:

$$v \propto I$$

- If the wire is thinner, the electrons move more quickly for a given current. That is:

$$v \propto \frac{I}{A}$$

- There are fewer electrons in a thinner piece of wire, so an individual electron must travel more quickly.
- In a material with a lower density of electrons (smaller *n*), the mean drift velocity must be greater for a given current. That is:

$$v \propto \frac{1}{n}$$

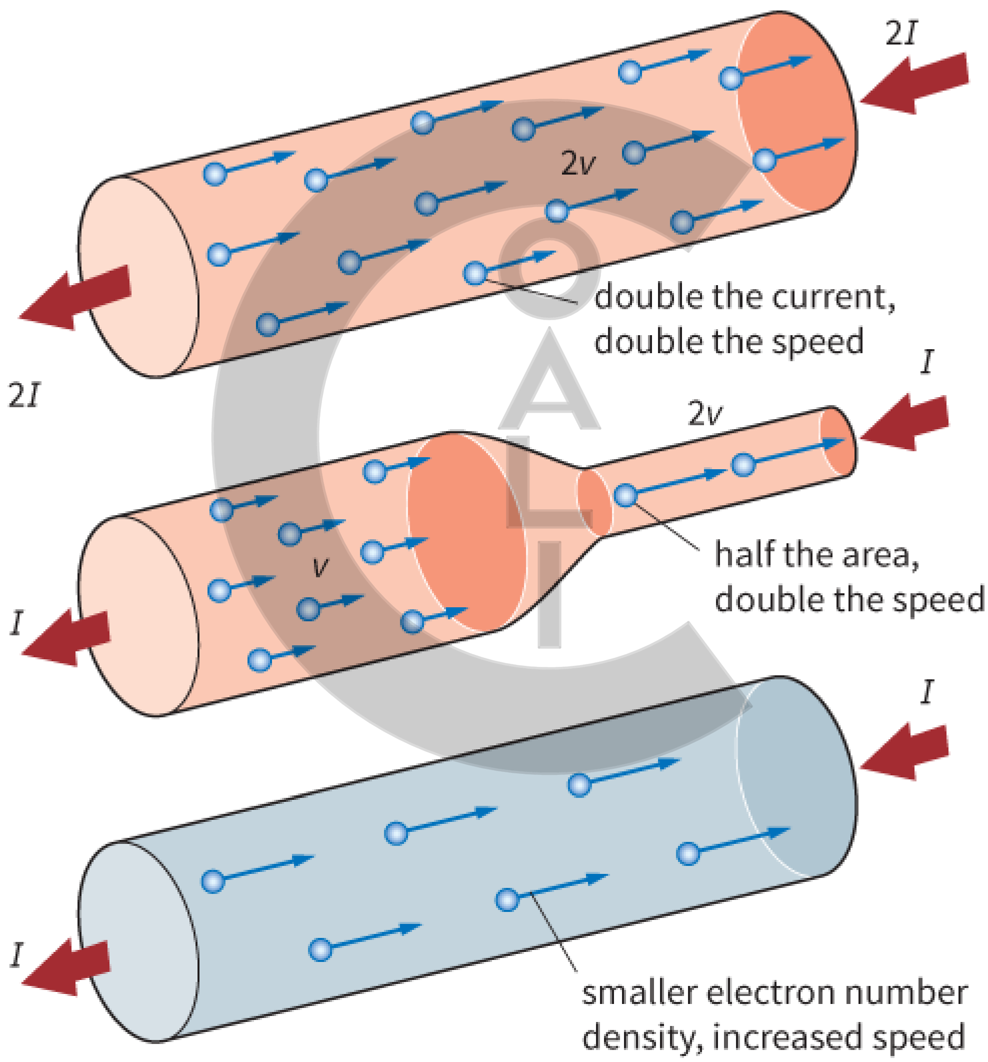


Figure 8.10: The mean drift velocity of electrons depends on the current, the cross-sectional area and the electron density of the material.

Questions

9 Calculate the current in a gold wire of cross-sectional area 2.0 mm^2 when the mean drift velocity of the electrons in the wire is 0.10 mm s^{-1}. The electron number density for gold is 5.9×10^{28} m^{-3}.

10 Calculate the mean drift velocity of electrons in a copper wire of diameter 1.0 mm with a current of 5.0 A. The electron number density for copper is 8.5×10^{28} m^{-3}.

11 A length of copper wire is joined in series to a length of silver wire of the same diameter. Both wires have a current in them when connected to a battery. Explain how the mean drift velocity of the electrons will change as they travel from the copper into the silver. Electron number densities:

copper $n = 8.5 \times 10^{28}$ m^{-3}

silver $n = 5.9 \times 10^{28}$ m^{-3}.

It may help you to picture how the drift velocity of electrons changes by thinking about the flow of water in a river. For a high rate of flow, the water moves fast – this corresponds to a greater current I. If the course of the river narrows, it speeds up – this corresponds to a smaller cross-sectional area A.

Metals have a high electron number density–typically of the order of 10^{28} or 10^{29} m^{-3}. Semiconductors, such as silicon and germanium, have much lower values of n–perhaps 10^{23} m^{-3}. In a semiconductor, electron mean drift velocities are typically a million times greater than those in metals for the same current. Electrical insulators, such as rubber and plastic, have very few conduction electrons per unit volume to act as charge carriers.

8.4 The meaning of voltage

The term **voltage** is often used in a rather casual way. In everyday life, the word is used in a less scientific and often incorrect sense – for example, 'A big voltage can go through you and kill you.' In this topic, we will consider a bit more carefully just what we mean by voltage and potential difference in relation to electric circuits.

Look at the simple circuit in Figure 8.11. Assume the power supply has negligible internal resistance. (We look at internal resistance later in Chapter 10). The three voltmeters are measuring three voltages or potential differences. With the switch open, the voltmeter placed across the supply measures 12 V. With the switch closed, the voltmeter across the power supply still measures 12 V and the voltmeters placed across the resistors measure 8 V and 4 V. You will not be surprised to see that the voltage across the power supply is equal to the sum of the voltages across the resistors.

Earlier in this chapter we saw that electric current is the rate of flow of electric charge. Figure 8.12 shows the same circuit as in Figure 8.11, but here we are looking at the movement of one coulomb (1 C) of charge round the circuit. Electrical energy is transferred to the charge by the power supply. The charge flows round the circuit, transferring some of its electrical energy to internal energy in the first resistor, and the rest to internal energy in the second resistor.

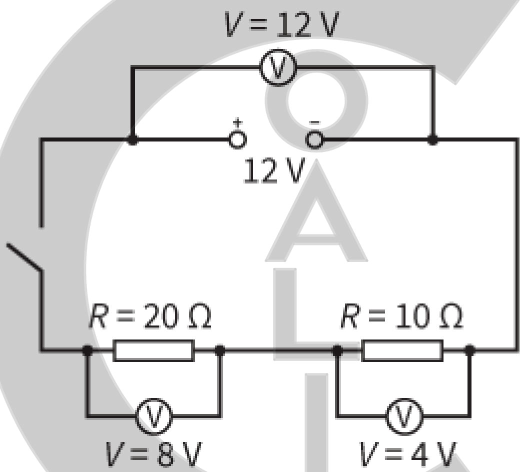


Figure 8.11: Measuring voltages in a circuit. Note that each voltmeter is connected across the component.

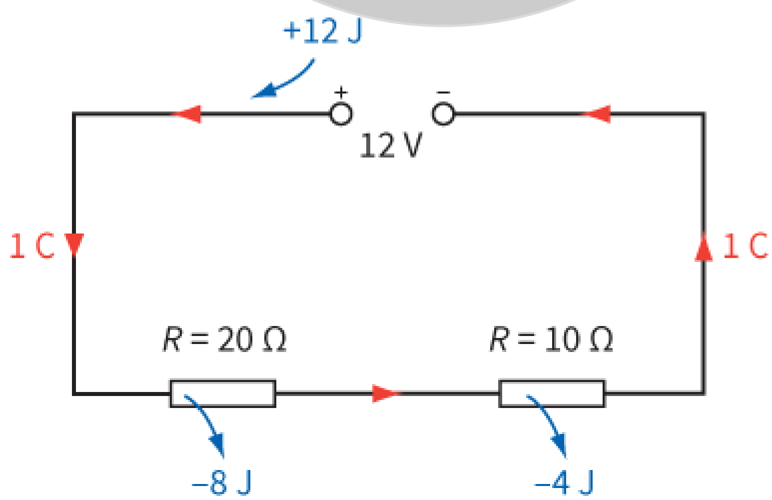


Figure 8.12: Energy transfers as 1 C of charge flows round a circuit. This circuit is the same as that shown in Figure 8.11.

The voltmeter readings indicate the energy transferred to the component by each unit of charge. The voltmeter placed across the power supply measures the e.m.f. of the supply, whereas the voltmeters placed across the resistors measure the potential difference (p.d.) across these components. The terms e.m.f. and potential difference have different meanings, so you have to be very vigilant.

The term **potential difference** is used when charges **lose** energy by transferring electrical energy to other forms of energy in a component, such as thermal energy or kinetic energy. Potential difference, *V*, is defined as the energy transferred per unit charge.

The potential difference between two points, A and B, is the energy transferred per unit charge as it moves from point A to point B.

$$\text{potential difference} = \frac{\text{energy transferred}}{\text{charge}} \equiv V = \frac{\Delta W}{\Delta Q}$$

This equation can be rearranged to calculate the energy transferred in a component:

$$\Delta W = V \Delta Q$$

KEY EQUATION

$$\text{energy transferred} = \text{potential difference} \times \text{charge}$$
$$\Delta W = V \Delta Q$$

A power supply or a battery transfers energy to electrical charges in a circuit. The electromotive force (**e.m.f.**), *E*, of the supply is also defined as the energy transferred per unit charge. However, this refers to the energy given to the charge by the supply. The e.m.f. of a source is the energy transferred per unit charge in driving charge around a complete circuit.

Note that e.m.f. stands for electromotive *force*. This is a misleading term. It has nothing at all to do with force. This term is a legacy from the past and we are stuck with it! It is best to forget where it comes from and simply use the term e.m.f.

8.5 Electrical resistance

If you connect a lamp to a battery, a current in the lamp causes it to glow. But what determines the size of the current? This depends on two factors:

- the potential difference or voltage *V* across the lamp – the greater the potential difference, the greater the current for a given lamp
- the resistance *R* of the lamp – the greater the resistance, the smaller the current for a given potential difference.

Now we need to think about the meaning of **electrical resistance**. The resistance of any component is defined as the ratio of the potential difference to the current.

This is written as:

$$\text{resistance} = \frac{\text{potential difference}}{\text{current}} = R = \frac{V}{I}$$

where *R* is the resistance of the component, *V* is the potential difference across the component and *I* is the current in the component.

KEY EQUATION

$$\text{resistance} = \frac{\text{potential difference}}{\text{current}}$$

$$R = \frac{V}{I}$$

You can rearrange the equation to give:

$$I = \frac{V}{R} \text{ or } V = IR$$

Table 8.2 summarises these quantities and their units.

Defining the ohm

The unit of resistance, the **ohm** (Ω), can be determined from the equation that defines resistance:

$$\text{resistance} = \frac{\text{potential difference}}{\text{current}}$$

The ohm is equivalent to 1 volt per ampere:

$$1\ \Omega = 1\ \text{V A}^{-1}$$

Quantity	Symbol for quantity	Unit	Symbol for unit
current	*I*	ampere (amp)	A
voltage (p.d., e.m.f.)	*V*	volt	V
resistance	*R*	ohm	Ω

Table 8.2: Basic electrical quantities, their symbols and SI units. Take care to understand the difference between V (in italics) meaning the quantity voltage and V meaning the unit volt.

Questions

12 A car headlamp bulb has a resistance of 36 Ω. Calculate the current in the lamp when connected to a '12 V' battery.

13 You can buy lamps of different brightness to fit in light fittings at home (Figure 8.13). A '100 watt' lamp glows more brightly than a '60 watt' lamp. Explain which of the lamps has the higher resistance.

14 **a** Calculate the potential difference across a motor carrying a current of 1.0 A and having a resistance of 50 Ω.

b Calculate the potential difference across the same motor when the current is doubled. Assume its resistance remains constant.

15 Calculate the resistance of a lamp carrying a current of 0.40 A when connected to a 230 V supply.

Figure 8.13: Both of these lamps work from the 230 V mains supply, but one has a higher resistance than the other. For Question 13.

WORKED EXAMPLE

5 Calculate the current in a lamp given that its resistance is 15 Ω and the potential difference across its ends is 3.0 V.

Step 1 Here we have V = 3.0 V and R = 15 Ω.

Step 2 Substituting in $I = \frac{V}{R}$ gives:

$$\text{current } I = \frac{3.0}{15} = 0.20 \text{ A}$$

So the current in the lamp is 0.20 A.

PRACTICAL ACTIVITY 8.1

Determining resistance

As we have seen, the equation for resistance is:

$$R = \frac{V}{I}$$

To determine the resistance of a component, we therefore need to measure both the potential difference *V* across it and the current *I* through it. To measure the current, we need an ammeter. To measure the potential difference, we need a voltmeter. Figure 8.14 shows how these meters should be connected to determine the resistance of a metallic conductor, such as a length of wire.

- The ammeter is connected **in series** with the conductor, so that there is the same current in both.
- The voltmeter is connected across (**in parallel** with) the conductor, to measure the potential difference across it.

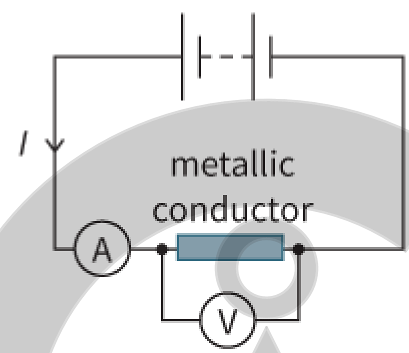


Figure 8.14: Connecting an ammeter and a voltmeter to determine the resistance of a metallic conductor in a circuit.

Question

16 In Figure 8.14 the reading on the ammeter is 2.4 A and the reading on the voltmeter is 6.0 V. Calculate the resistance of the metallic conductor.

8.6 Electrical power

The rate at which energy is transferred is known as power. Power P is measured in watts (W). (If you are not sure about this, refer back to Chapter 5, where we looked at the concept of power in relation to forces and work done.)

$$\text{power} = \frac{\text{energy transferred}}{\text{time taken}} \equiv P = \frac{\Delta W}{\Delta t}$$

where P is the power and ΔW is the energy transferred in a time Δt.

Take care not to confuse W for energy transferred or work done with W for watts.

Refer back to the equation derived from the definition of potential difference:

$$V = \frac{\Delta W}{\Delta Q}$$

This can be rearranged as:

$$\Delta W = V\,\Delta Q$$

Thus:

$$P = \frac{W}{\Delta t} = \frac{V\,\Delta Q}{\Delta t} = V\left(\frac{\Delta Q}{\Delta t}\right)$$

The ratio of charge to time, $\frac{\Delta Q}{\Delta t}$, is the current I in the component. Therefore:

$$P = VI$$

By substituting from the resistance equation $V = IR$, we get the alternative equations for power:

$$P = I^2R \text{ and } P = \frac{V^2}{R}$$

KEY EQUATIONS

Equations for power:

$$P = VI$$

$$P = I^2R$$

$$P = \frac{V^2}{R}$$

WORKED EXAMPLE

6 Calculate the rate at which energy is transferred by a 230 V mains supply that provides a current of 8.0 A to an electric heater.

Step 1 Use the equation for power:

$P = VI$

with V = 230 V and I = 8.0 A.

Step 2 Substitute values:

$P = 8 \times 230 = 1840$ W (1.84 kW)

7 a A power station produces 20 MW of power at a voltage of 200 kV. Calculate the current supplied to the grid cables.

Step 1 Here we have *P* and *V* and we have to find *I*, so we can use $P = VI$.

Step 2 Rearranging the equation and substituting the values we know gives:

$$\text{current } I = \frac{P}{V} = \frac{20 \times 10^6}{200 \times 10^3} = 100 \text{ A}$$

Hint: *Remember to convert megawatts into watts and kilovolts into volts.*

So, the power station supplies a current of 100 A.

b The grid cables are 15 km long, with a resistance per unit length of 0.20 Ω km^{-1}. How much power is wasted as heat in these cables?

Step 1 First, we must calculate the resistance of the cables:

resistance $R = 15 \text{ km} \times 0.20\ \Omega\ \text{km}^{-1} = 3.0\ \Omega$

Step 2 Now we know *I* and *R* and we want to find *P*. We can use $P = I^2R$:

power wasted as heat,

$$P = I^2R = (100)^2 \times 3.0$$
$$= 3.0 \times 10^4 \text{ W}$$
$$= 30 \text{ kW}$$

Hence, of the 20 MW of power produced by the power station, 30 kW is wasted – just 0.15%.

8 A bathroom heater, when connected to a 230 V supply has an output power of 1.0 kW. Calculate the resistance of the heater.

Step 1 We have *P* and *V* and have to find *R*, so we can use $P = \frac{V^2}{R}$

Step 2 Rearrange the equation and substitute in the known values:

$$\text{resistance } R = \frac{V^2}{P} = \frac{230^2}{1000} = 53\Omega$$

Note: *The kilowatts were converted to watts in a similar way to the previous example.*

Questions

17 Calculate the current in a 60 W light bulb when it is connected to a 230 V power supply.

18 A power station supplies electrical energy to the grid at a voltage of 25 kV. Calculate the output power of the station when the current it supplies is 40 kA.

Power and resistance

A current *I* in a resistor of resistance *R* transfers energy to it. The resistor dissipates energy heating the resistor and the surroundings.. The p.d. *V* across the resistor is given by $V = IR$. Combining this with the equation for power, $P = VI$, gives us two further forms of the equation for power dissipated in the resistor:

$$P = I^2R$$
$$P = \frac{V^2}{R}$$

Which form of the equation we use in any particular situation depends on the information we have available to us. This is illustrated in Worked examples 7a and 7b, which relate to a power station and to the grid cables that lead from it (Figure 8.15).

Figure 8.15: A power station and electrical transmission lines. How much electrical power is lost as heat in these cables? (See Worked examples 7a and 7b.)

Questions

19 A calculator is powered by a 3.0 V battery. The calculator's resistance is 20 kΩ. Calculate the power transferred to the calculator.

20 An energy-efficient light bulb is labelled '230 V, 15 W'. This means that when connected to the 230 V mains supply it is fully lit and changes electrical energy to heat and light at the rate of 15 W. Calculate:

- **a** the current in the bulb when fully lit
- **b** its resistance when fully lit.

21 Calculate the resistance of a 100 W light bulb that draws a current of 0.43 A from a power supply.

Calculating energy

We can use the relationship for power as energy transferred per unit time and the equation for electrical power to find the energy transferred in a circuit.

Since:

$$\text{power} = \text{current} \times \text{voltage}$$

and:

$$\text{energy} = \text{power} \times \text{time}$$

we have:

$$\text{energy transferred} = \text{current} \times \text{voltage} \times \text{time}$$
$$W = IV\Delta t$$

Working in SI units, this gives energy transferred in joules.

Questions

22 A 12 V car battery can supply a current of 10 A for 5.0 hours. Calculate how many joules of energy the battery transfers in this time.

23 A lamp is operated for 20 s. The current in the lamp is 10 A. In this time, it transfers 400 J of energy to the lamp. Calculate:

- **a** how much charge flows through the lamp
- **b** how much energy each coulomb of charge transfers to the lamp
- **c** the p.d. across the lamp.

REFLECTION

Without referring back to your textbook, explain to a classmate the difference between potential difference and electromotive force.

A common error is to think that the higher the resistance between two points, the greater the power output. Explain to someone, without using mathematics, why this is incorrect.

As you look at this activity, what is one thing you would like to change?

SUMMARY

Electric current is the rate of flow of charge. In a metal, the charge is electrons; in an electrolyte, it is both positive and negative ions.

The direction of conventional current is from positive to negative; because electrons are negative, they move in the opposite direction.

The SI unit of charge is the coulomb (C). One coulomb is the charge passing a point when there is a current of one ampere at that point for one second:

$$\text{charge} = \text{current} \times \text{time} \ (\Delta Q = I\Delta t)$$

The current *I* in a conductor of cross-sectional area *A* depends on the mean drift velocity (*v*) of the charge carriers and the number density (*n*):

$$I = nAvQ$$

The term potential difference is used when charge transfers energy to the component or the surroundings. It is defined as energy transferred per unit charge:

$$V = \frac{\Delta W}{\Delta Q} \text{ or } \Delta W = V\Delta Q$$

The term electromotive force is used when describing the maximum energy per unit charge that a source can provide:

$$E = \frac{\Delta W}{\Delta Q} \text{ or } \Delta W = E\Delta Q$$

A volt is a joule per coulomb (1 J C^{-1}).

Power is the energy transferred per unit time. There are three formulae to calculate power used according to the quantities that are given:

$$P = VI \text{ or } P = I^2R \text{ or } P = \frac{V^2}{R}$$

Resistance is the ratio of voltage to current:

$$R = \frac{V}{I}$$

The resistance of a component is 1 ohm when the voltage of 1 V produces a current of 1 ampere.

Energy transferred in the circuit in a time Δt is given by the equation:

$$PV = \frac{\Delta W}{\Delta Q} \text{ or } \Delta W = V\Delta Q$$

$$W = IV\Delta t$$

COALI

EXAM-STYLE QUESTIONS

1 A small immersion heater is connected to a power supply of e.m.f. of 12 V for a time of 150 s. The output power of the heater is 100 W.

What charge passes through the heater? **[1]**

A 1.4 C

B 8.0 C

C 1250 C

D 1800 C

2 Which statement defines e.m.f.? **[1]**

A The e.m.f. of a source is the energy transferred when charge is driven through a resistor.

B The e.m.f. of a source is the energy transferred when charge is driven round a complete circuit.

C The e.m.f. of a source is the energy transferred when unit charge is driven round a complete circuit.

D The e.m.f. of a source is the energy transferred when unit charge is driven through a resistor.

3 Calculate the charge that passes through a lamp when there is a current of 150 mA for 40 minutes. **[3]**

4 A generator produces a current of 40 A. Calculate how long will it take for a total of 2000 C to flow through the output. **[2]**

5 In a lightning strike there is an average current of 30 kA, which lasts for 2000 μs. Calculate the charge that is transferred in this process. **[3]**

6 **a** A lamp of resistance 15 Ω is connected to a battery of e.m.f. 4.5 V. Calculate the current in the lamp. **[2]**

b Calculate the resistance of the filament of an electric heater that takes a current of 6.5 A when it is connected across a mains supply of 230 V. **[2]**

c Calculate the voltage that is required to drive a current of 2.4 A through a wire of resistance 3.5 Ω. **[2]**

[Total: 6]

7 A battery of e.m.f. 6 V produces a steady current of 2.4 A for 10 minutes. Calculate:

a the charge that it supplied **[2]**

b the energy that it transferred. **[2]**

[Total: 4]

8 Calculate the energy gained by an electron when it is accelerated through a potential difference of 50 kV. (Charge on the electron = -1.6×10^{-19} C.) **[2]**

9 A woman has available 1 A, 3 A, 5 A, 10 A and 13 A fuses. Explain which fuse she should use for a 120 V, 450 W hairdryer. **[3]**

10 This diagram shows the electrolysis of copper chloride.

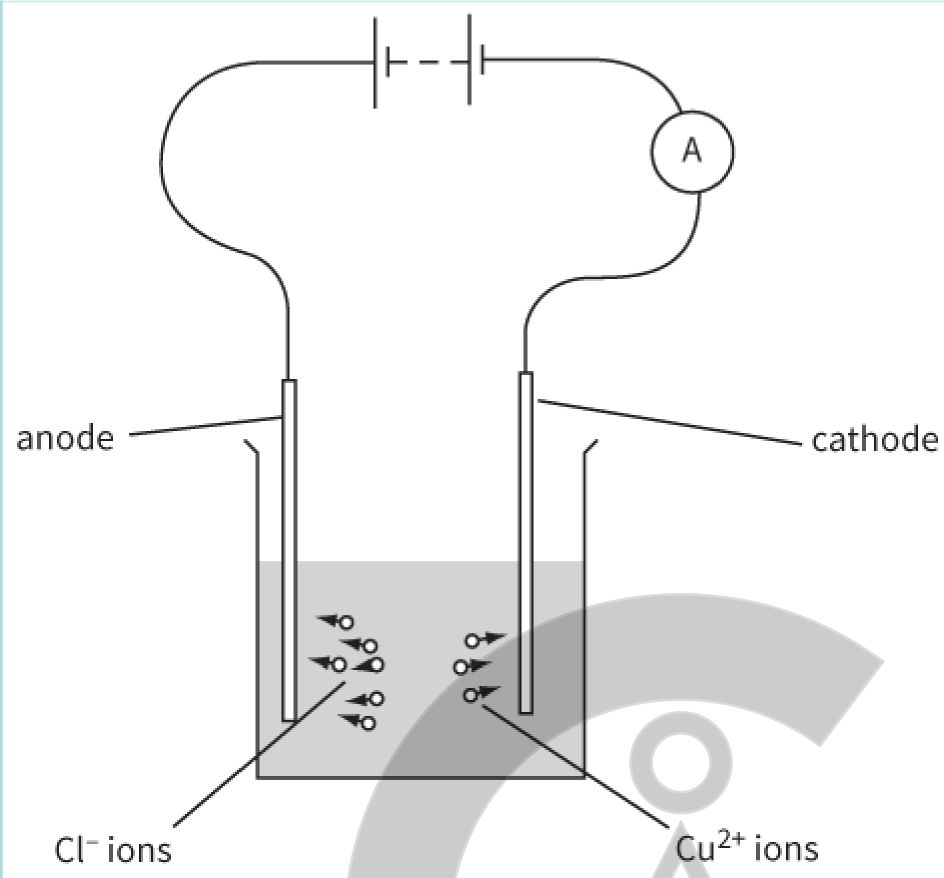


Figure 8.16

a **i** On a copy of the diagram, mark the direction of the conventional current in the electrolyte. Label it conventional current. **[1]**

ii Mark the direction of the electron flow in the connecting wires. Label this electron flow. **[1]**

b In a time period of 8 minutes, 3.6×10^{16} chloride (Cl^-) ions are neutralised and liberated at the anode and 1.8×10^{16} copper (Cu^{2+}) ions are neutralised and deposited on the cathode.

i Calculate the total charge passing through the electrolyte in this time. **[2]**

ii Calculate the current in the circuit. **[2]**

[Total: 6]

11 This diagram shows an electron tube. Electrons moving from the cathode to the anode constitute a current. The current in the ammeter is 4.5 mA.

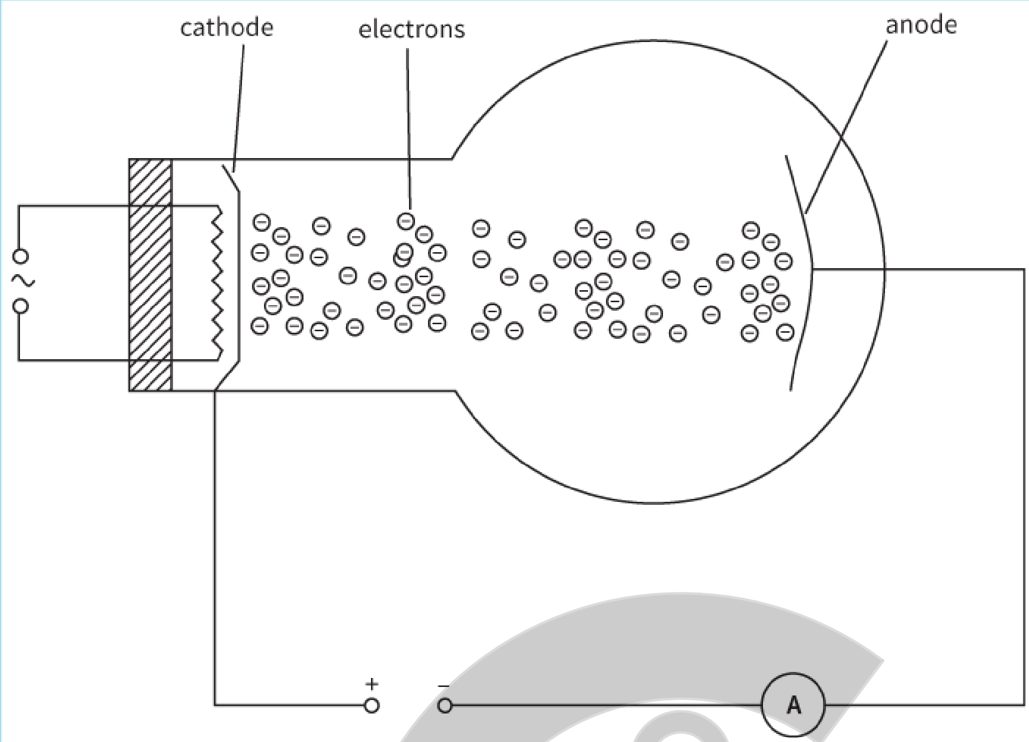


Figure 8.17

a Calculate the charge passing through the ammeter in 3 minutes. **[3]**

b Calculate the number of electrons that hit the anode in 3 minutes. **[3]**

c The potential difference between the cathode and the anode is 75 V. Calculate the energy gained by an electron as it travels from the cathode to the anode. **[2]**

[Total: 8]

12 A length of copper track on a printed circuit board has a cross-sectional area of 5.0×10^{-8} m^2. The current in the track is 3.5 mA. You are provided with some useful information about copper:

1 m^3 of copper has a mass of 8.9×10^3 kg

54 kg of copper contains 6.0×10^{26} atoms

In copper, there is roughly one electron liberated from each copper atom.

a Show that the electron number density n for copper is about 10^{29} m^{-3}. **[2]**

b Calculate the mean drift velocity of the electrons. **[3]**

[Total: 5]

13 a Explain the difference between **potential difference** and **e.m.f.** **[2]**

b A battery has negligible internal resistance, an e.m.f. of 12.0 V and a capacity of 100 A h (ampere-hours). Calculate:

i the total charge that it can supply **[2]**

ii the total energy that it can transfer. **[2]**

c The battery is connected to a 27 W lamp. Calculate the resistance of the lamp. **[2]**

[Total: 8]

14 Some electricity-generating companies use a unit called the kilowatt-hour (kW h) to calculate energy bills. 1 kWh is the energy a kilowatt appliance transfers in 1 hour.

a Show that 1 kWh is equal to 3.6 MJ. **[2]**

b An electric shower heater is rated at 230 V, 9.5 kW.

i Calculate the current it will take from the mains supply. **[2]**

ii Suggest why the shower requires a separate circuit from other appliances. **[1]**

iii Suggest a suitable current rating for the fuse in this circuit. **[1]**

c Calculate the energy transferred when a boy uses the shower for 5 minutes. **[2]**

[Total: 8]

15 A student is measuring the resistance per unit length of a resistance wire. He takes the following measurements.

Quantity	**Value**	**Uncertainty**
length of wire	80 mm	±2%
current in the wire	2.4 A	±0.1 A
potential difference across the wire	8.9 V	±5%

a Calculate the percentage uncertainty in the measurement of the current. **[1]**

b Calculate the value of the resistance per unit length of the wire. **[1]**

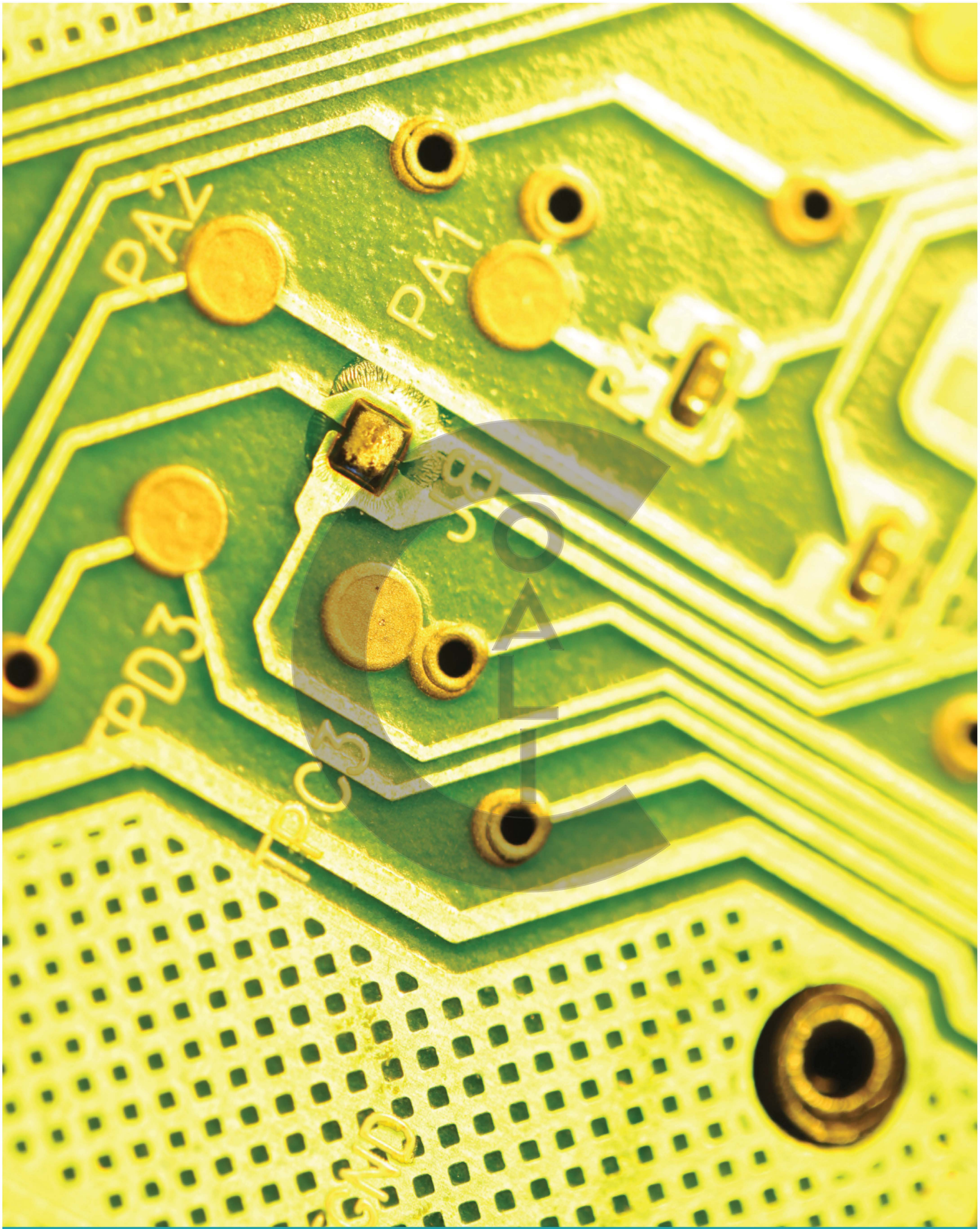
c Calculate the absolute uncertainty of the resistance per unit length of the wire. **[2]**

[Total: 4]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand of the nature of electric current	8.2			
understand the term charge and recognise its unit, the coulomb	8.2			
understand that charge is quantised	8.2			
solve problems using the equation $\Delta Q = I\Delta t$	8.2			
solve problems using the formula $I = nAve$	8.3			
solve problems involving the mean drift velocity of charge carriers	8.3			
understand the terms potential difference, e.m.f. and the volt	8.4			
use energy considerations to distinguish between p.d. and e.m.f.	8.4			
define resistance and recognise its unit, the ohm	8.5			
solve problems using the formula $V = IR$	8.5			
solve problems concerning energy and power in electric circuits.	8.6			


PA2
PA1
PD3
C3
GND

〉Chapter 9

Kirchhoff's laws

LEARNING INTENTIONS

In this chapter you will learn how to:

- recall and apply Kirchhoff's laws
- use Kirchhoff's laws to derive the formulae for the combined resistance of two or more resistors in series and in parallel
- recognise that ammeters are connected in series within a circuit and therefore should have low resistance
- recognise that voltmeters are connected in parallel across a component, or components, and therefore should have high resistance.

BEFORE YOU START

- Write down the name(s) of the meters you use to measure current in a component and potential difference across it.
- Draw a circuit diagram showing a circuit in which a battery is used to drive a current through a variable resistor in series with a lamp. Show on your circuit how you would connect the meters named in your list.
- Try to draw a circuit diagram to measure the potential difference of a component and the current in it. Swap with a classmate to check.

CIRCUIT DESIGN

Over the years, electrical circuits have become increasingly complex, with more and more components combining to achieve very precise results (Figure 9.1). Such circuits typically include power supplies, sensing devices, potential dividers and output devices. At one time, circuit designers would start with a simple circuit and gradually modify it until the desired result was achieved. This is impossible today when circuits include many hundreds or thousands of components.

Instead, electronics engineers (Figure 9.2) rely on computer-based design software that can work out the effect of any combination of components. This is only possible because computers can be programmed with the equations that describe how current and voltage behave in a circuit. These equations, which include Ohm's law and Kirchhoff's two laws, were established in the 18th century, but they have come into their own in the 21st century through their use in computer-aided design (CAD) systems.

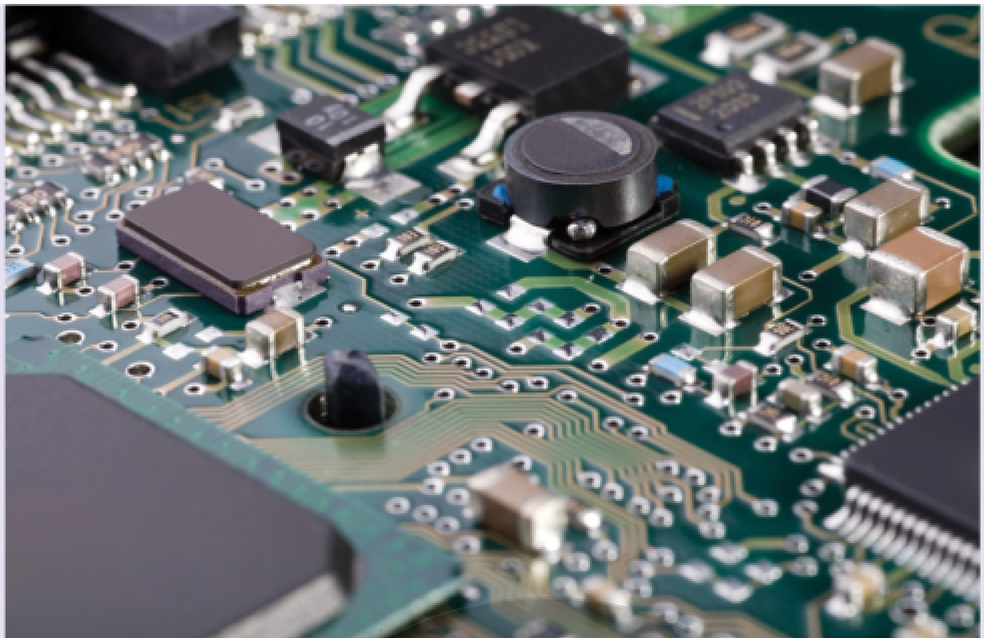


Figure 9.1: A complex electronic circuit – this is the circuit board that controls a computer's hard drive.

Think about other areas of industry. How have computers changed those industrial practices in the last 30 years?

Figure 9.2: A computer engineer uses a computer-aided design (CAD) software tool to design a circuit that will form part of a microprocessor, the device at the heart of every computer.

COALI

9.1 Kirchhoff's first law

You will have learnt that current may divide up where a circuit splits into two separate branches. For example, a current of 5.0 A may split at a junction or a point in a circuit into two separate currents of 2.0 A and 3.0 A. The total amount of current remains the same after it splits. We would not expect some of the current to disappear, or extra current to appear from nowhere. This is the basis of **Kirchhoff's first law**, which states that the sum of the currents entering any point in a circuit is equal to the sum of the currents leaving that same point.

This is illustrated in Figure 9.3. In the first part, the current into point P must equal the current out, so:

$$I_1 = I_2$$

In the second part of the figure, we have one current coming into point Q, and two currents leaving. The current divides at Q. Kirchhoff's first law gives:

$$I_1 = I_2 + I_3$$

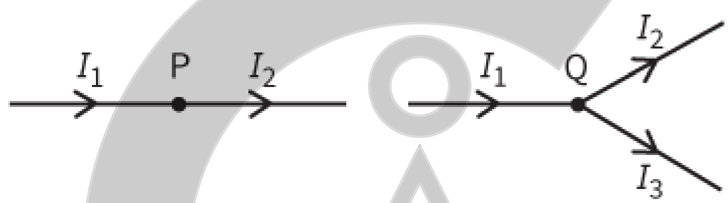


Figure 9.3: Kirchhoff's first law: current is conserved because charge is conserved.

Kirchhoff's first law is an expression of the **conservation of charge**. The idea is that the total amount of charge entering a point must exit the point. To put it another way, if a billion electrons enter a point in a circuit in a time interval of 1.0 s, then one billion electrons must exit this point in 1.0 s. The law can be tested by connecting ammeters at different points in a circuit where the current divides. You should recall that an ammeter must be connected in series so the current to be measured passes through it.

Questions

1 Use Kirchhoff's first law to deduce the value of the current I in Figure 9.4.

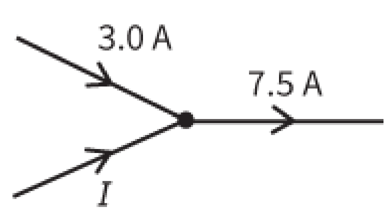


Figure 9.4: For Question 1.

2 In Figure 9.5, calculate the current in the wire X. State the direction of this current (towards P or away from P).

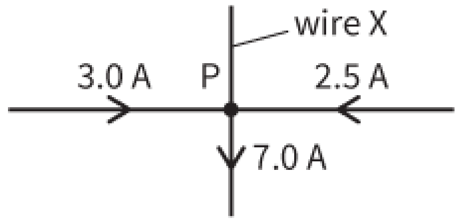


Figure 9.5: For Question 2.

Formal statement of Kirchhoff's first law

We can write Kirchhoff's first law as an equation:

$$\Sigma I_{in} = \Sigma I_{out}$$

Here, the symbol Σ (Greek letter sigma) means 'the sum of all', so ΣI_{in} means 'the sum of all currents entering into a point' and ΣI_{out} means 'the sum of all currents leaving that point'. This is the sort of equation that a computer program can use to predict the behaviour of a complex circuit.

KEY EQUATIONS

Kirchhoff's first law:

$$\Sigma I_{in} = \Sigma I_{out}$$

Questions

3 Calculate ΣI_{in} and ΣI_{out} in Figure 9.6. Is Kirchhoff's first law satisfied?

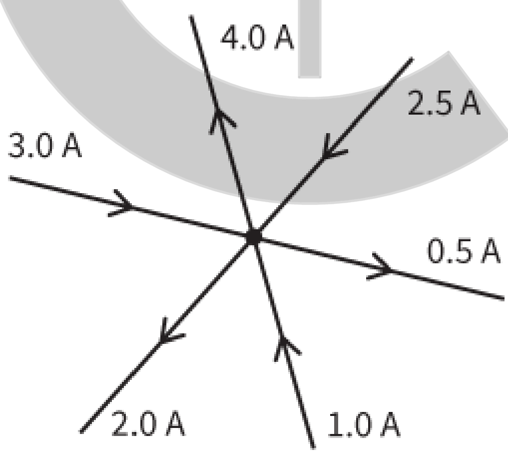


Figure 9.6: For Question 3.

4 Use Kirchhoff's first law to deduce the value and direction of the current I in Figure 9.7.

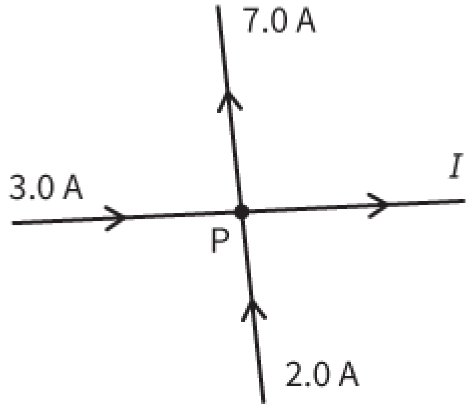


Figure 9.7: For Question 4.

9.2 Kirchhoff's second law

This law deals with e.m.f.s and voltages in a circuit. We will start by considering a simple circuit that contains a cell and two resistors of resistances R_1 and R_2 (Figure 9.8). Since this is a simple series circuit, the current I must be the same all the way around, and we need not concern ourselves further with Kirchhoff's first law. For this circuit, we can write the following equation:

$$E = IR_1 + IR_2$$

e.m.f. of battery = sum of p.d.s across the resistors

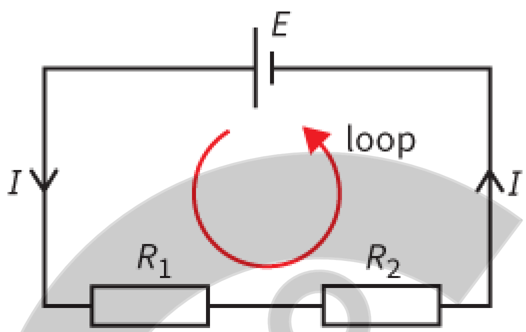


Figure 9.8: A simple series circuit.

You should not find these equations surprising. However, you may not realise that they are a consequence of applying **Kirchhoff's second law** to the circuit. This law states that the sum of the e.m.f.s around any loop in a circuit is equal to the sum of the p.d.s around the loop.

WORKED EXAMPLE

1 Use Kirchhoff's laws to find the current in the circuit in Figure 9.9.

This is a series circuit so the current is the same all the way round the circuit.

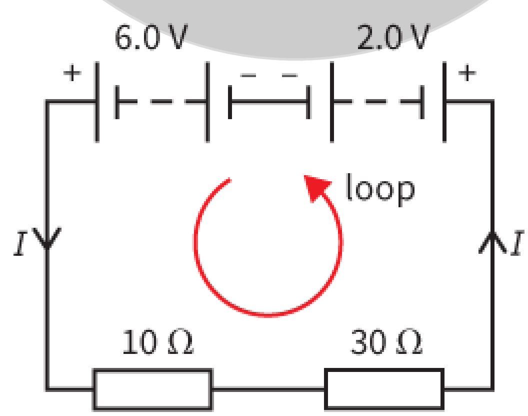


Figure 9.9: A circuit with two opposing batteries.

Step 1 We calculate the sum of the e.m.f.s:

sum of e.m.f.s = 6.0 V − 2.0 V = 4.0 V

The batteries are connected in opposite directions so we must consider one of the e.m.f.s as negative.

Step 2 We calculate the sum of the p.d.s.

sum of p.d.s = $(I \times 10) + (I \times 30) = 40\,I$

Step 3 We equate these:

$4.0 = 40\,I$

and so $I = 0.1$ A

No doubt, you could have solved this problem without formally applying Kirchhoff's second law, but you will find that in more complex problems the use of these laws will help you to avoid errors.

You will see later that Kirchhoff's second law is an expression of the conservation of energy. We shall look at another example of how this law can be applied, and then look at how it can be applied in general.

Question

5 Use Kirchhoff's second law to deduce the p.d. across the resistor of resistance *R* in the circuit shown in Figure 9.10, and hence find the value of *R*. (Assume the battery of e.m.f. 10 V has negligible internal resistance.)

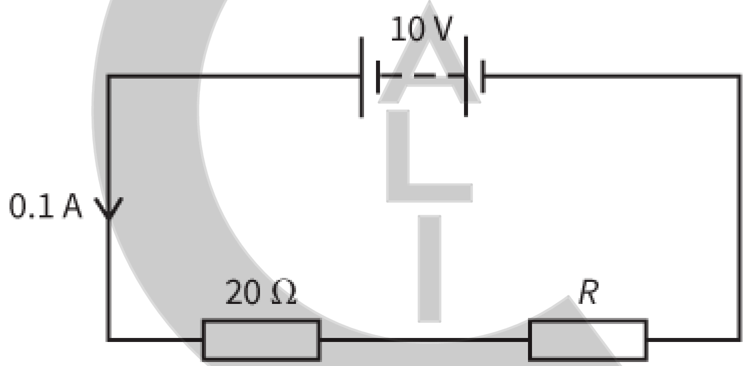


Figure 9.10: Circuit for Question 5.

An equation for Kirchhoff's second law

In a similar manner to the formal statement of the first law, the second law can be written as an equation:

$$\Sigma E = \Sigma V$$

where ΣE is the sum of the e.m.f.s and ΣV is the sum of the potential differences.

KEY EQUATION

Kirchhoff's second law:

$$\Sigma E = \Sigma V$$

COALI

9.3 Applying Kirchhoff's laws

Figure 9.11 shows a more complex circuit, with more than one 'loop'. Again, there are two batteries and two resistors. The problem is to find the current in each resistor. There are several steps in this; Worked example 2 shows how such a problem is solved.

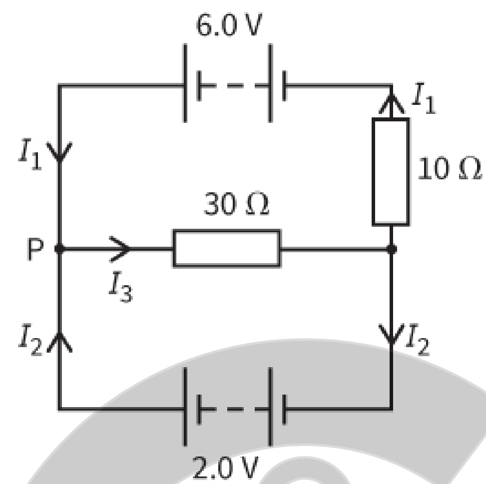


Figure 9.11: Kirchhoff's laws are needed to determine the currents in this circuit.

Signs and directions

Caution is necessary when applying Kirchhoff's second law. You need to take account of the ways in which the sources of e.m.f. are connected and the directions of the currents. Figure 9.12 shows one loop from a larger complicated circuit to illustrate this point. Only the components and currents in this particular loop are shown.

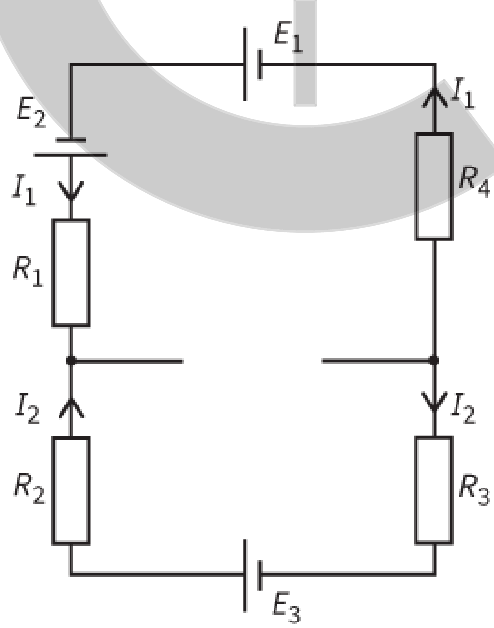


Figure 9.12: A loop extracted from a complicated circuit.

e.m.f.s

Starting with the cell of e.m.f. E_1 and working **anticlockwise** around the loop (because E_1 is 'pushing current' anticlockwise):

$$\text{sum of e.m.f.s} = E_1 + E_2 - E_3$$

Note that E_3 is opposing the other two e.m.f.s.

p.d.s

Starting from the same point, and working **anticlockwise** again:

$$\text{sum of p.d.s} = I_1R_1 - I_2R_2 - I_2R_3 + I_1R_4$$

Note that the direction of current I_2 is clockwise, so the p.d.s that involve I_2 are negative.

WORKED EXAMPLE

2 Calculate the current in each of the resistors in the circuit shown in Figure 9.11.

Step 1 Mark the currents. The diagram shows I_1, I_2 and I_3.

Hint: *It does not matter if we mark these in the wrong directions, as they will simply appear as negative quantities in the solutions.*

Step 2 Apply Kirchhoff's first law. At point P, this gives:

$I_1 + I_2 = I_3$ (1)

Step 3 Choose a loop and apply Kirchhoff's second law. Around the upper loop, this gives:

$6.0 = (I_3 \times 30) + (I_1 \times 10)$ (2)

Step 4 Repeat step 3 around other loops until there are the same number of equations as unknown currents. Around the lower loop, this gives:

$2.0 = I_3 \times 30$ (3)

We now have three equations with three unknowns (the three currents).

Step 5 Solve these equations as simultaneous equations. In this case, the situation has been chosen to give simple solutions. Equation 3 gives $I_3 = 0.067$ A, and substituting this value in Equation 2 gives $I_1 = 0.400$ A. We can now find I_2 by substituting in equation 1:

$I_2 = I_3 - I_1 = 0.067 - 0.400 = -0.333 \text{ A} \approx -0.33 \text{ A}$

Thus I_2 is negative–it is in the opposite direction to the arrow shown in Figure 9.11.

Note that there is a third 'loop' in this circuit; we could have applied Kirchhoff's second law to the outermost loop of the circuit. This would give a fourth equation:

$6 - 2 = I_1 \times 10$

However, this is not an independent equation; we could have arrived at it by subtracting equation 3 from equation 2.

Questions

6 You can use Kirchhoff's second law to find the current I in the circuit shown in Figure 9.13. Choosing the best loop can simplify the problem.

a Which loop in the circuit should you choose?

b Calculate the current I.

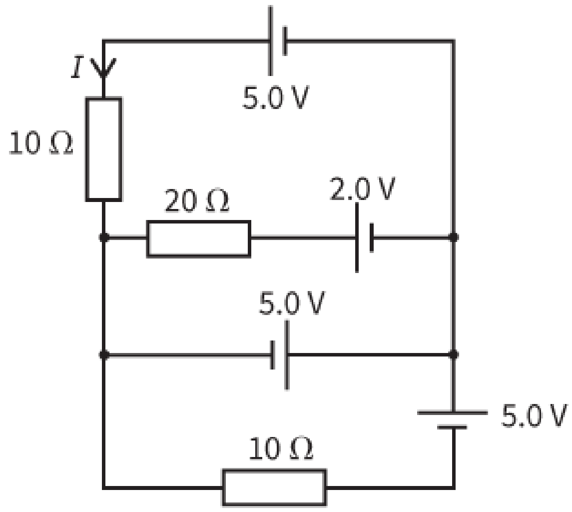


Figure 9.13: Careful choice of a suitable loop can make it easier to solve problems like this. For Question 6.

7 Use Kirchhoff's second law to deduce the resistance *R* of the resistor shown in the circuit loop of Figure 9.14.

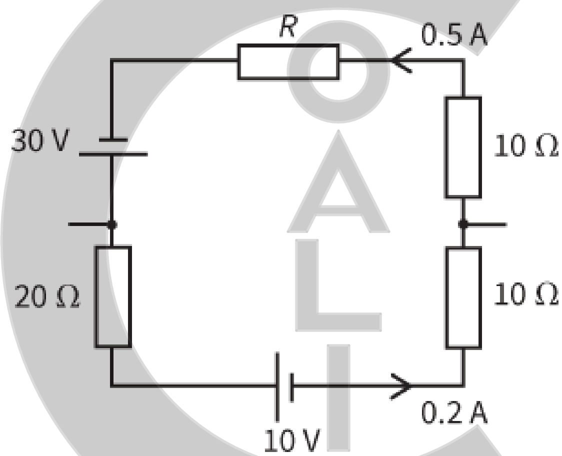


Figure 9.14: For Question 7.

Conservation of energy

Kirchhoff's second law is a consequence of the principle of conservation of energy. If a charge, say 1 C, moves around the circuit, it **gains** energy as it moves through each source of e.m.f. and loses energy as it passes through each p.d. If the charge moves all the way round the circuit so that it ends up where it started, it must have the same energy at the end as at the beginning. (Otherwise we would be able to create energy from nothing simply by moving charges around circuits.)

So:

energy gained passing through sources of e.m.f. = energy lost passing through components with p.d.s

You should recall that an e.m.f. in volts is simply the energy gained per 1 C of charge as it passes through a source. Similarly, a p.d. is the energy lost per 1 C as it passes through a component.

1 volt = 1 joule per coulomb

Hence, we can think of Kirchhoff's second law as:

energy gained per coulomb around loop = energy lost per coulomb around loop

Here is another way to think of the meaning of e.m.f. A 1.5 V cell gives 1.5 J of energy to each coulomb of charge that passes through it. The charge then moves round the circuit, transferring the energy to components in the circuit. The consequence is that, by driving 1 C of charge around the circuit, the cell transfers 1.5 J of energy. Hence, the e.m.f. of a source simply tells us the amount of energy (in joules) transferred by the source in driving unit charge (1 C) around a circuit.

Questions

8 Use the idea of the energy gained and lost by a 1 C charge to explain why two 6 V batteries connected together in series can give an e.m.f. of 12 V or 0 V, but connected in parallel they give an e.m.f. of 6 V.

9 Apply Kirchhoff's laws to the circuit shown in Figure 9.15 to determine the current that will be shown by the ammeters A_1, A_2 and A_3.

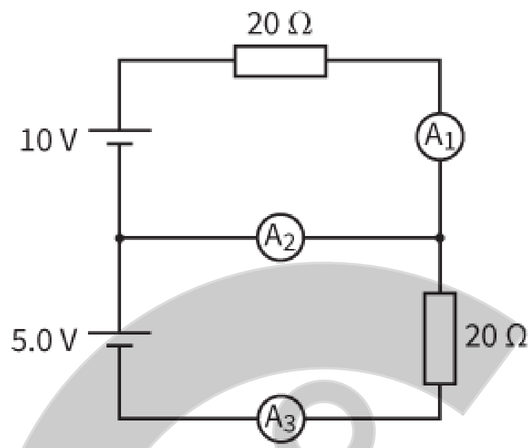


Figure 9.15: Kirchhoff's laws make it possible to deduce the ammeter readings.

9.4 Resistor combinations

You are already familiar with the formulae used to calculate the combined resistance R of two or more resistors connected in series or in parallel. To derive these formulae we have to use Kirchhoff's laws.

Resistors in series

Take two resistors of resistances R_1 and R_2 connected in series (Figure 9.16). According to Kirchhoff's first law, the current in each resistor is the same. The p.d. V across the combination is equal to the sum of the p.d.s across the two resistors:

$$V = V_1 + V_2$$

Since $V = IR$, $V_1 = IR_1$ and $V_2 = IR_2$, we can write:

$$IR = IR_1 + IR_2$$

Cancelling the common factor of current I gives:

$$R = R_1 + R_2$$

KEY EQUATION

Total resistance R of three or more resistors in series = $R_1 + R_2 + R_3 + \ldots$

For three or more resistors, the equation for total resistance R becomes:

$$R = R_1 + R_2 + R_3 + \ldots$$

You must learn how to derive this equation using Kirchhoff's laws.

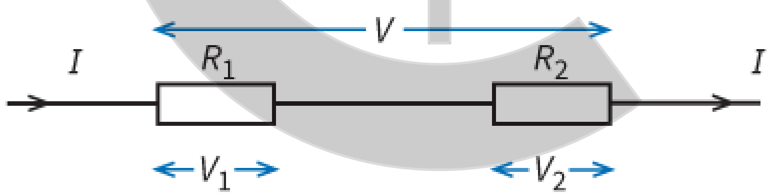


Figure 9.16: Resistors in series.

Questions

10 Calculate the combined resistance of two 5 Ω resistors and a 10 Ω resistor connected in series.

11 The cell shown in Figure 9.17 provides an e.m.f. of 2.0 V. The p.d. across one lamp is 1.2 V. Determine the p.d. across the other lamp.

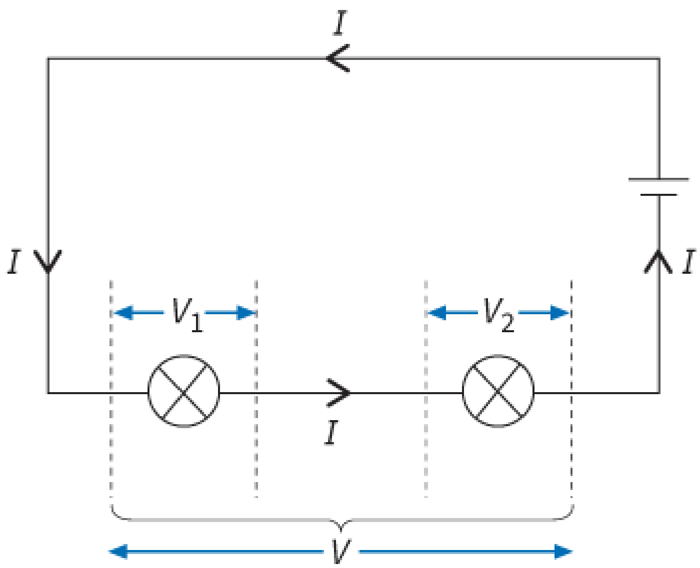


Figure 9.17: A series circuit for Question 11.

12 You have five 1.5 V cells. How would you connect all five of them to give an e.m.f. of:

a 7.5 V

b 1.5 V

c 4.5 V?

Resistors in parallel

For two resistors of resistances R_1 and R_2 connected in parallel (Figure 9.18), we have a situation where the current divides between them. Hence, using Kirchhoff's first law, we can write:

$$I = I_1 + I_2$$

If we apply Kirchhoff's second law to the loop that contains the two resistors, we have:

$$I_1R_1 - I_2R_2 = 0\text{ V}$$

(because there is no source of e.m.f. in the loop).

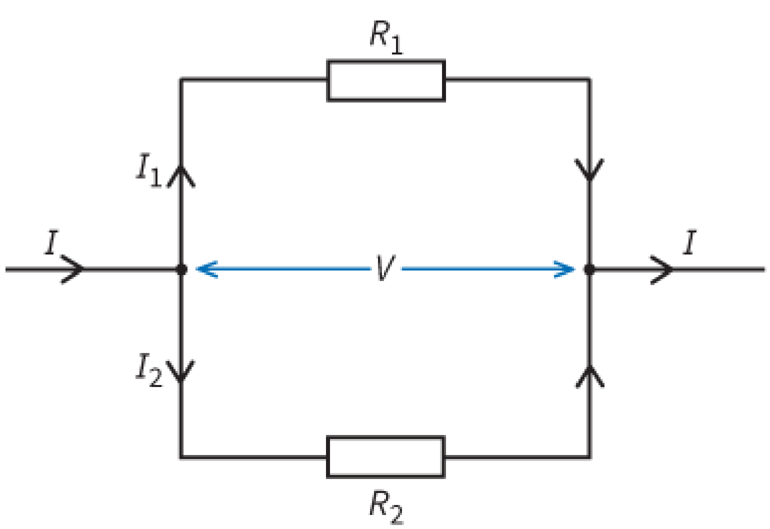


Figure 9.18: Resistors connected in parallel.

This equation states that the two resistors have the same p.d. V across them. Hence we can write:

$$I = \frac{V}{R}$$
$$I_1 = \frac{V}{R_1}$$
$$I_2 = \frac{V}{R_2}$$

Substituting in $I = I_1 + I_2$ and cancelling the common factor V gives:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

For three or more resistors, the equation for total resistance R becomes:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$$

KEY EQUATION

Total resistance R of three or more resistors in parallel is given by the equation

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$$

You must learn how to derive this equation using Kirchhoff's laws.

To summarise, when components are connected in parallel:

- all have the same p.d. across their ends
- the current is shared between them
- we use the reciprocal formula to calculate their combined resistance.

WORKED EXAMPLE

3 Two 10 Ω resistors are connected in parallel. Calculate the total resistance.

Step 1 We have $R_1 = R_2 = 10\ \Omega$, so:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$
$$\frac{1}{R} = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \frac{1}{5}$$

Step 2 Inverting both sides of the equation gives:

$R = 5\ \Omega$

Hint: *Take care not to forget this step! Nor should you write* $\frac{1}{R} = \frac{1}{5} = 5\ \Omega$, *as then you are saying* $\frac{1}{5} = 5$.

You can also determine the resistance as follows:

$$R = (R_1^{-1} + R_2^{-1})^{-1}$$
$$= (10^{-1} + 10^{-1})^{-1} = 5\Omega$$

Questions

13 Calculate the total resistance of four 10 Ω resistors connected in parallel.

14 Calculate the resistances of the following combinations:

- **a** 100 Ω and 200 Ω in series
- **b** 100 Ω and 200 Ω in parallel
- **c** 100 Ω and 200 Ω in series and this in parallel with 200 Ω.

15 Calculate the current drawn from a 12 V battery of negligible internal resistance connected to the ends of the following:

- **a** 500 Ω resistor
- **b** 500 Ω and 1000 Ω resistors in series
- **c** 500 Ω and 1000 Ω resistors in parallel.

16 You are given one 200 Ω resistor and two 100 Ω resistors. What total resistances can you obtain by connecting some, none, or all of these resistors in various combinations?

Solving problems with parallel circuits

Here are some useful ideas that may help when you are solving problems with parallel circuits (or checking your answers to see whether they seem reasonable).

- When two or more resistors are connected in parallel, their combined resistance is smaller than any of their individual resistances. For example, three resistors of 2 Ω, 3 Ω and 6 Ω connected together in parallel have a combined resistance of 1 Ω. This is less than the smallest of the individual resistances. This comes about because, by connecting the resistors in parallel, you are providing extra pathways for the current. Since the combined resistance is lower than the individual resistances, it follows that connecting two or more resistors in parallel will increase the current drawn from a supply. Figure 9.19 shows a hazard that can arise when electrical appliances are connected in parallel.
- When components are connected in parallel, they all have the same p.d. across them. This means that you can often ignore parts of the circuit that are not relevant to your calculation.
- Similarly, for resistors in parallel, you may be able to calculate the current in each one individually, then add them up to find the total current. This may be easier than working out their combined resistance using the reciprocal formula. (This is illustrated in Question 19.)

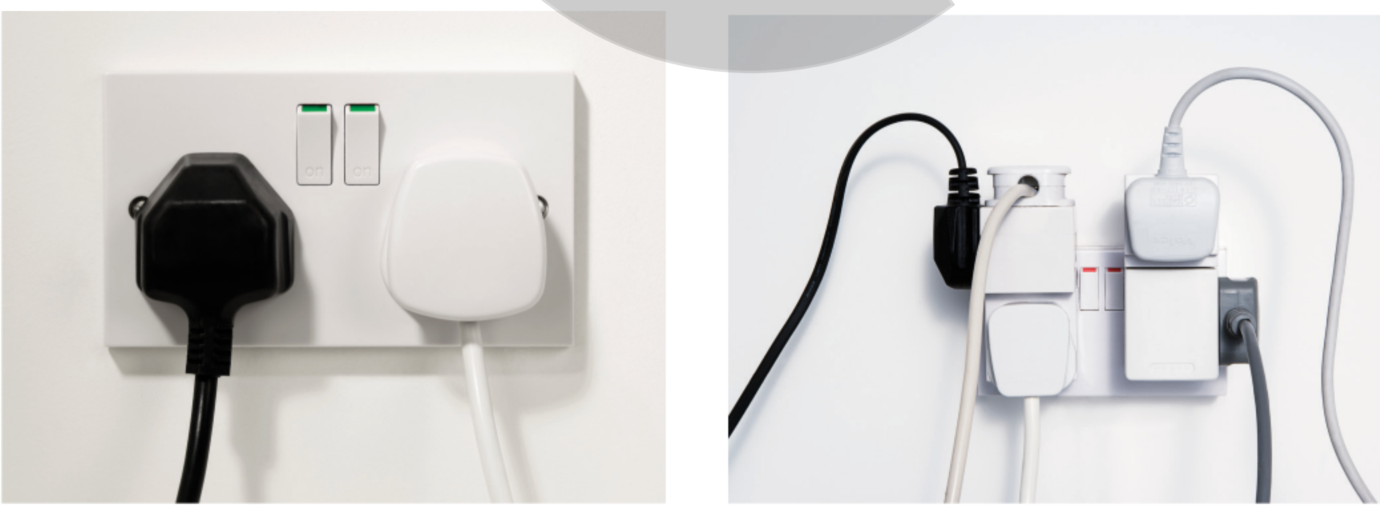


Figure 9.19: a Correct use of an electrical socket. **b** Here, too many appliances (resistances) are connected in parallel. This reduces the total resistance and increases the current drawn, to the point where it becomes dangerous.

Questions

17 Three resistors of resistances 20 Ω, 30 Ω and 60 Ω are connected together in parallel. Select which of the following gives their combined resistance:

110 Ω, 50 Ω, 20 Ω, 10 Ω

(No need to do the calculation!)

18 In the circuit in Figure 9.20 the battery of e.m.f. 10 V has negligible internal resistance. Calculate the current in the 20 Ω resistor shown in the circuit.

19 Determine the current drawn from the battery in Figure 9.20.

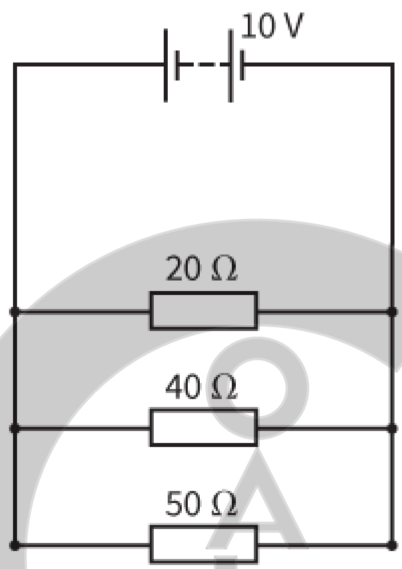


Figure 9.20: Circuit diagram for Questions 18 and 19.

20 What value of resistor must be connected in parallel with a 20 Ω resistor so that their combined resistance is 10 Ω?

21 You are supplied with a number of 100 Ω resistors. Describe how you could combine the minimum number of these to make a 250 Ω resistor.

22 Calculate the current at each point (A–E) in the circuit shown in Figure 9.21.

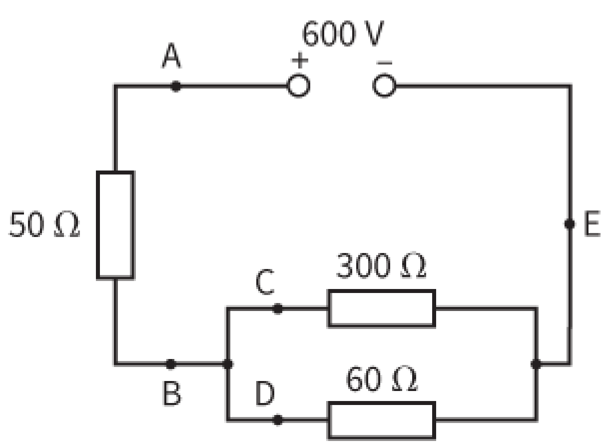


Figure 9.21: For Question 22.

PRACTICAL ACTIVITY 10.1

Ammeters and voltmeters

Ammeters and voltmeters are connected differently in circuits (Figure 9.22). Ammeters are always connected in series, since they measure the current in a circuit. For this reason, an ammeter should have as low a resistance as possible so that as little energy as possible is dissipated in the ammeter itself. Inserting an ammeter with a higher resistance could significantly reduce the current flowing in the circuit. The ideal resistance of an ammeter is zero. Digital ammeters have very low resistances.

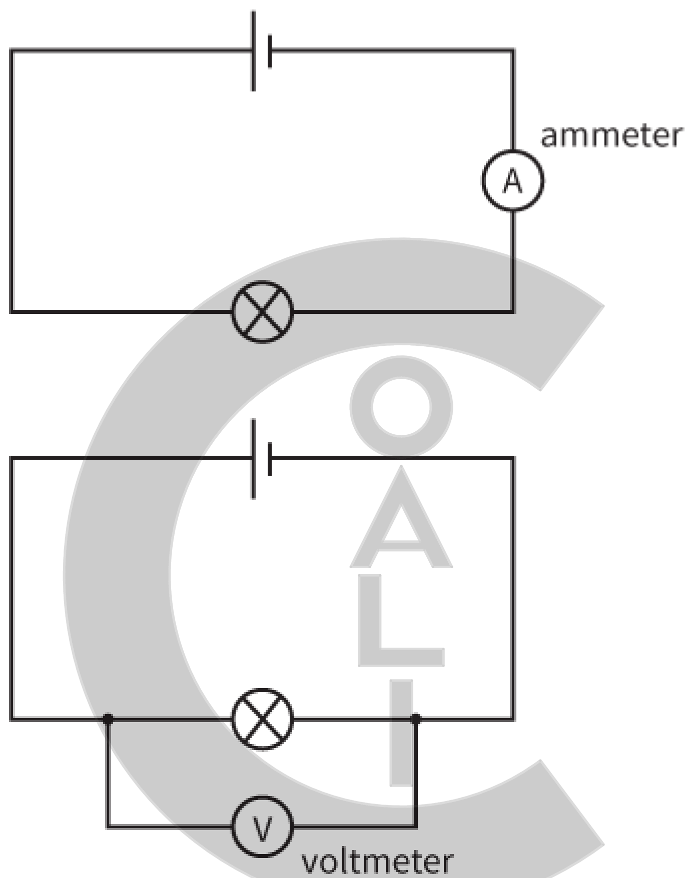


Figure 9.22: How to connect up an ammeter and a voltmeter.

Voltmeters measure the potential difference between two points in the circuit. For this reason, they are connected in parallel (i.e., between the two points), and they should have a very high resistance to take as little current as possible. The ideal resistance of a voltmeter would be infinite. In practice, voltmeters have typical resistance of about 1 MΩ. A voltmeter with a resistance of 10 MΩ measuring a p.d. of 2.5 V will take a current of 2.5×10^{-7} A and dissipate just 0.625 μJ of heat energy from the circuit every second.

Figure 9.23 shows some measuring instruments.

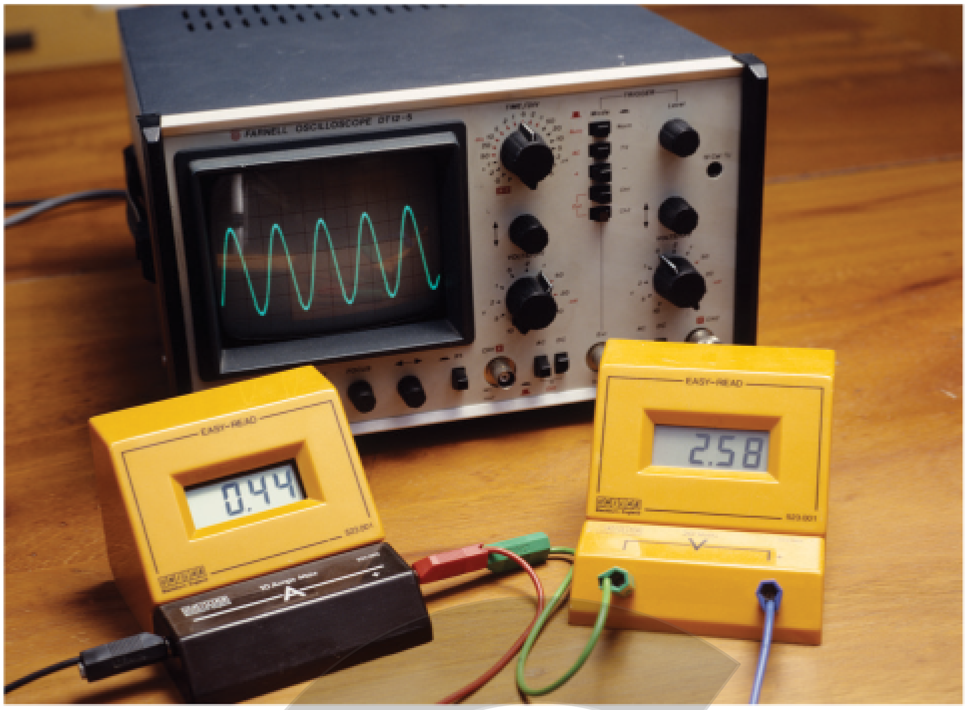


Figure 9.23: Electrical measuring instruments: an ammeter, a voltmeter and an oscilloscope. The oscilloscope can display rapidly changing voltages.

Question

23 **a** A 10 V power supply of negligible internal resistance is connected to a 100 Ω resistor. Calculate the current in the resistor.

b An ammeter is now connected in the circuit, to measure the current. The resistance of the ammeter is 5.0 Ω. Calculate the ammeter reading.

REFLECTION

Kirchhoff's Laws formalise facts that you might already have been familiar with.

Make a list of the main points that these laws have helped clarify in your mind.

Compare your list with two or three other people's lists.

Are they identical?

Thinking back on this chapter, what things might you want more help with?

SUMMARY

Kirchhoff's first law states that the sum of the current currents entering any point in a circuit is equal to the sum of the currents leaving that point.
Kirchhoff's second law states that the sum of the e.m.f.s around any loop in a circuit is equal to the sum of the p.d.s around the loop.
The combined resistance of resistors in series is given by the formula: $$R = R_1 + R_2 + \ldots$$
The combined resistance of resistors in parallel is given by the formula: $$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots$$
Ammeters have a low resistance and are connected in series in a circuit.
Voltmeters have a high resistance and are connected in parallel in a circuit.

EXAM-STYLE QUESTIONS

1 Which row in this table is correct? **[1]**

A	Kirchhoff's first law is an expression of the conservation of charge.	Kirchhoff's second law is an expression of the conservation of charge.
B	Kirchhoff's first law is an expression of the conservation of charge.	Kirchhoff's second law is an expression of the conservation of energy.
C	Kirchhoff's first law is an expression of the conservation of energy.	Kirchhoff's second law is an expression of the conservation of charge.
D	Kirchhoff's first law is an expression of the conservation of energy.	Kirchhoff's second law is an expression of the conservation of energy.

Table 9.1

2 What is the current I_1 in this circuit diagram? **[1]**

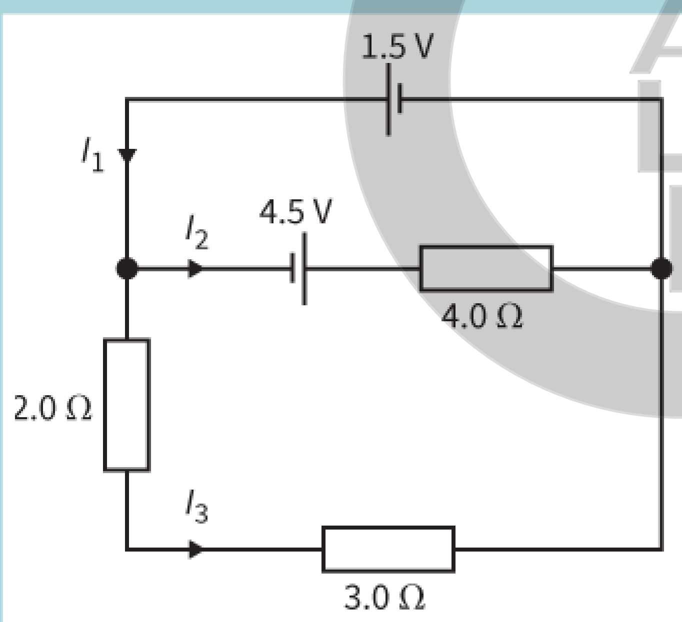


Figure 9.24

A −0.45 A

B +0.45 A

C +1.2 A

D +1.8 A

3 Use Kirchhoff's first law to calculate the unknown currents in these examples.

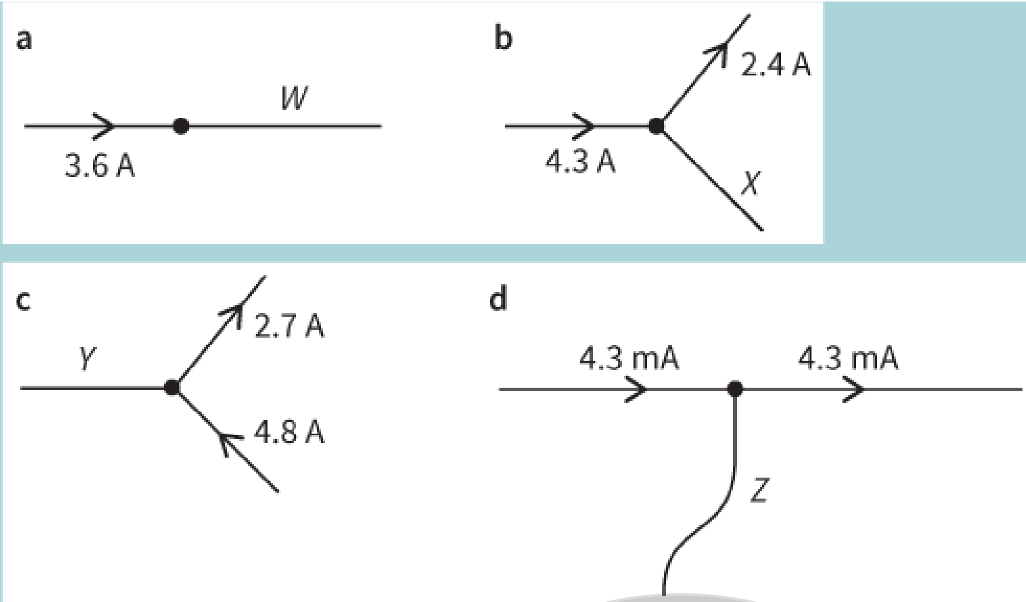


Figure 9.25

For each example, state the direction of the current. **[4]**

4 This diagram shows a part of a circuit.

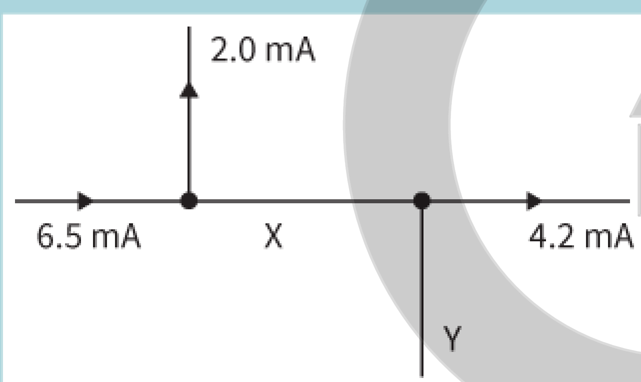


Figure 9.26

Copy the circuit and write in the currents at X and at Y, and show their directions. **[2]**

5 Look at these four circuits.

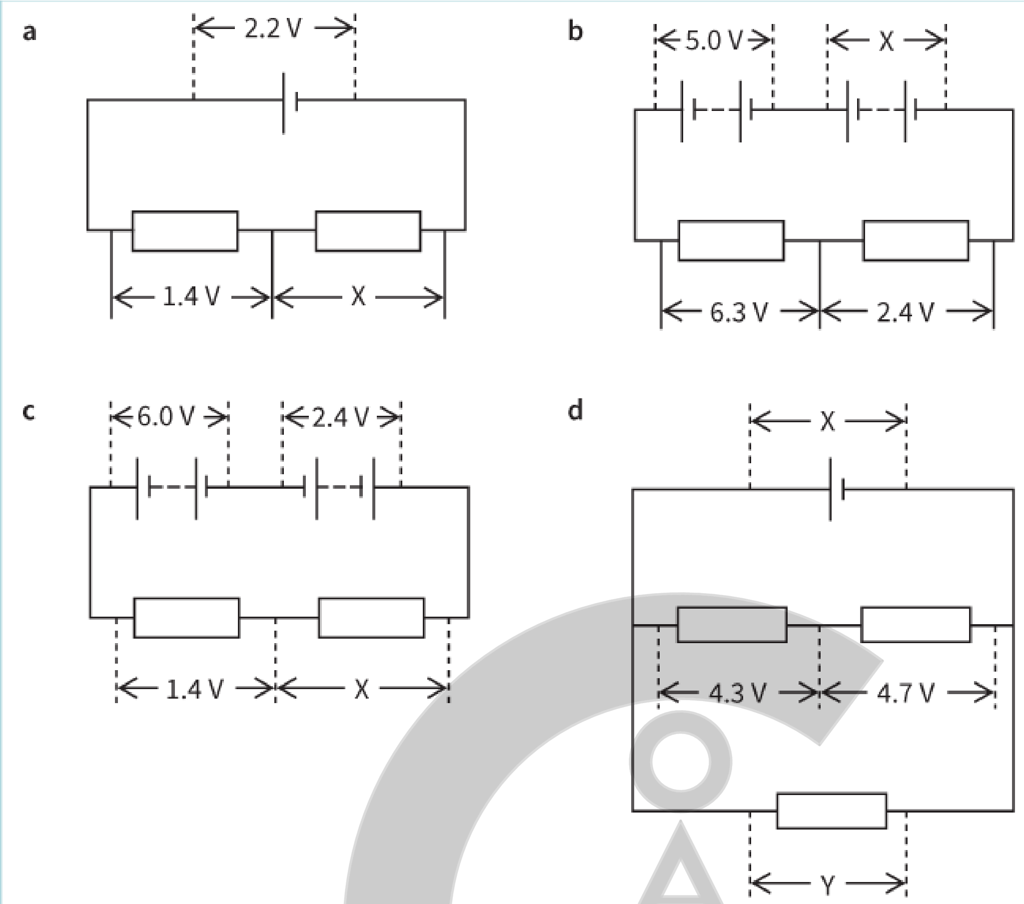


Figure 9.27

Determine the unknown potential difference (or differences) in each case. **[5]**

6 A filament lamp and a 220 Ω resistor are connected in series to a battery of e.m.f. 6.0 V. The battery has negligible internal resistance. A high-resistance voltmeter placed across the resistor measures 1.8 V.

Calculate:

a the current drawn from the battery **[1]**

b the p.d. across the lamp **[1]**

c the total resistance of the circuit **[1]**

d the number of electrons passing through the battery in a time of 1.0 minute. **[4]**

(The elementary charge is 1.6×10^{-19} C.)

[Total: 7]

7 The circuit diagram shows a 12 V power supply connected to some resistors.

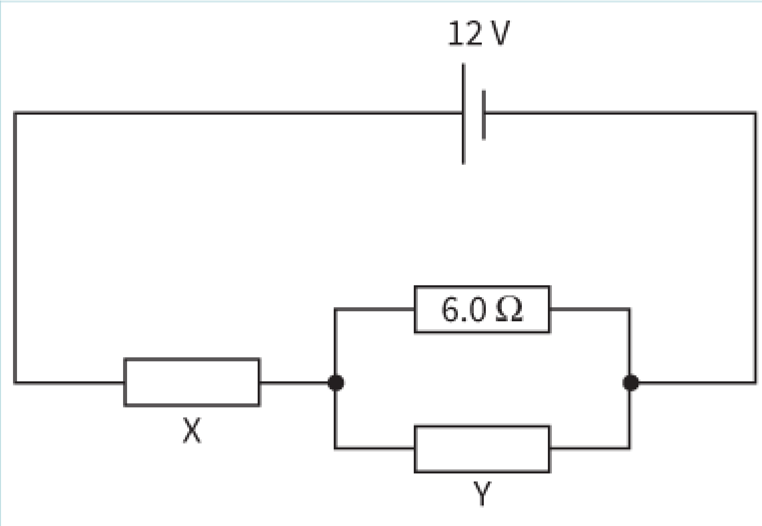


Figure 9.28

The current in the resistor X is 2.0 A and the current in the 6.0 Ω resistor is 0.5 A. Calculate:

a the current in resistor Y **[1]**

b the resistance of resistor Y **[2]**

c the resistance of resistor X. **[2]**

[Total: 5]

8 a Explain the difference between the terms e.m.f. and potential difference. **[2]**

b This circuit contains batteries and resistors. You may assume that the batteries have negligible internal resistance.

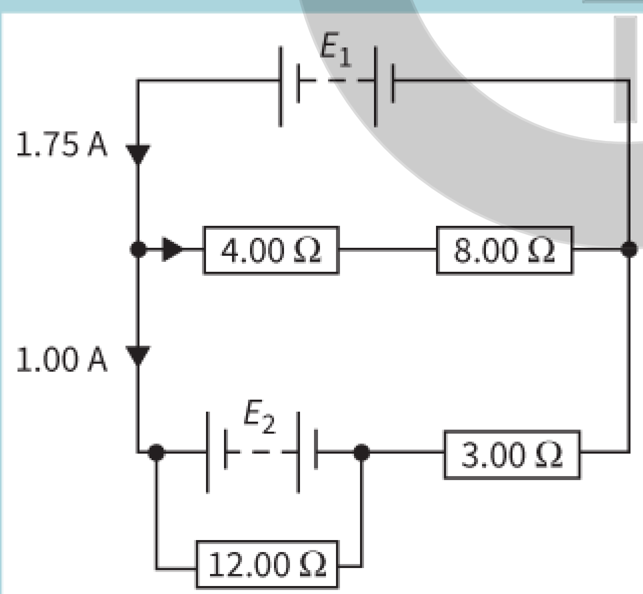


Figure 9.29

i Use Kirchhoff's first law to find the current in the 4.00 Ω and 8.00 Ω resistors. **[1]**

ii Calculate the e.m.f. of E_1. **[2]**

iii Calculate the value of E_2. **[2]**

iv Calculate the current in the 12.00 Ω resistor. **[2]**

[Total: 9]

9 a Explain why an ammeter is designed to have a low resistance. **[1]**

A student builds the circuit, as shown, using a battery of negligible internal resistance. The reading on the voltmeter is 9.0 V.

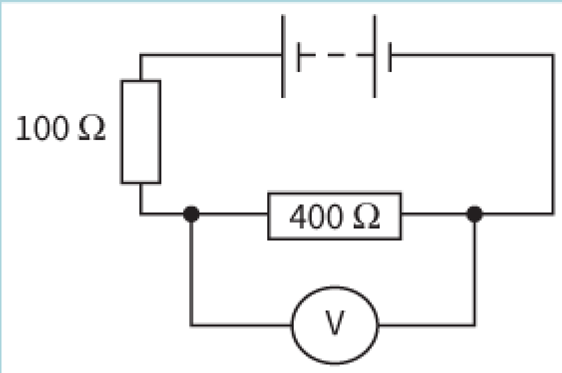


Figure 9.30

b i The voltmeter has a resistance of 1200 Ω. Calculate the e.m.f. of the battery. **[4]**

ii The student now repeats the experiment using a voltmeter of resistance 12 kΩ. Show that the reading on this voltmeter would be 9.5 V. **[3]**

iii Refer to your answers to **i** and **ii** and explain why a voltmeter should have as high a resistance as possible. **[2]**

[Total: 10]

10 a Explain what is meant by the resistance of a resistor. **[1]**

b This diagram shows a network of resistors connected to a cell of e.m.f. 6.0 V.

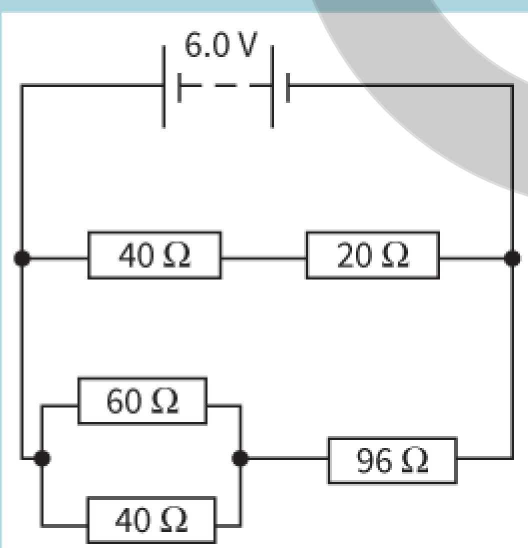


Figure 9.31

Show that the resistance of the network of resistors is 40 Ω. **[3]**

c Calculate the current in the 60 Ω resistor. **[3]**

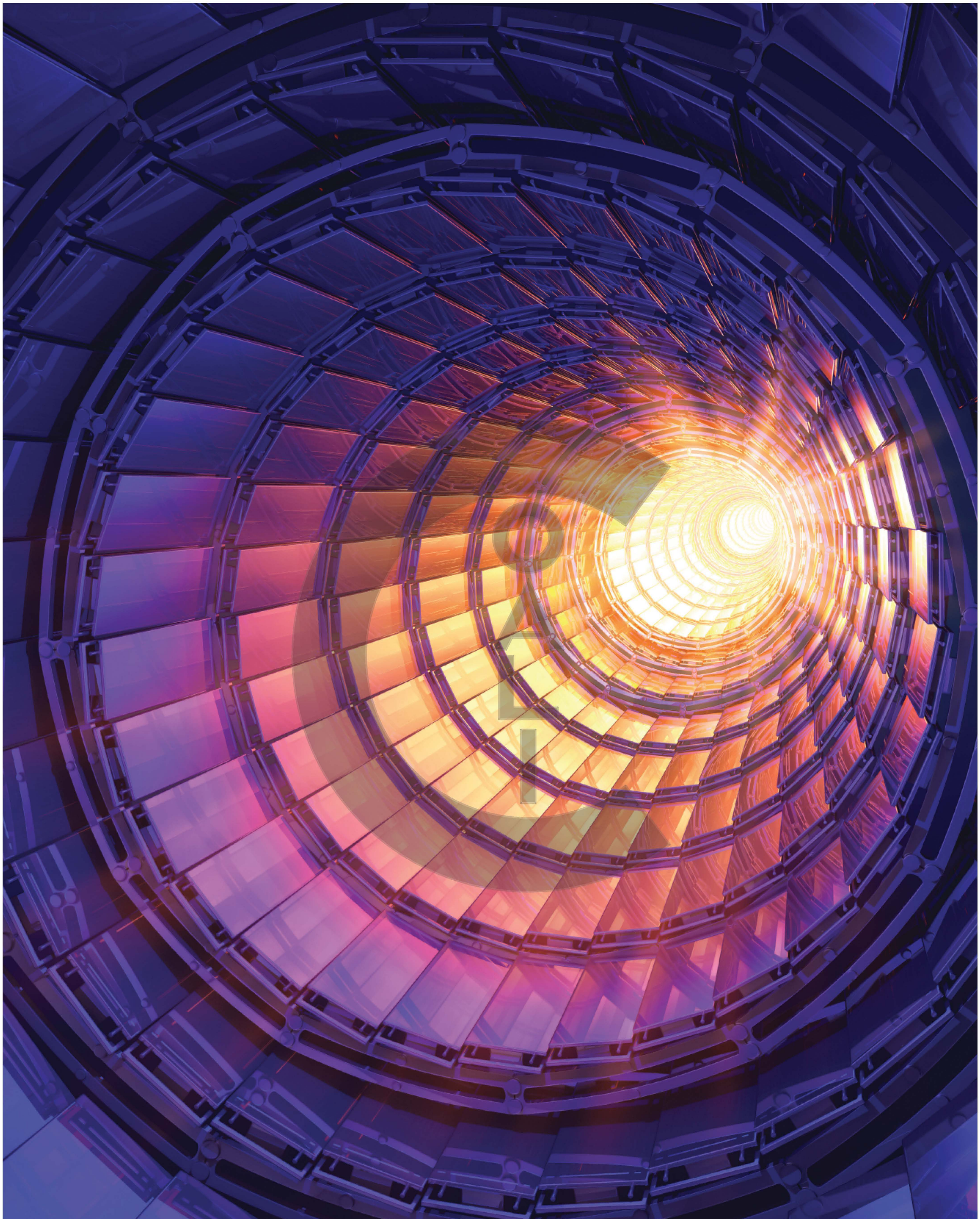
[Total: 7]

COALI

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
state and use Kirchhoff's first law	9.1, 9.3			
state and use Kirchhoff's second law that states that the sum of the e.m.f.s around any loop in a circuit is equal to the sum of the p.d.s around the loop	9.2, 9.3			
calculate the total resistance of two or more resistors in series	9.4			
calculate the resistance of two or more resistors in parallel	9.4			
understand that ammeters have a low resistance and are connected in series in a circuit	9.4			
understand that voltmeters have a high resistance and are connected in parallel in a circuit.	9.4			



〉 Chapter 10

Resistance and resistivity

LEARNING INTENTIONS

In this chapter you will learn how to:

- state Ohm's law
- sketch and explain the *I–V* characteristics for various components
- sketch the temperature characteristic for an NTC thermistor
- solve problems involving the resistivity of a material.

BEFORE YOU START

- Do you understand the terms introduced in Chapters 8 and 9: current, charge, potential difference, e.m.f., resistance and their relationships to one another?
- What are their units?
- Take turns in challenging a partner to define a term or to write down an equation linking different terms. Do not use the textbook or your notes to look up the terms.

SUPERCONDUCTIVITY

As metals are cooled, their resistance decreases. It was discovered as long ago as 1911 that when mercury was cooled using liquid helium to 4.1 K (4.1 degrees above absolute zero), its resistance suddenly fell to zero. This phenomenon was named **superconductivity**. Other metals, such as lead at 7.2 K, also become superconductors.

When charge flows in a superconductor, it can continue in that superconductor without the need for any potential difference and without dissipating any energy. This means that large currents can occur without the unwanted heating effect that would occur in a normal metallic or semiconducting conductor.

Initially, superconductivity was only of scientific interest and had little practical use, as the liquid helium that was required to cool the superconductors is very expensive to produce. In 1986, it was discovered that particular ceramics became superconducting at much higher temperatures – above 77 K, the boiling point of liquid nitrogen. This meant that liquid nitrogen, which is readily available, could be used to cool the superconductors and expensive liquid helium was no longer needed. Consequently, superconductor technology became a feasible proposition.

Uses of superconductors

The JR-Maglev train in Japan's Yamanashi province floats above the track using superconducting magnets (Figure 10.1). This means that not only is the heating effect of the current in the magnet coils reduced to zero – it also means that the friction between the train and the track is eliminated and that the train can reach incredibly high speeds of up to 580 km h^{-1}.

Particle accelerators, such as the Large Hadron Collider (LHC) at the CERN research facility in Switzerland, accelerate beams of charged particles to very high energies by making them orbit around a circular track many times. The particles are kept moving in the circular path by very strong magnetic fields produced by electromagnets whose coils are made from superconductors. Much of our understanding of the fundamental nature of matter is from doing experiments in which beams of these very high speed particles are made to collide with each other.

Figure 10.1: The Japanese JR-Maglev train, capable of speeds approaching 600 km h^{-1}.

Magnetic resonance imaging (MRI) was developed in the 1940s. It is used by doctors to examine internal organs without invasive surgery.

Superconducting magnets can be made much smaller than conventional magnets, and this has enabled the magnetic fields produced to be much more precise, resulting in better imaging.

Imagine you are a scientific consultant for a new science fiction film. You have been instructed to find a use of a superconductor to enable the hero to escape from a villain who is about to destroy the world. What use would you come up with?

10.1 The *I–V* characteristic for a metallic conductor

In Chapter 8, we saw how we could measure the resistance of a resistor using a voltmeter and ammeter. In this topic we are going to investigate the variation of the current – and, therefore, resistance – as the potential difference across a conductor changes.

The potential difference across a metal conductor can be altered using a variable power supply or by placing a variable resistor in series with the conductor. This allows us to measure the current at different potential differences across the conductor. The results of such a series of measurements are shown graphically in Figure 10.2.

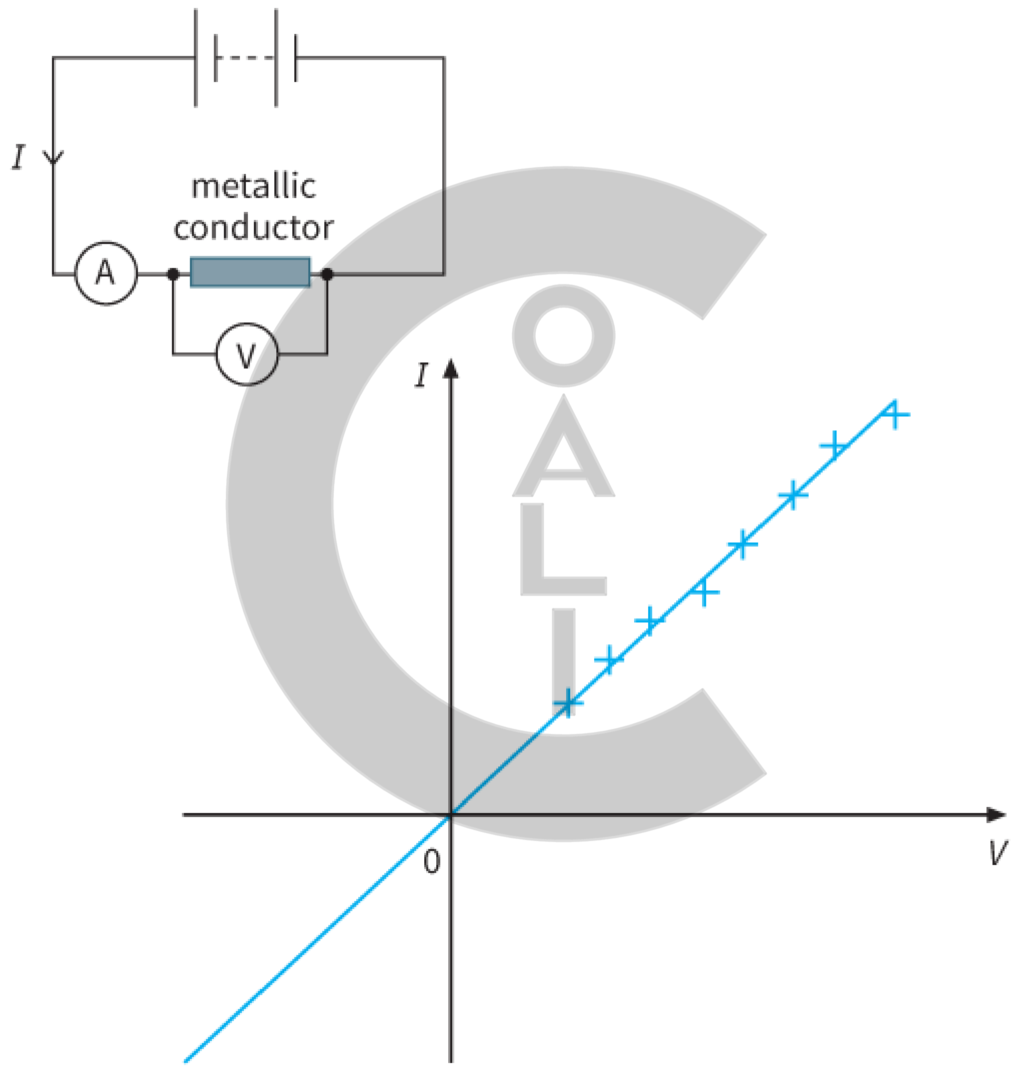


Figure 10.2: To determine the resistance of a component, you need to measure both current and potential difference.

Look at the graph of Figure 10.2. Such a graph is known as an **I–V characteristic**. The points are slightly scattered, but they clearly lie on a straight line. A line of best fit has been drawn. You will see that it passes through the origin of the graph. In other words, the current *I* is directly proportional to the voltage *V*.

The straight-line graph passing through the origin shows that the resistance of the conductor remains constant. If you double the current, the voltage will also double. However, its resistance, which is the ratio of the voltage to the current, remains the same. Instead of using:

$$R = \frac{V}{I}$$

to determine the resistance, for a graph of *I* against *V* that is a straight line passing through the origin, you can also use:

$$\text{resistance} = \frac{1}{\text{gradient of graph}}$$

(This will give a more accurate value for *R* than if you were to take a single experimental data point. Take care! You can only find resistance from the gradient if the *I*–*V* graph is a straight line through the origin.)

By reversing the connections to the resistor, the p.d. across it will be reversed (in other words, it becomes negative). The current will be in the opposite direction – it is also negative. The graph is symmetrical, showing that if a p.d. of, say, 2.0 V produces a current of 0.5 A, then a p.d. of −2.0 V will produce a current of −0.5 A. This is true for most simple metallic conductors but is not true for some electronic components, such as diodes.

You get results similar to those shown in Figure 10.2 for a commercial **resistor**. Resistors have different resistances, so the gradient of the *I*–*V* graph will be different for different resistors.

Question

1 Table 10.1 shows the results of an experiment to measure the resistance of a carbon resistor whose resistance is given by the manufacturer as 47 Ω ± 10%.

a Plot a graph to show the *I*–*V* characteristic of this resistor.

b Do the points appear to fall on a straight line that passes through the origin of the graph?

c Use the graph to determine the resistance of the resistor.

d Does the value of the resistance fall within the range given by the manufacturer?

Potential difference / V	Current / A
2.1	0.040
4.0	0.079
6.3	0.128
7.9	0.192
10.0	0.202
12.1	0.250

Table 10.1: Potential difference V and current I

10.2 Ohm's law

For the metallic conductor whose *I–V* characteristic is shown in Figure 10.2, the current in it is directly proportional to the p.d. across it. This means that its resistance is independent of both the current and the p.d.

This is because the ratio $\frac{V}{I}$ is a constant. Any component that behaves like this is described as an **ohmic** component, and we say that it obeys **Ohm's law**. The statement of Ohm's law is very precise and you must not confuse this with the equation $\frac{V}{I} = R$

A conductor obeys Ohm's law if the current in it is directly proportional to the potential difference across its ends.

Question

2 An electrical component allows a current of 10 mA through it when a voltage of 2.0 V is applied. When the voltage is increased to 8.0 V, the current becomes 60 mA. Does the component obey Ohm's law? Give numerical values for the resistance to justify your answer.

10.3 Resistance and temperature

A conductor that does not obey Ohm's law is described as **non-ohmic**. An example is a filament lamp. Figure 10.3 shows such a lamp; you can clearly see the wire filament glowing as the current passes through it. Figure 10.4 shows the *I*–*V* characteristic for a similar lamp.

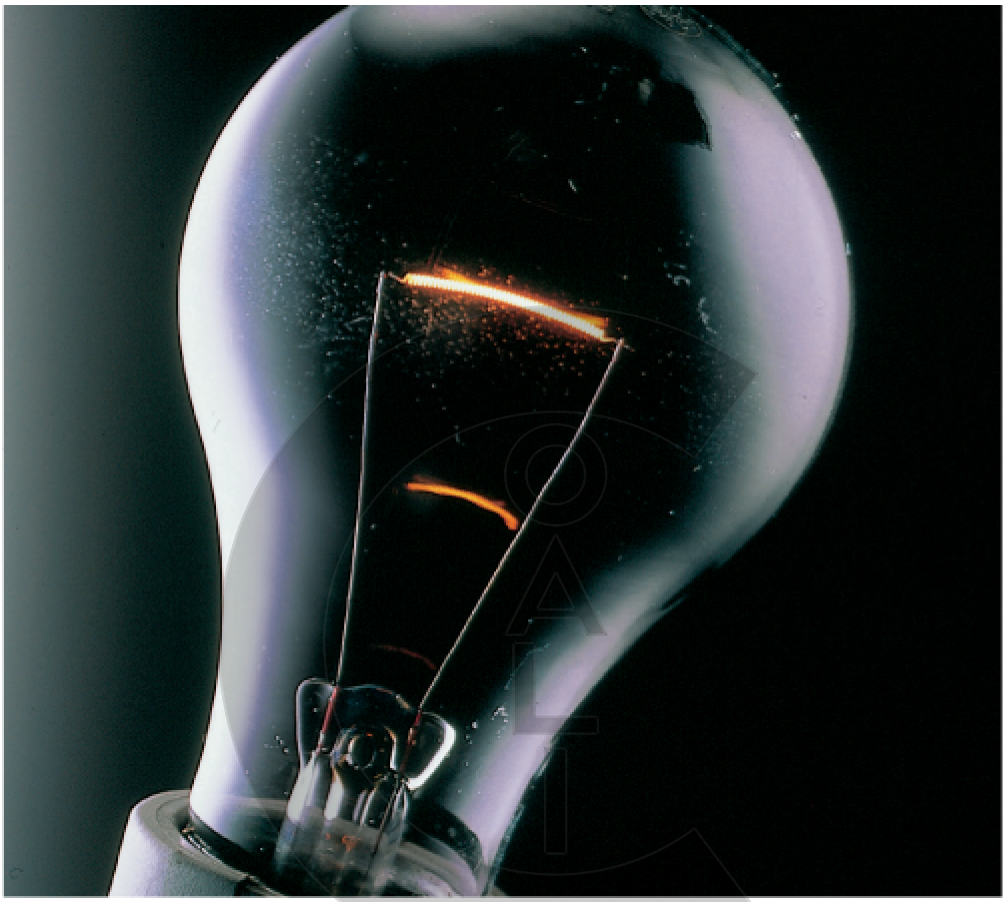


Figure 10.3: The metal filament in a lamp glows as the current passes through it. It also feels warm. This shows that the lamp produces both heat and light.

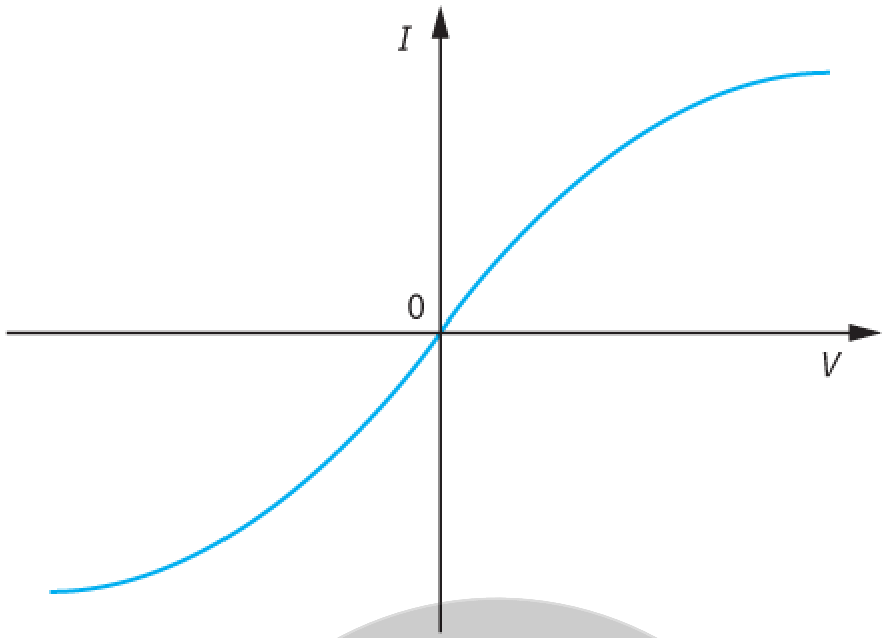


Figure 10.4: The *I*–*V* characteristic for a filament lamp.

There are some points you should notice about the graph in Figure 10.4:

- The line passes through the origin (as for an ohmic component).
- For very small currents and voltages, the graph is roughly a straight line.
- At higher voltages, the line starts to curve. The current is a bit less than we would have expected from a straight line. This suggests that the lamp's resistance has increased. You can also tell that the resistance has increased because the ratio $\frac{V}{I}$ is larger for higher voltages than for low voltages.

The graph of Figure 10.4 is not a straight line–this shows that the resistance of the lamp depends on the temperature of its filament. Its resistance may increase by a factor as large as ten between when it is cold and when it is brightest (when its temperature may be as high as 1750 °C).

Thermistors

Thermistors are components that are designed to have a resistance that changes rapidly with temperature. Thermistors ('**thermal** res**istors**') are made from metal oxides such as those of manganese and nickel.

There are two different types of thermistor:

- Negative temperature coefficient **(NTC) thermistors** – the resistance of this type of thermistor decreases with increasing temperature. Those commonly used for physics teaching may have a resistance of many thousands of ohms at room temperature, falling to a few tens of ohms at 100 °C. You should become familiar with the properties of NTC thermistors.
- Positive temperature coefficient (PTC) thermistors–the resistance of this type of thermistor rises abruptly at a definite temperature, usually around 100–150 °C.

In this course, you only need to know about NTC thermistors. So, whenever thermistors are mentioned, assume that it refers to an NTC thermistor.

The change in their resistance with temperature gives thermistors many uses. Examples include:

- water temperature sensors in cars and ice sensors on aircraft wings – if ice builds up on the wings, the thermistor 'senses' this temperature drop and a small heater is activated to melt the ice
- baby breathing monitors–the baby rests on an air-filled pad, and as he or she breathes, air from the pad passes over a thermistor, keeping it cool; if the baby stops breathing, the air movement stops, the

thermistor warms up and an alarm sounds

- fire sensors – a rise in temperature activates an alarm
- overload protection in electric razor sockets – if the razor overheats, the thermistor's resistance decreases, the current increases rapidly and cuts off the circuit.

Questions

3 The two graphs in Figure 10.5 show the *I*–*V* characteristics of a metal wire at two different temperatures, θ_1 and θ_2.

a Calculate the resistance of the wire at each temperature.

b State which is the higher temperature, θ_1 or θ_2.

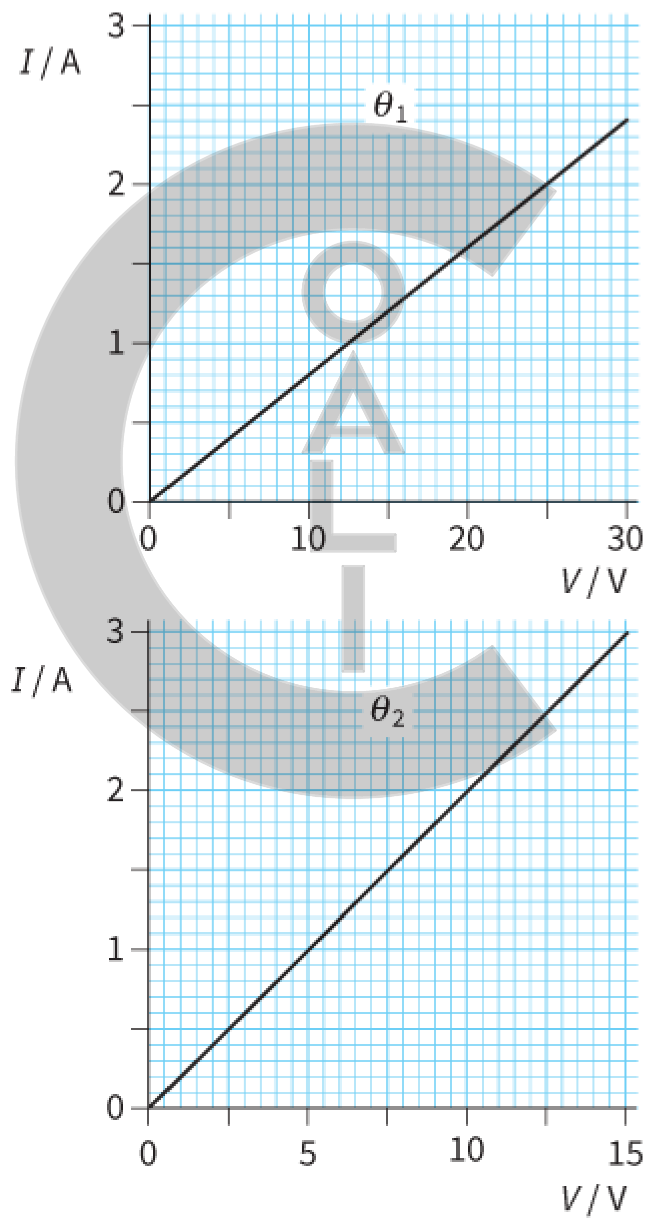


Figure 10.5: *I*–*V* graphs for a wire at two different temperatures. For Question 3.

4 The graph in Figure 10.6 shows the *I*–*V* characteristics of two electrical components, a filament lamp and a length of steel wire.

a Identify which curve relates to each component.

b State the voltage at which both have the same resistance.

c Determine the resistance at the voltage stated in part **b**.

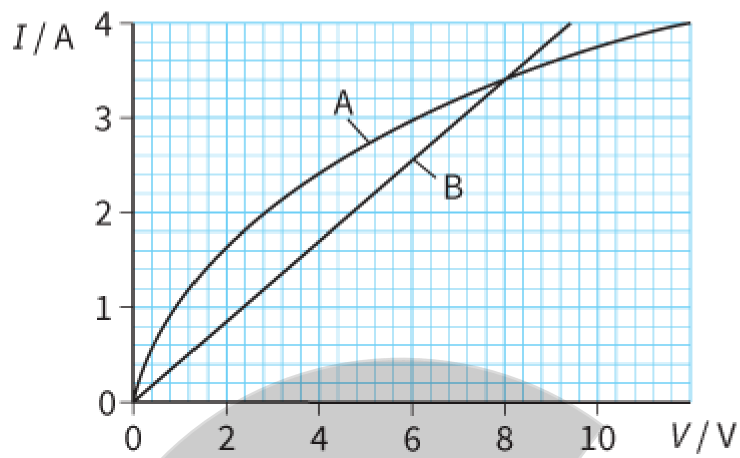


Figure 10.6: For Question 4.

Diodes

The semiconductor diode is another example of a non-ohmic conductor. A diode is any component that allows electric current in only one direction. Most diodes are made of semiconductor materials. One type, the light-emitting diode or LED, gives out light when it conducts.

Figure 10.7 shows the *I*–*V* characteristic for a diode. There are some points you should notice about this graph.

- We have included positive and negative values of current and voltage. This is because, when connected one way round, forward-biased, the diode conducts and has a fairly low resistance. Connected the other way round, reverse-biased, it allows only a tiny current and has almost infinite resistance.
- For positive voltages less than about 0.6 V, the current is almost zero and hence the diode has almost infinite resistance. It starts to conduct suddenly at its **threshold voltage**. The resistance of the diode decreases dramatically for voltages greater than 0.6 V.

KEY IDEA

Most modern diodes are made from silicon and will start conducting when there is a potential difference of about 0.6 V across them. You need to remember this key 0.6 V value.

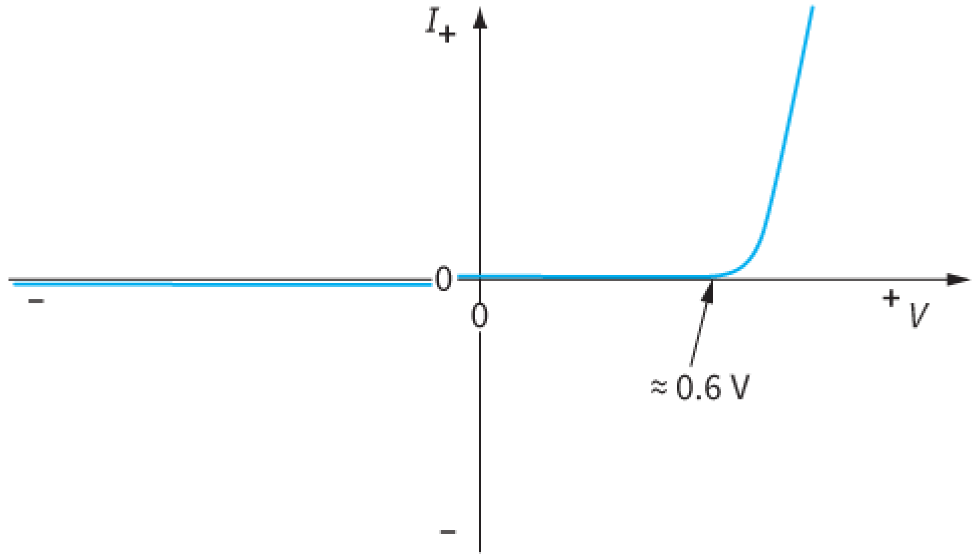


Figure 10.7: The current against potential difference (I–V) characteristic for a diode. The graph is not a straight line. A diode does not obey Ohm's law.

The resistance of a diode depends on the potential difference across it. From this we can conclude that it does not obey Ohm's law; it is a non-ohmic component.

Diodes are used as rectifiers. They allow current to pass in one direction only and so can be used to convert alternating current into direct current. (There is more about this in Chapter 27.) Most modern diodes are made from silicon and will start conducting when there is a potential difference of about 0.6 V across them. You need to remember this key 0.6 V value.

LEDs have traditionally been used as indicator lamps to show when an appliance is switched on. Newer versions, some of which produce white light, are replacing filament lamps, for example, in traffic lights and torches (flashlights) – see Figure 10.8. Although they are more expensive to manufacture, they are more energy-efficient and hence cheaper to run, so that the overall cost is less.

The threshold voltage at which an LED starts to conduct and emit light is higher than 0.6 V and depends on the colour of light it emits, but may be taken to be about 2 V.

Figure 10.8: This torch has seven white LEDs, giving a brighter, whiter light than a traditional filament lamp.

Questions

5 The graph in Figure 10.9 was obtained by measuring the resistance *R* of a particular thermistor as its temperature θ changed.

a Determine its resistance at:

i 20 °C

ii 45 °C.

b Determine the temperature when its resistance is:

i 5000 Ω

ii 2000 Ω.

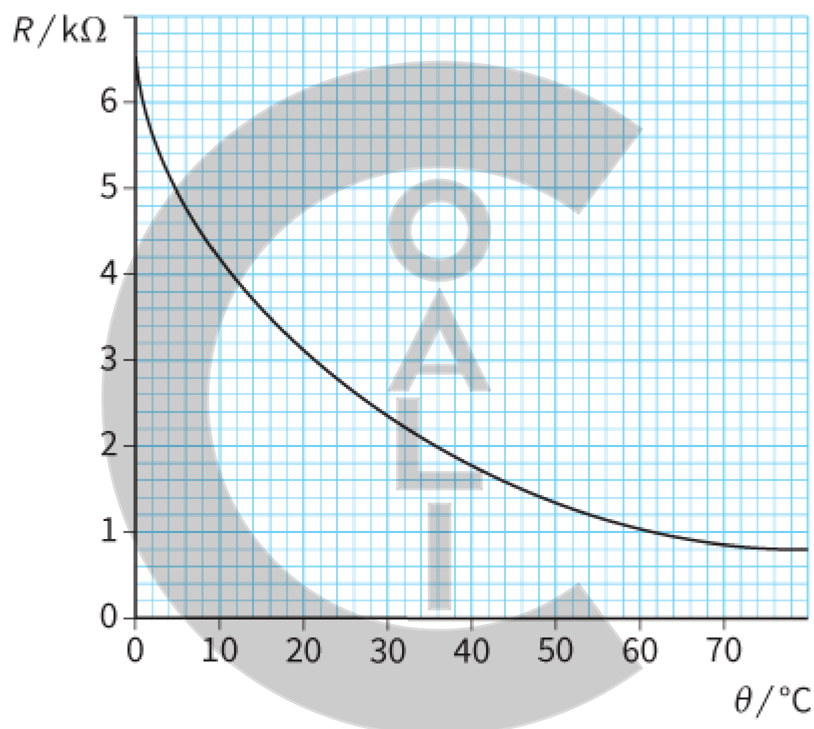


Figure 10.9: The resistance of an NTC thermistor decreases as the temperature increases. For Question 5.

6 A student connects a circuit with an NTC thermistor, a filament lamp and a battery in series. The lamp glows dimly. The student warms the thermistor with a hair dryer. What change will the student notice in the brightness of the lamp? Explain your answer.

The light-dependent resistor (LDR)

A **light-dependent resistor (LDR)** is made of a high-resistance semiconductor. If light falling on the LDR is of a high enough frequency, photons are absorbed by the semiconductor. As some photons are absorbed, electrons are released from atoms in the semiconductor. The resulting free electrons conduct electricity and the resistance of the semiconductor is reduced.

The graph in Figure 10.10 shows the variation of the resistance of a typical LDR with light intensity. Only a narrow range of light intensity, measured in lux, is shown. A typical LDR will have a resistance of a few hundred ohms in sunlight, but in the dark its resistance will be millions of ohms.

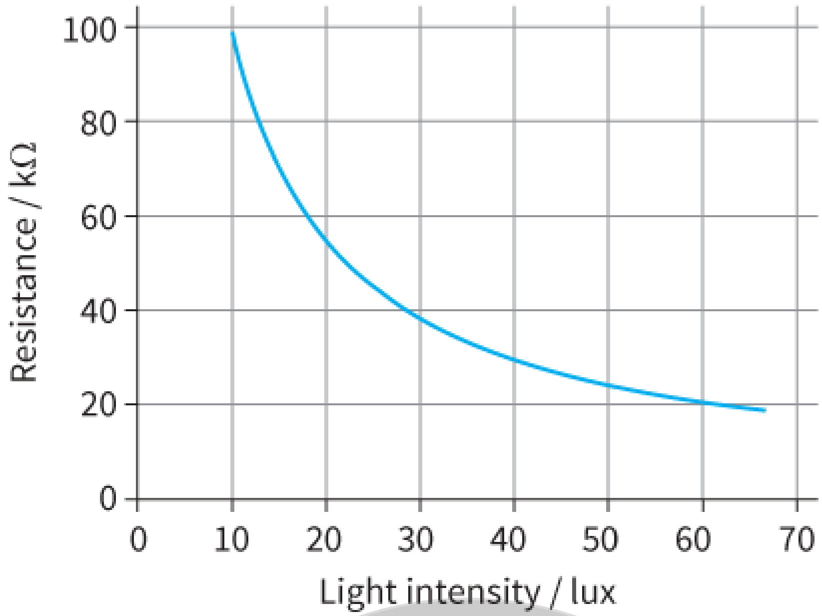


Figure 10.10: Resistance plotted against light intensity for an LDR.

Understanding the origin of resistance

To understand a little more about the origins of resistance, it is helpful to look at how the resistance of a pure metal wire changes as its temperature is increased. This is shown in the graph in Figure 10.11. You will see that the resistance of the pure metal increases linearly as the temperature increases from 0 °C to 100 °C. Compare this with the graph in Figure 10.9 for an NTC thermistor; the thermistor's resistance decreases very dramatically over a narrow temperature range.

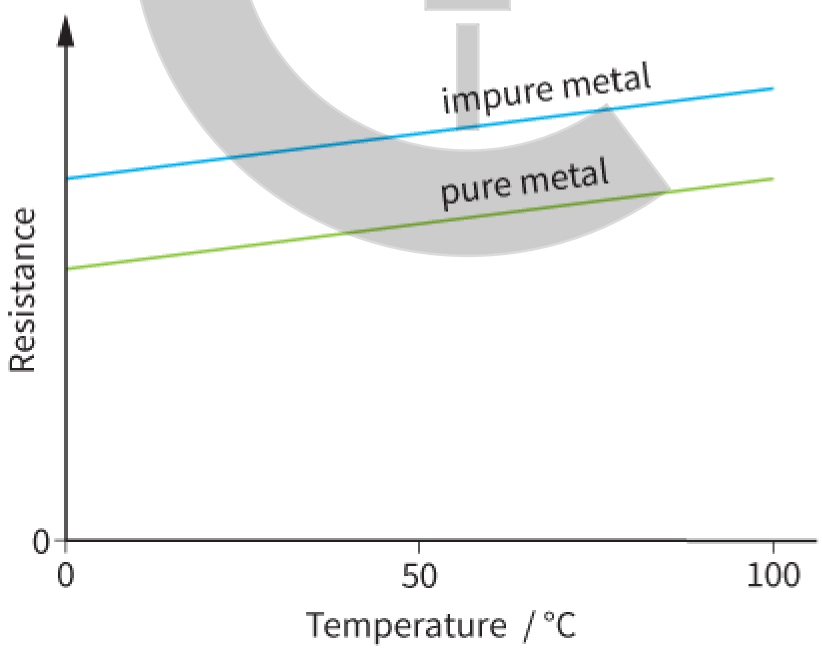


Figure 10.11: The resistance of a metal increases gradually as its temperature is increased. The resistance of an impure metal wire is greater than that of a pure metal wire of the same dimensions.

Figure 10.11 also shows how the resistance of the metal changes if it is slightly impure. The resistance of an impure metal is greater than that of the pure metal and follows the same gradual upward slope. The resistance of a metal changes in this gradual way over a wide range of temperatures–from close to absolute zero up to its melting point, which may be over 2000 °C.

This suggests there are two factors that affect the resistance of a metal:

- the temperature
- the presence of impurities.

Figure 10.12 shows a simple model that explains what happens in a metal when electrons flow through it.

In a metal, a current is due to the movement of free electrons. At low temperatures, they can move easily past the positive ions (Figure 10.12a). However, as the temperature is raised, the ions vibrate with larger amplitudes. The electrons collide more frequently with the vibrating ions, and this decreases their mean drift velocity. They lose energy to the vibrating ions (Figure 10.12b).

If the metal contains impurities, some of the atoms will be of different sizes (Figure 10.12c). Again, this disrupts the free flow of electrons. In colliding with impurity atoms, the electrons lose energy to the vibrating atoms.

You can see that electrons tend to lose energy when they collide with vibrating ions or impurity atoms. They give up energy to the metal, so it gets hotter. The resistance of the metal increases with the temperature of the wire because of the decrease in the mean drift velocity of the electrons.

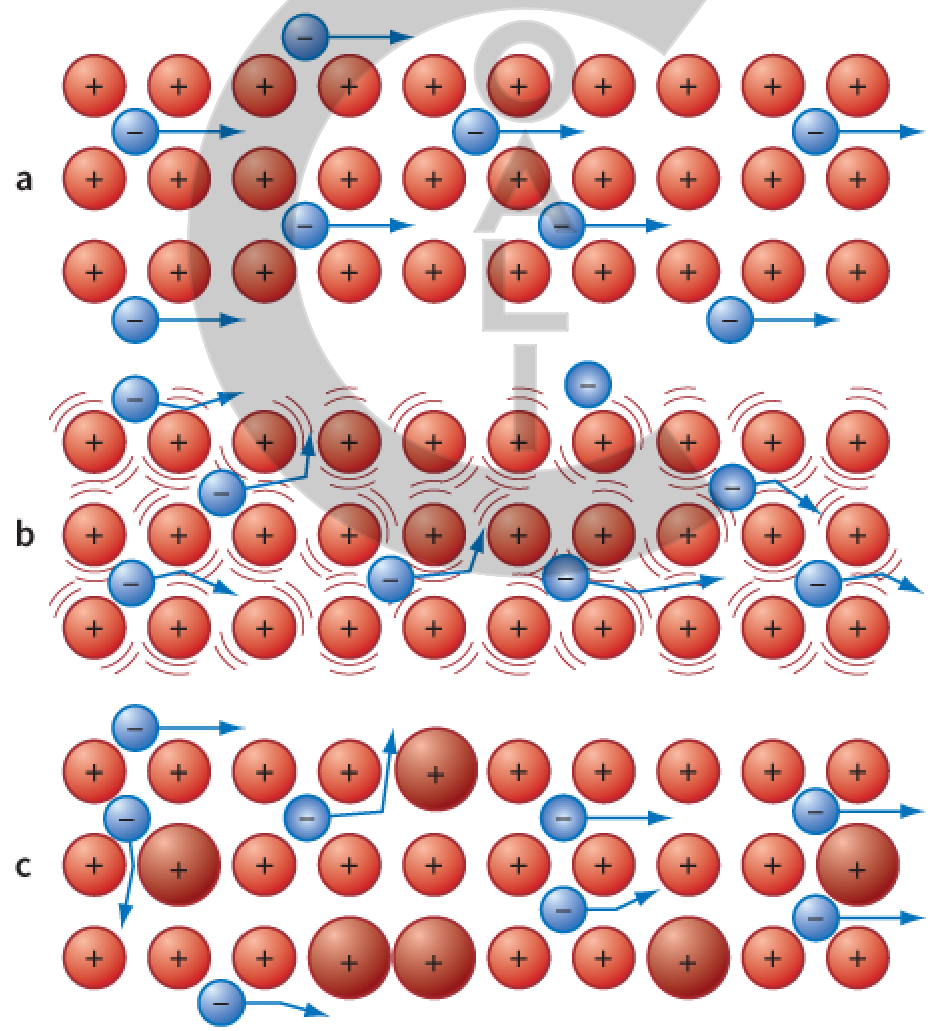


Figure 10.12: A model of the origins of resistance in a metal. **a:** At low temperatures, electrons flow relatively freely. **b:** At higher temperatures, the electrons are obstructed by the vibrating ions and they make very frequent collisions with the ions. **c:** Impurity atoms can also obstruct the free flow of electrons.

Conduction in semiconductors is different. At low temperatures, there are few **delocalised**, or free, electrons. For conduction to occur, electrons must have sufficient energy to free themselves from the atom they are bound to. As the temperature increases, a few electrons gain enough energy to break free of their atoms to become conduction electrons. The number of conduction electrons thus increases and so the material becomes a better conductor. At the same time, there are more electron–ion collisions, but this effect is small compared with the increase in the number of conduction electrons.

Question

7 The resistance of a metal wire changes with temperature. This means that a wire could be used to sense changes in temperature, in the same way that a thermistor is used.

- **a** Suggest **one** advantage a thermistor has over a metal wire for this purpose.
- **b** Suggest **one** advantage a metal wire has over a thermistor.

10.4 Resistivity

The resistance of a particular wire depends on its size and shape. A long wire has a greater resistance than a short one, provided it is of the same thickness and material. A thick wire has less resistance than a thin one. For a metal in the shape of a wire, *R* depends on the following factors:

- length *L*
- cross-sectional area *A*
- the material the wire is made from
- the temperature of the wire.

At a constant temperature, the resistance is directly proportional to the length of the wire and inversely proportional to its cross-sectional area:

$$\text{resistance} \propto \text{length}$$

and

$$\text{resistance} \propto \frac{1}{\text{cross-sectional area}}$$

We can see how these relate to the formulae for adding resistors in series and in parallel:

- If we double the length of a wire it is like connecting two identical resistors in series; their resistances add to give double the resistance. The resistance is proportional to the length.
- Doubling the cross-sectional area of a wire is like connecting two identical resistors in parallel; their combined resistance is halved (since $\frac{1}{R_{total}} = \frac{1}{R} + \frac{1}{R}$).

  Hence the resistance is inversely proportional to the cross-sectional area.

Combining the two proportionalities for length and cross-sectional area, we get:

$$\text{resistance} \propto \frac{1}{\text{cross-sectional area}}$$

or

$$R \propto \frac{L}{A}$$

But the resistance of a wire also depends on the material it is made of. Copper is a better conductor than steel, steel is a better conductor than silicon, and so on. So if we are to determine the resistance *R* of a particular wire, we need to take into account its length, its cross-sectional area and the material. The relevant property of the material is its **resistivity**, for which the symbol is ρ (Greek letter **rho**).

The word equation for resistance is:

$$\text{resistance} = \frac{\text{resistivity} \times \text{length}}{\text{cross-sectional area}}$$

$$R = \frac{\rho L}{A}$$

KEY EQUATION

$$\text{resistance} = \frac{\text{resistivity} \times \text{length}}{\text{cross-sectional area}}$$

$$R = \frac{\rho L}{A}$$

We can rearrange this equation to give an equation for resistivity. The resistivity of a material is defined by the following word equation:

$$\text{resitivity} = \frac{\text{resistance} \times \text{cross-sectional area}}{\text{length}}$$

$$\rho = \frac{RA}{L}$$

KEY EQUATION

$$\text{resitivity} = \frac{\text{resistance} \times \text{cross-sectional area}}{\text{length}}$$

$$\rho = \frac{RA}{L}$$

Values of the resistivities of some typical materials are shown in Table 10.2. Notice that the units of resistivity are ohm metres (Ω m); this is not the same as ohms per metre.

Material	**Resistivity / Ω m**
silver	1.60×10^{-8}
copper	1.69×10^{-8}
nichrome[(a)]	1.30×10^{-8}
aluminium	3.21×10^{-8}
lead	20.8×10^{-8}
manganin[(b)]	44.0×10^{-8}
eureka[(c)]	49.0×10^{-8}
mercury	69.0×10^{-8}
graphite	800×10^{-8}
germanium	0.65
silicon	2.3×10^{3}
Pyrex glass	10^{12}
PTFE[(d)]	10^{13}–10^{16}
quartz	5×10^{16}

[(a)] Nichrome – an alloy of nickel, copper and aluminium used in electric heaters because it does not oxidise at 1000 °C.

[(b)] Manganin – an alloy of 84% copper, 12% manganese and 4% nickel.

[(c)] Eureka (constantan) – an alloy of 60% copper and 40% nickel.

[(d)] PTFE – Poly(tetrafluoroethene) or Teflon.

Table 10.2: Resistivities of various materials at 20 °C.

WORKED EXAMPLE

1 Find the resistance of a 2.6 m length of eureka wire with cross-sectional area 2.5×10^{-7} m^2.

Step 1 Use the equation for resistance:

$$\text{resistance} = \frac{\text{resistivity} \times \text{length}}{\text{cross-sectional area}}$$

$$R = \frac{\rho L}{A}$$

Step 2 Substitute values from the question and use the value for ρ from Table 10.2:

$$R = \frac{49.0 \times 10^{-8} \times 2.6}{2.5 \times 10^{-7}}$$

$$= 5.1\Omega$$

So the wire has a resistance of 5.1 Ω.

Resistivity and temperature

Resistivity, like resistance, depends on temperature. For a metal, resistivity increases with temperature. As we saw earlier, this is because there are more frequent collisions between the conduction electrons and the vibrating ions of the metal.

Questions

8 Use the resistivity value quoted in Table 10.2 to calculate the lengths of 0.50 mm diameter manganin wire needed to make resistance coils with resistances of:

a 1.0 Ω

b 5.0 Ω

c 10 Ω.

9 1.0 cm^3 of copper is drawn out into the form of a long wire of cross-sectional area 4.0×10^{-7} m^2. Calculate its resistance. (Use the resistivity value for copper from Table 10.2.)

10 A 1.0 m length of copper wire has a resistance of 0.50 Ω.

a Calculate the resistance of a 5.0 m length of the same wire.

b What will be the resistance of a 1.0 m length of copper wire having half the diameter of the original wire?

11 A piece of steel wire has a resistance of 10 Ω. It is stretched to twice its original length. Compare its new resistance with its original resistance.

REFLECTION

Imagine you are helping a younger cousin who is studying for her IGCSE (or similar course). She finds it difficult to understand why the resistivity does not change when the dimensions of a sample are changed, but resistance does.

Think about how you might help her understand.

Now that it is completed, what are your first thoughts about this activity? Are they mostly positive or negative?

SUMMARY

A conductor obeys Ohm's law if the current in it is directly proportional to the potential difference across its ends.
Ohmic components include a wire at constant temperature and a resistor.
Non-ohmic components include a filament lamp and a light-emitting diode.
A semiconductor diode allows current in one direction only.
As the temperature of a metal increases, so does its resistance.
A thermistor is a component that shows a rapid change in resistance over a narrow temperature range. The resistance of an NTC thermistor decreases as its temperature is increased.
The resistivity ρ of a material is defined as: $$\rho = \frac{RA}{L}$$ where R is the resistance of a wire of that material, A is its cross-sectional area and L is its length. The unit of resistivity is the ohm metre (Ω m).

EXAM-STYLE QUESTIONS

1 An element of an electric fire is made up from a length of nichrome wire of diameter 0.40 mm and length 5.0 m.

The resistance of this element is R_1.

Another element, also made from nichrome, for a different electric fire, has a length of 2.0 m and a diameter of 0.20 mm. This element has a resistance of R_2.

What is the relationship between R_1 and R_2? **[1]**

A $R_2 = 0.80\ R_1$

B $R_2 = 1.6\ R_1$

C $R_2 = 5.0\ R_1$

D $R_2 = 10\ R_1$

2 This is a circuit.

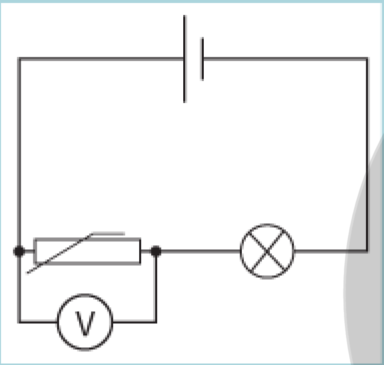


Figure 10.13

Which line in the table shows the changes to the lamp and the voltmeter reading when the temperature rises? **[1]**

	Lamp	**Voltmeter reading**
A	gets brighter	decreases
B	gets brighter	increases
C	gets dimmer	decreases
D	gets dimmer	increases

Table 10.3

3 This shows the I–V characteristic of an electrical component.

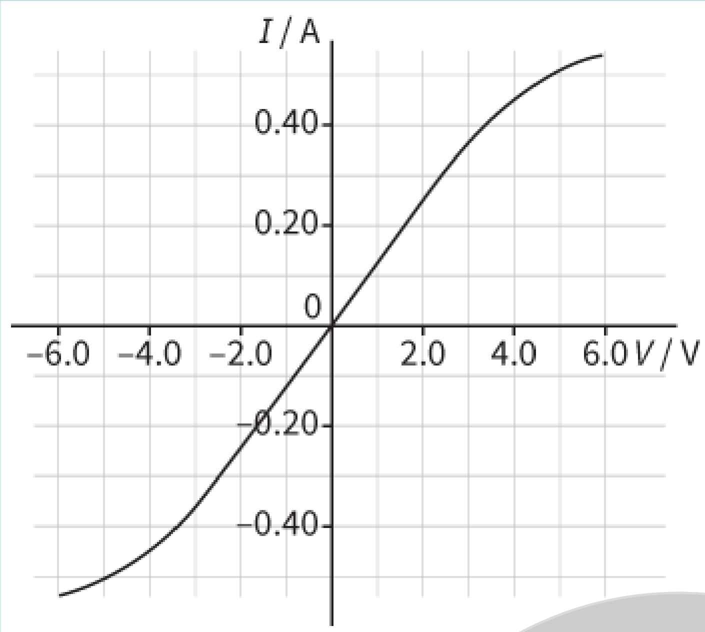


Figure 10.14

a Calculate the resistance of the component when the potential difference across it is:

i 2.0 V **[2]**

ii 5.0 V. **[1]**

b Suggest what the component is. **[1]**

[Total: 4]

4 A student connects a thermistor to a battery and an ammeter. He places the thermistor in a beaker of water and gradually heats the water from 10 °C to its boiling point, recording the value of the current as he does so. He then plots a graph of the current in the thermistor against the temperature of the water.

a Sketch the graph you would expect the student to obtain from the experiment. **[1]**

b Explain how the student could now use the thermistor as a thermometer. **[2]**

[Total: 3]

5 a Describe the difference between the conduction processes in copper and in silicon, a semiconductor. **[3]**

b Explain why the resistance of a metallic conductor increases with temperature while that of a semiconductor decreases. **[3]**

[Total: 6]

6 A nichrome wire has a length of 1.5 m and a cross-sectional area of 0.0080 mm^2. The resistivity of nichrome is $1.30 \times 10^{-8}\ \Omega$ m.

a Calculate the resistance of the wire. **[2]**

b Calculate the length of this wire that would be needed to make an element of an electric heater of resistance 30 Ω. **[2]**

[Total: 4]

7 This is a circuit.

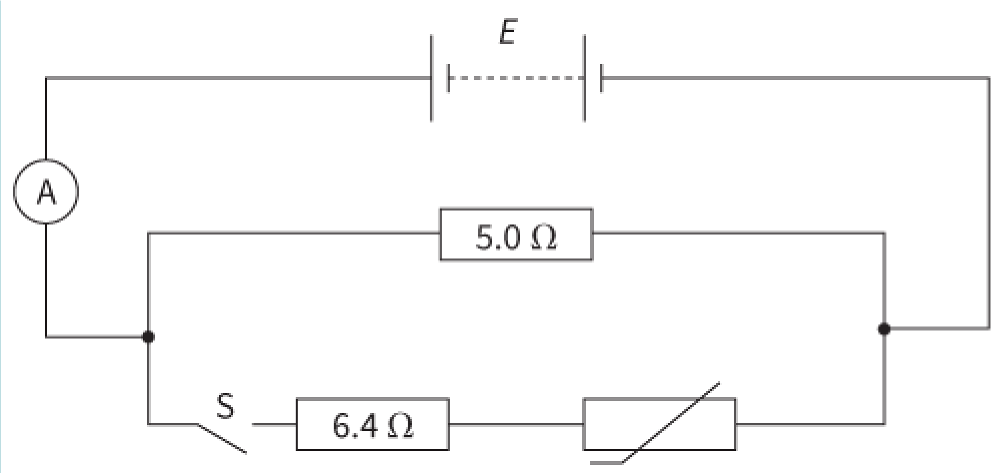


Figure 10.15

a When switch S is open the current in ammeter A is 0.48 A. Calculate the e.m.f. of the battery. You may assume the battery has negligible internal resistance. **[2]**

b When switch S is closed the current in the ammeter increases to 0.72 A.

i Determine the current in the 6.4 Ω resistor. **[1]**

ii State the current in the thermistor. **[1]**

c State and explain how the reading on the ammeter changes when the temperature of the thermistor is increased. **[3]**

[Total: 7]

8 **a** Explain why the resistance of a metal increases when its temperature increases. **[2]**

b State **two** other factors that determine the resistance of a stated length of wire. **[2]**

c When a potential difference of 1.5 V is applied across a 5.0 m length of insulated copper wire, a current of 0.24 A is measured in it.

i Calculate the resistance of the length of wire. **[2]**

ii The resistivity of copper is 1.69×10^{-8} Ω m. Calculate the diameter of the wire. **[3]**

d The wire is now made into a tight bundle. State and explain how you would expect the current in it to change. **[3]**

[Total: 12]

9 This diagram shows a piece of silicon of width 32 mm and length 36 mm. The resistance of the silicon between the points P and Q is 1.1 MΩ. Silicon has a resistivity of 2.3×10^{3} Ω m.

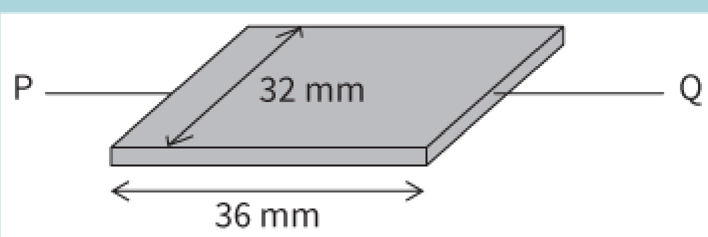


Figure 10.16

a Calculate the thickness of the piece of silicon. **[3]**

b Calculate the current that would pass through the silicon if a potential difference of 12 V were applied across P and Q. **[2]**

c Describe how the current would change if it were large enough to cause the silicon to become significantly warmer. **[3]**

[Total: 8]

10 A student is investigating the properties of a semiconducting diode. This diagram shows the circuit she builds.

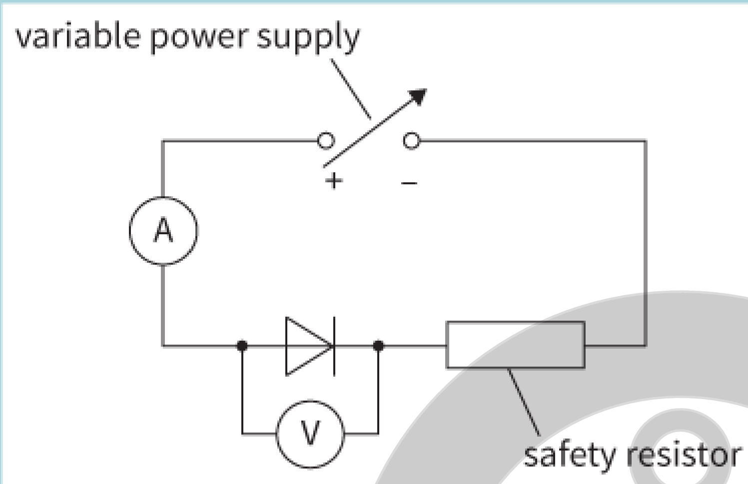


Figure 10.16

a **i** Sketch a graph to show how the current in the diode would vary as the voltage across it is increased from 0 V to 1.0 V. **[1]**

ii The supply is now connected in the reverse direction and once more the potential difference across the diode is increased from 0 V to 1.0 V. Complete the *I*–*V* graph. **[1]**

b Suggest why the safety resistor is required. **[2]**

c When the potential difference across the safety resistor is 1.4 V, the current in it is 20 mA. Calculate the resistance of the safety resistor. **[2]**

[Total: 6]

11 **a** Explain what is meant by an **ohmic conductor**. **[2]**

b **i** Sketch a graph of resistance *R* against voltage *V* for a wire of pure iron kept at constant temperature. Label this line X. **[1]**

ii Sketch a graph of resistance *R* against voltage *V* for a second wire of impure iron, of the same diameter and the same length, which is kept at the same temperature. Label this line Y. **[1]**

iii Explain how the graphs would change if the wires were kept at a higher, but still constant, temperature. **[1]**

c Deduce how the resistance of a wire made of pure iron would change if both the diameter and the length were doubled. **[3]**

[Total: 8]

12 The readings in this table are recorded from an experiment to measure the resistivity of silver.

Diameter of the wire	0.40 ± 0.02 mm

Length of the wire	2.25 ± 0.05 m
Resistance of the wire	0.28 ± 0.01 Ω

Table 10.4

a Calculate the resistivity of silver. **[2]**

b **i** Calculate the percentage uncertainty in each of the variables. **[2]**

ii Use your answers to **i** to calculate the absolute uncertainty in the value of the resistivity obtained in the experiment. **[2]**

[Total: 6]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
state and understand Ohm's law	10.2			
recognise ohmic and non-ohmic components	10.2, 10.3			
recognise and understand the changes in the resistance of metals and thermistors when there is a change in their temperature	10.3, 10.4			
understand that a light-dependent resistor is a component whose resistance decreases as the light level increases	10.3			
understand that resistivity ρ of a material is defined as: $\rho = \frac{RA}{L}$ where *R* is the resistance of a wire of that material, *A* is its cross-sectional area and *L* is its length. The unit of resistivity is the ohm metre (Ω m).	10.4			


PE5

〉Chapter 11

Practical circuits

LEARNING INTENTIONS

In this chapter you will learn how to:

- explain the effects of internal resistance on terminal p.d. and power output of a source of e.m.f.
- explain the use of potential divider circuits
- solve problems involving the potentiometer as a means of comparing voltages.

BEFORE YOU START

How confident are you on the concepts of terminal potential difference and e.m.f.? Without looking at a textbook, either write down the meaning of each or discuss it with a partner. This will help you in the first part of this chapter, which further develops the idea of e.m.f. and illustrates why the terminal p.d. and the e.m.f. are different.

THE FIRST ELECTRICAL CELL: AN HISTORICAL MYSTERY

The Italian Alessandro Volta (Figure 11.1a) is generally credited with inventing the first battery. He devised it after his friend and rival Luigi Galvani had shown that a (dead) frog's leg could be made to twitch if an electrically charged plate was connected to it. Volta's battery consisted of alternate discs of copper and zinc, separated by felt soaked in brine–see Figure 11.1b.

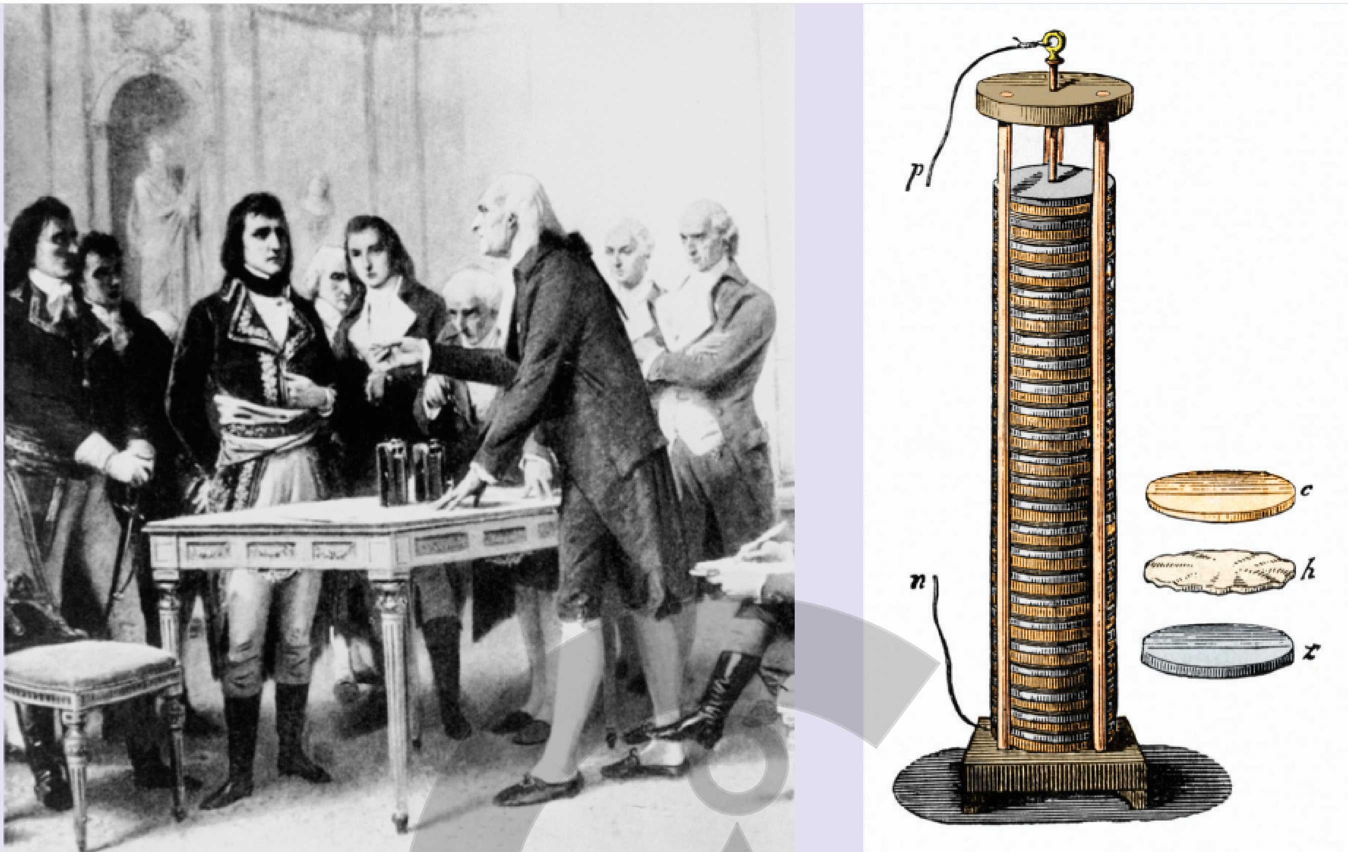


Figure 11.1: a Alessandro Volta demonstrating his newly invented pile (battery) to the French Emperor Napoleon. **b** Volta's pile, showing (top to bottom) discs of copper, wet felt and zinc.

However, there is evidence that earlier technologists may have beaten him by over 1000 years. In 1936, a small pot was discovered during an archaeological dig near Baghdad. The pot was sealed with pitch, and inside the pot there was a copper cylinder surrounding an iron rod. When filled with an acid, perhaps vinegar, a potential difference of around 1.5 volts could be produced between the copper and the iron.

It has been suggested that this battery might have been used to electroplate metal objects with gold. So, did Volta really invent the battery, or did he just rekindle an art that had been lost for more than a millennium?

11.1 Internal resistance

You will have learnt that, when you use a power supply or other source of e.m.f., you cannot assume that it is providing you with the exact voltage across its terminals as suggested by the value of its e.m.f. There are several reasons for this. For example, the supply may not be made to a high degree of precision, or the batteries may have become flat, and so on. However, there is a more important factor, which is that all sources of e.m.f. have an **internal resistance**. For a power supply, this may be due to the wires and components inside, whereas for a cell the internal resistance is due to the chemicals within it. Experiments show that the voltage across the terminals of the power supply depends on the circuit of which it is part. In particular, the voltage across the power supply terminals decreases if it is required to supply more current.

Figure 11.2 shows a circuit you can use to investigate this effect, and a sketch graph showing how the voltage across the terminals of a power supply might decrease as the supplied current increases.

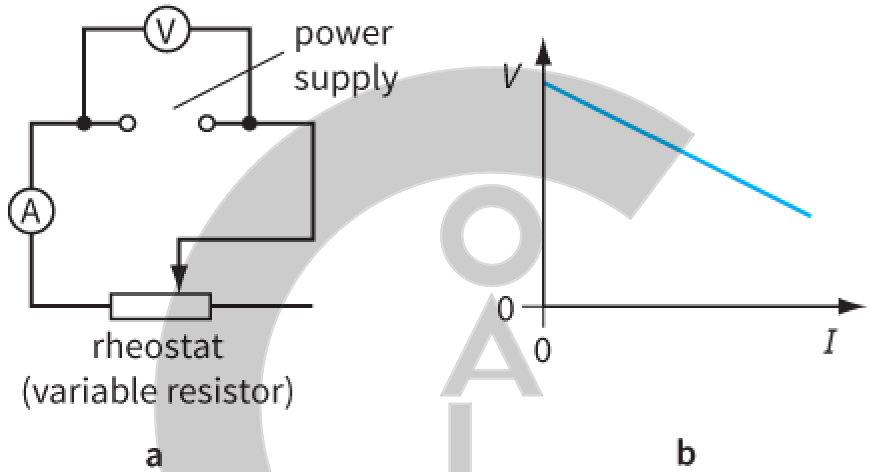


Figure 11.2: a A circuit for determining the e.m.f. and internal resistance of a supply; **b** typical form of results.

The charges moving round a circuit have to pass through the external components **and** through the internal resistance of the power supply. These charges gain electrical energy from the power supply. This energy is lost as thermal energy as the charges pass through the external components and through the internal resistance of the power supply. Power supplies and batteries get warm when they are being used. (Try using a cell to light a small torch bulb; feel the cell before connecting to the bulb, and then feel it again after the bulb has been lit for about 15 seconds.)

The reason for this heating effect is that some of the electrical potential energy of the charges is transformed to internal energy as they do work against the internal resistance of the cell.

It can often help to solve problems if we show the internal resistance r of a source of e.m.f. explicitly in circuit diagrams (Figure 11.3). Here, we are representing a cell as if it were a 'perfect' cell of e.m.f. E, together with a separate resistor of resistance r. The dashed line enclosing E and r represents the fact that these two are, in fact, a single component.

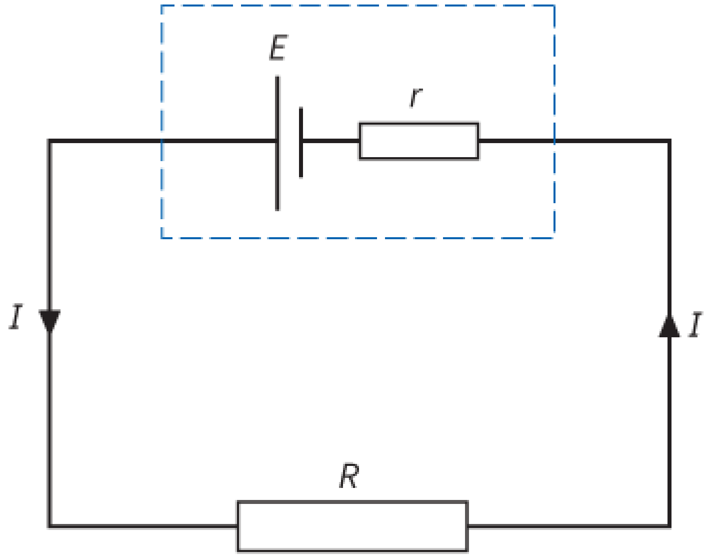


Figure 11.3: It can be helpful to show the internal resistance *r* of a cell (or a supply) in a circuit diagram.

Now we can determine the current when this cell is connected to an external resistor of resistance *R*. You can see that *R* and *r* are in series with each other. The current *I* is the same for both of these resistors. The combined resistance of the circuit is thus $R + r$, and we can write:

$$E = I(R + r) \quad \text{or} \quad E = IR + Ir$$

We cannot measure the e.m.f. *E* of the cell directly, because we can only connect a voltmeter across its terminals. This **terminal p.d.** *V* across the cell is always the same as the p.d. across the external resistor.

Therefore, we have:

$$V = IR$$

This will be less than the e.m.f. *E* by an amount *Ir*. The quantity *Ir* is the potential difference across the internal resistor. If we combine these two equations, we get:

$$V = E - Ir$$

where *E* is the emf of the source, *I* is the current in the source and *r* is the internal resistance of the source.

or

terminal p.d. = e.m.f. − p.d across the internal resistance

The potential difference across the internal resistance indicates the energy transferred to the internal resistance of the supply. If you short-circuit a battery with a piece of wire, a large current will flow, and the battery will get warm as energy is transferred within it. This is also why you may damage a power supply by trying to make it supply a larger current than it is designed to give.

KEY EQUATION

Potential difference across a power source:

$$V = E - Ir$$

WORKED EXAMPLE

1 There is a current of 0.40 A when a battery of e.m.f. 6.0 V is connected to a resistor of 13.5 Ω. Calculate the internal resistance of the cell.

Step 1 Substitute values from the question in the equation for e.m.f.:

$$E = 6.0\ \text{V}, \quad I = 0.40\ \text{A}, \quad R = 13.5\Omega$$
$$E = IR + Ir$$

$$6.0 = 0.40 \times 13.5 + 0.40 \times r$$
$$= 5.4 + 0.40r$$

Step 2 Rearrange the equation to make *r* the subject and solve:

$$6.0 - 5.4 = 0.40r$$
$$0.60 = 0.40r$$
$$r = \frac{0.60}{0.40} = 1.5\Omega$$

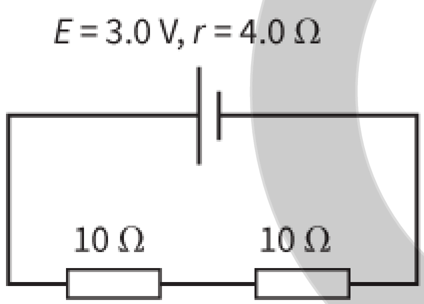
Questions

1 A battery of e.m.f. 5.0 V and internal resistance 2.0 Ω is connected to an 8.0 Ω resistor. Draw a circuit diagram and calculate the current in the circuit.

2 **a** Calculate the current in each circuit in Figure 11.4.

b Calculate also the potential difference across the internal resistance for each cell, and the terminal p.d.

i



ii

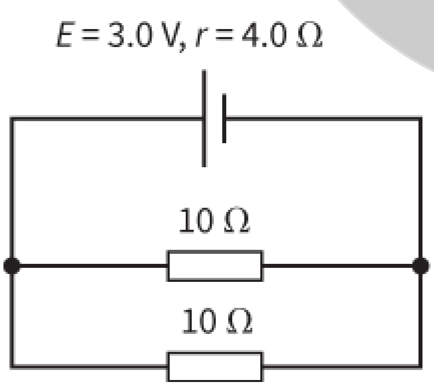


Figure 11.4: For Question 2.

3 Four identical cells, each of e.m.f. 1.5 V and internal resistance 0.10 Ω, are connected in series. A lamp of resistance 2.0 Ω is connected across the four cells. Calculate the current in the lamp.

PRACTICAL ACTIVITY 12.1

Determining e.m.f. and internal resistance

You can get a good idea of the e.m.f. of an isolated power supply or a battery by connecting a digital voltmeter across it. A digital voltmeter has a very high resistance (~10^7 Ω), so only a tiny current will pass through it. The potential difference across the internal resistance will then only be a tiny fraction of the e.m.f. If you want to determine the internal resistance r as well as the e.m.f. E, you need to use a circuit like that shown in Figure 11.2. When the variable resistor is altered, the current in the circuit changes and measurements can be recorded of the circuit current I and terminal p.d. V. The internal resistance r can be found from a graph of V against I (Figure 11.5).

Compare the equation $V = E - Ir$ with the equation of a straight line $y = mx + c$. By plotting V on the y-axis and I on the x-axis, a straight line should result. The intercept on the y-axis is E, and the gradient is $-r$.

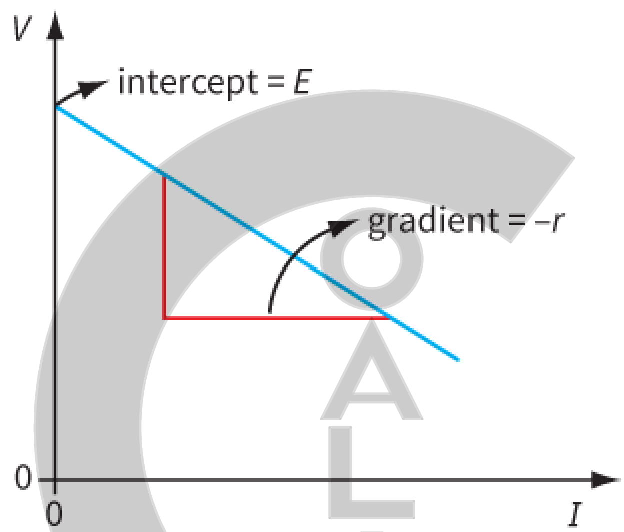


Figure 11.5: E and r can be found from this graph.

Questions

4 When a high-resistance voltmeter is placed across an isolated battery, its reading is 3.0 V. When a 10 Ω resistor is connected across the terminals of the battery, the voltmeter reading drops to 2.8 V. Use this information to determine the internal resistance of the battery.

5 The results of an experiment to determine the e.m.f. E and internal resistance r of a power supply are shown in Table 11.1. Plot a suitable graph and use it to find E and r.

V / V	1.43	1.33	1.18	1.10	0.98
I / A	0.10	0.30	0.60	0.75	1.00

Table 11.1: Results for Question 5.

The effects of internal resistance

You cannot ignore the effects of internal resistance. Consider a battery of e.m.f. 3.0 V and of internal resistance 1.0 Ω. The **maximum current** that can be drawn from this battery is when its terminals are shorted-out. (The external resistance $R \approx 0$.) The maximum current is given by:

$$\text{maximum current} = \frac{E}{r} = \frac{3.0}{1.0} = 3.0\text{ A}$$

The **terminal p.d.** of the battery depends on the resistance of the external resistor. For an external resistor of resistance 1.0 Ω, the terminal p.d. is 1.5 V – half of the e.m.f. The terminal p.d. approaches the value of the e.m.f. when the external resistance *R* is very much greater than the internal resistance of the battery. For example, a resistor of resistance 1000 Ω connected to the battery gives a terminal p.d. of 2.997 V. This is almost equal to the e.m.f. of the battery. The more current a battery supplies, the more its terminal p.d. will decrease. An example of this can be seen when a driver tries to start a car with the headlamps on. The starter motor requires a large current from the battery, the battery's terminal p.d. drops and the headlamps dim.

Question

6 A car battery has an e.m.f. of 12 V and an internal resistance of 0.04 Ω. The starter motor draws a current of 100 A.

a Calculate the terminal p.d. of the battery when the starter motor is in operation.

b Each headlamp is rated as '12 V, 36 W'. Calculate the resistance of a headlamp.

c To what value will the power output of each headlamp decrease when the starter motor is in operation? (Assume that the resistance of the headlamp remains constant.)

11.2 Potential dividers

How can we get an output of 3.0 V from a battery of e.m.f. 6.0 V? Sometimes we want to use only part of the e.m.f. of a supply. To do this, we use an arrangement of resistors called a **potential divider** circuit.

Figure 11.6 shows two potential divider circuits, each connected across a battery of e.m.f. 6.0 V and of negligible internal resistance. The high-resistance voltmeter measures the voltage across the resistor of resistance R_2. We refer to this voltage as the output voltage, V_{out}, of the circuit. The first circuit, **a**, consists of two resistors of values R_1 and R_2. The voltage across the resistor of resistance R_2 is half of the 6.0 V of the battery. The second potential divider, **b**, is more useful. It consists of a single variable resistor. By moving the sliding contact, we can achieve any value of V_{out} between 0.0 V (slider at the bottom) and 6.0 V (slider at the top).

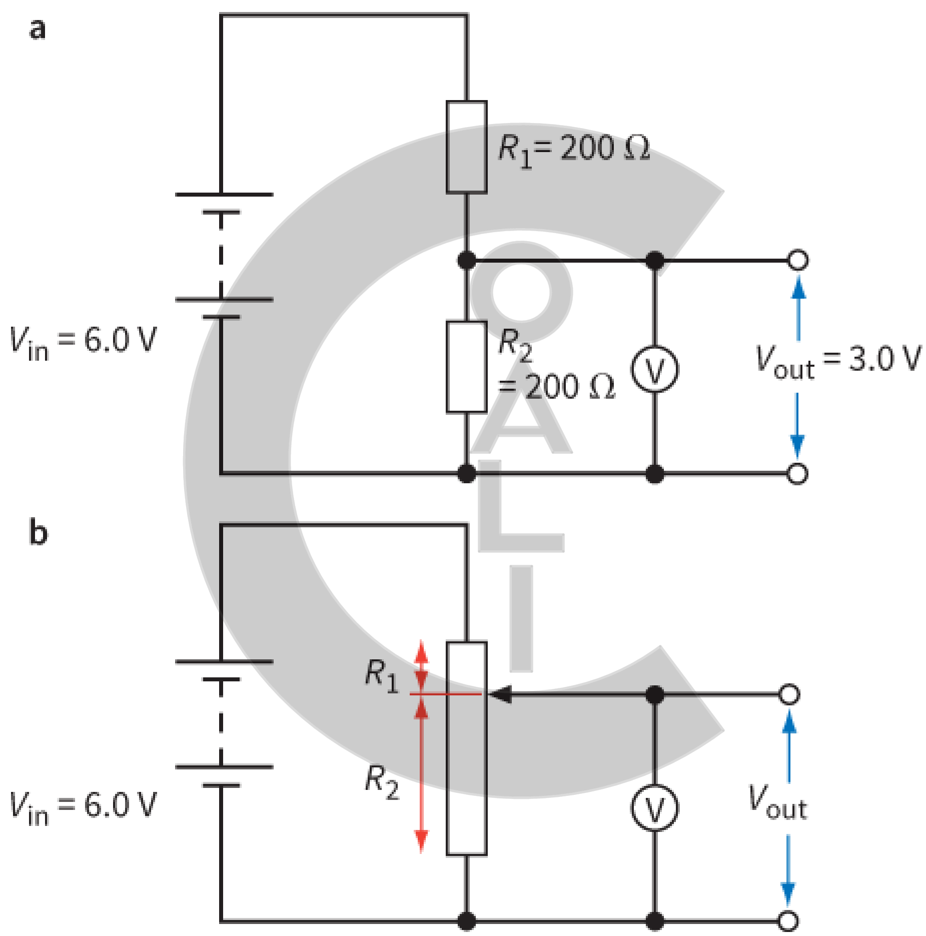


Figure 11.6: Two potential divider circuits.

The output voltage V_{out} depends on the relative values of R_1 and R_2. You can calculate the value of V_{out} using the **potential divider equation**:

$$V_{out} = \left(\frac{R_2}{R_1+R_2}\right) \times V_{in}$$

where R_2 is the resistance of the component over which the output is taken, R_1 is the resistance of the second component in the potential divider and V_{in} is the p.d. across the two components.

KEY EQUATION

Potential divider equation:

$$V_{out} = \left(\frac{R_2}{R_1+R_2}\right) \times V_{in}$$

Question

7 Determine the range of V_{out} for the circuit in Figure 11.7 as the variable resistor R_2 is adjusted over its full range from 0 Ω to 40 Ω. (Assume the supply of e.m.f. 10 V has negligible internal resistance.)

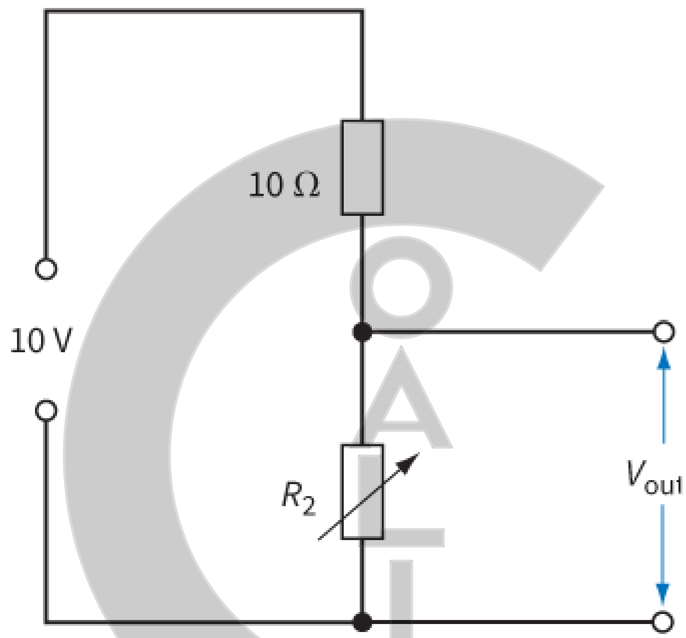


Figure 11.7: For Question 7.

11.3 Sensors

Light-dependent resistors as sensors

How is a light-dependent resistor (LDR) used as a **sensor** or **transducer**? A voltage is needed to drive the output device, such as a voltmeter, yet the LDR only produces a change in resistance. The sensor must use this change in resistance to generate the change in voltage. The solution is to place the LDR in series with a fixed resistor, as shown in Figure 11.8.

The voltage of the supply is shared between the two resistors in proportion to their resistance so, as the light level changes and the LDR's resistance changes, so does the voltage across each of the resistors. The two resistors form a potential divider whose output changes automatically with changing light intensities.

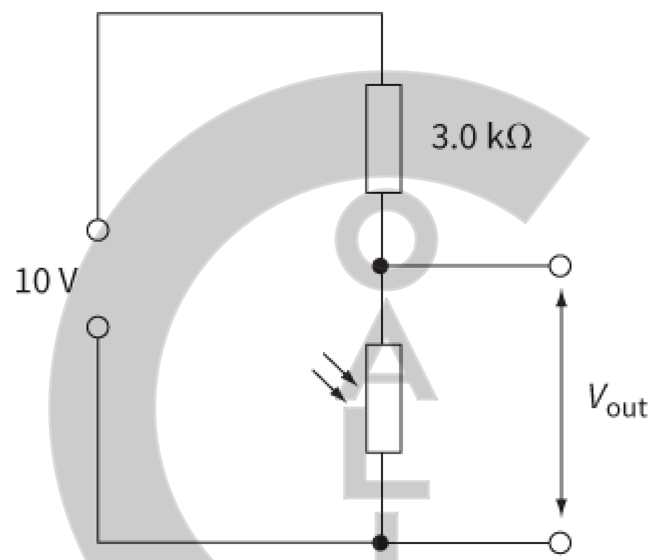


Figure 11.8: An LDR used as a sensor.

WORKED EXAMPLE

2 Using the graph in Figure 11.9, calculate V_{out} in Figure 11.8 when the light intensity is 60 lux.

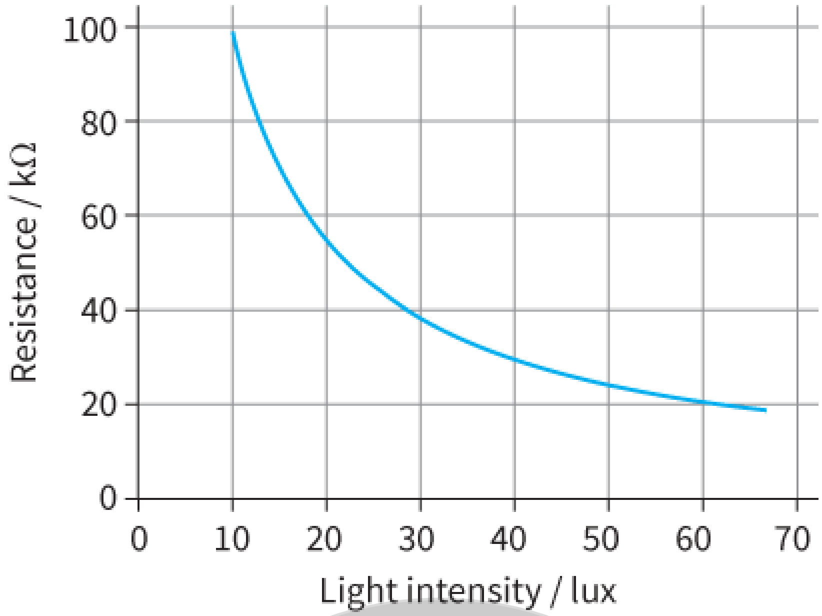


Figure 11.9: for Worked example 2 and question 8.

Step 1 Find the resistance of the LDR at 60 lux.

$R_{LDR} = 20\ \text{k}\Omega$

Step 2 Divide the total voltage of 10 V in the ratio 3 : 20. The total number of parts is 23 so:

$$V_{out} = \frac{20}{23} \times 10 = 8.70\text{V}$$

Hint: *The answer on your calculator might be 8.69565. When you give your answer to three significant figures, do not write 8.69 – you must round correctly.*

Questions

8 What is the voltage across the 3.0 kΩ resistor in Figure 11.9 when the light intensity is 10 lux?

9 The circuit shown in Figure 11.8 produces a decreasing output voltage when the light intensity increases. How can the circuit be altered to produce an increasing output voltage as the light intensity increases?

Thermistors as a sensors

The thermistors that we refer to in this course are known as **negative temperature coefficient** (NTC) thermistors. This means that, when the temperature rises, the resistance of the thermistor falls. This happens because the thermistor is made from a semiconductor material. One property of a semiconductor is that when the temperature rises the number of free electrons increases, and thus the resistance falls.

Figure 11.10 shows a graph of the resistance of a thermistor and the resistance of a metal wire plotted against temperature. You can see that the resistance of a metal wire increases with increase in temperature. A metal wire is not a negative temperature device, but it could be used as a sensing device. A thermistor is more useful than a metal wire because there is a much larger change in resistance with change in temperature. However, the change in resistance of a thermistor is not linear with temperature; indeed, it is likely to be an exponential decrease. This means that any device used to measure temperature electronically must be calibrated to take into account the resistance–temperature graph. The scale on an ordinary laboratory thermometer between 0 °C and 100 °C is divided up into 100 equal parts, each of which represents 1 °C. If the resistance of a thermistor were divided like this, the scale would be incorrect.

The thermistor can be used as a sensing device in the same way as an LDR. Instead of sensing a change in light level, it senses a change in temperature.

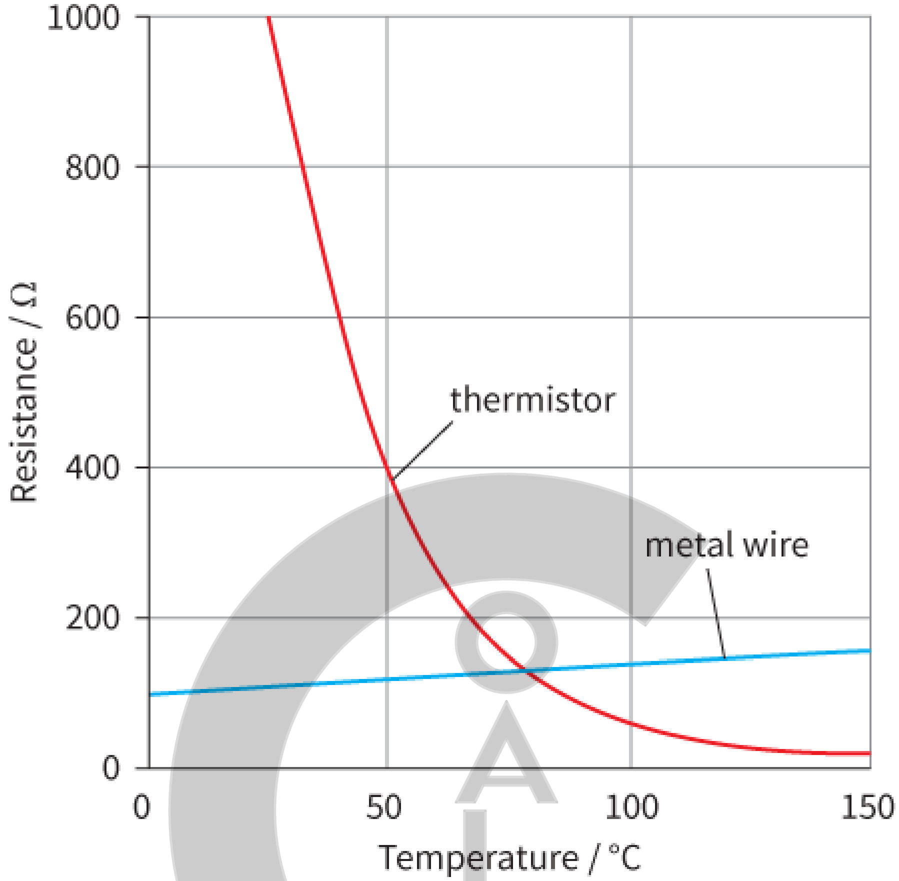


Figure 11.10: Variation of resistance with temperature.

Questions

10 Explain how a thermistor can be used as a transducer.

11 State **two** similarities between an LDR and a thermistor.

12 Design a circuit using the thermistor in Figure 11.10 that uses a cell of 10 V and produces an output voltage of 5 V at 50 °C. Explain whether the voltage output of your circuit increases or decreases as the temperature rises.

11.4 Potentiometer circuits

A **potentiometer** is a device used for comparing potential differences. For example, it can be used to measure the e.m.f. of a cell, provided you already have a source whose e.m.f. is known accurately. As we will see, a potentiometer can be thought of as a type of potential divider circuit.

A potentiometer consists of a piece of resistance wire, usually 1 m in length, stretched horizontally between two points. In Figure 11.11, the ends of the wire are labelled A and B. A **driver cell** is connected across the length of wire. Suppose this cell has an e.m.f. E_o of 2.0 V. We can then say that point A is at a voltage of 2.0 V, B is at 0 V, and the midpoint of the wire is at 1.0 V. In other words, the voltage decreases steadily along the length of the wire.

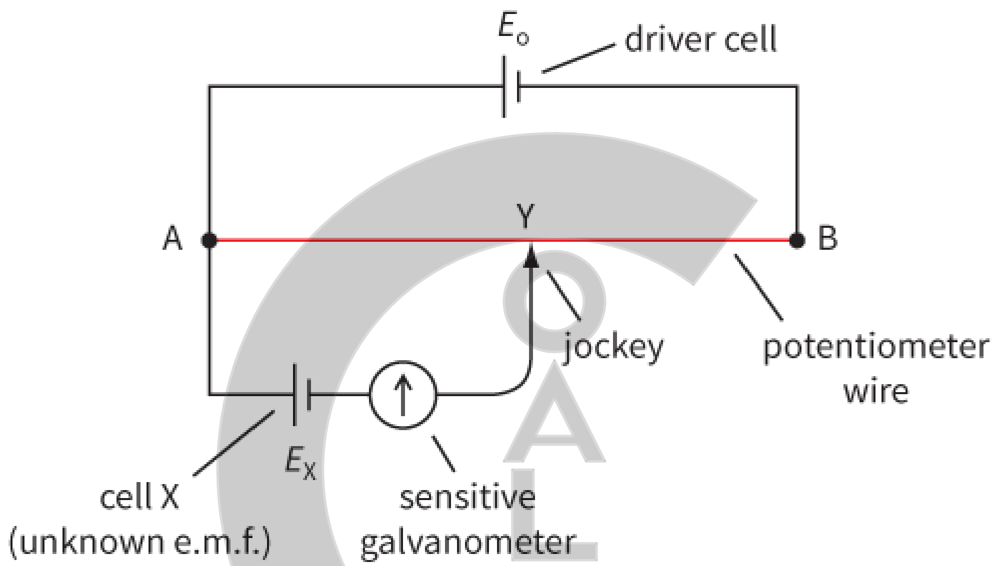


Figure 11.11: A potentiometer connected to measure the e.m.f. of cell X.

Now, suppose we wish to measure the e.m.f. E_X of cell X (this must have a value less than that of the driver cell). The positive terminal of cell X is connected to point A. (Note that both cells have their positive terminals connected to A.) A lead from the negative terminal is connected to a sensitive **galvanometer** (such as a microammeter), and a lead from the other terminal of the galvanometer ends with a metal **jockey**. This is a simple connecting device with a very sharp edge that allows very precise positioning on the wire.

If the jockey is touched onto the wire close to point A, the galvanometer needle will deflect in one direction. If the jockey is touched close to B, the galvanometer needle will deflect in the opposite direction. Clearly, there must be some point Y along the wire that, when touched by the jockey, gives zero deflection – the needle moves neither to the left nor the right.

In finding this position, the jockey must be touched gently and briefly onto the wire; the deflection of the galvanometer shows whether the jockey is too far to the left or right. It is important not to slide the jockey along the potentiometer wire as this may scrape its surface, making it non-uniform so that the voltage does not vary uniformly along its length.

When the jockey is positioned at Y, the galvanometer gives zero deflection, showing that there is no current through it. This can only happen if the potential difference across the length of wire AY is equal to the e.m.f. of cell X. We can say that the potentiometer is balanced. If the balance point was exactly half-way along the wire, we would be able to say that the e.m.f. of X was half that of the driver cell. This technique – finding a point where there is a reading of zero – is known as a **null method**.

To calculate the unknown e.m.f. E_X we measure the length AY. Then we have:

$$E_X = \frac{AY}{AB} \times E_o$$

where E_o is the e.m.f. of the driver cell.

KEY EQUATION

To compare two e.m.f.s E_x and E_o:

$$E_X = \frac{AY}{AB} \times E_o$$

The potentiometer can be thought of as a potential divider because the point of contact Y divides the resistance wire into two parts, equivalent to the two resistors of a potential divider.

Comparing e.m.f.s with a potentiometer

When a potentiometer is balanced, no current flows from the cell being investigated. This means that its terminal p.d. is equal to its e.m.f.; we do not have to worry about the potential difference across the internal resistance. This is a great advantage that a potentiometer has over a voltmeter, which must draw a small current in order to work.

However, there is a problem: the driver cell is supplying current to the potentiometer, and so the p.d. between A and B will be less than the e.m.f. of the driver cell (some volts are lost because of its internal resistance). To overcome this problem, we use the potentiometer to **compare** p.d.s. Suppose we have two cells whose e.m.f.s E_X and E_Y we want to compare. Each is connected in turn to the potentiometer, giving balance points at C and D–see Figure 11.12. (In the diagram, you can see immediately that E_Y must be greater than E_X because D is further to the right than C.)

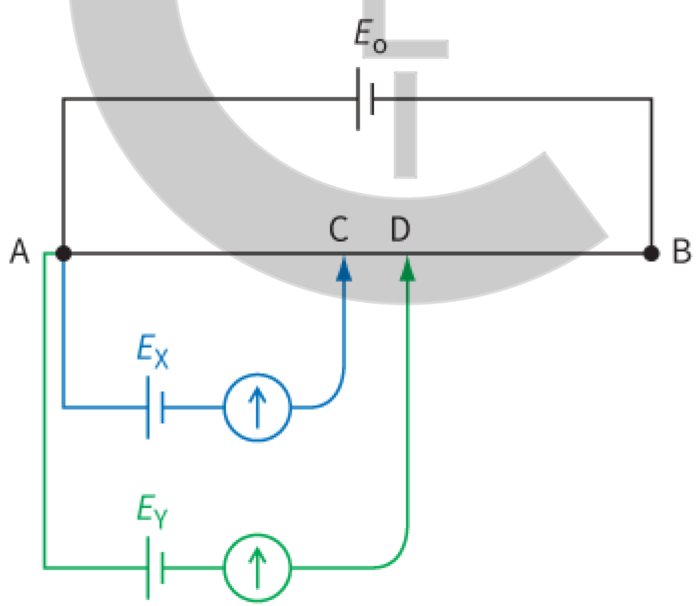


Figure 11.12: Comparing two e.m.f.s using a potentiometer.

The ratio of the e.m.f.s of the two cells will be equal to the ratio of the two lengths AC and AD:

$$\frac{E_X}{E_Y} = \frac{AC}{AD}$$

If one of the cells used has an accurately known e.m.f., the other can be calculated with the same degree of accuracy.

Comparing p.d.s

The same technique can be used to compare potential differences. For example, two resistors could be connected in series with a cell (Figure 11.13). The p.d. across one resistor is first connected to the potentiometer and the balance length found. This is repeated with the other resistor and the new balance point is found. The ratio of the lengths is the ratio of the p.d.s.

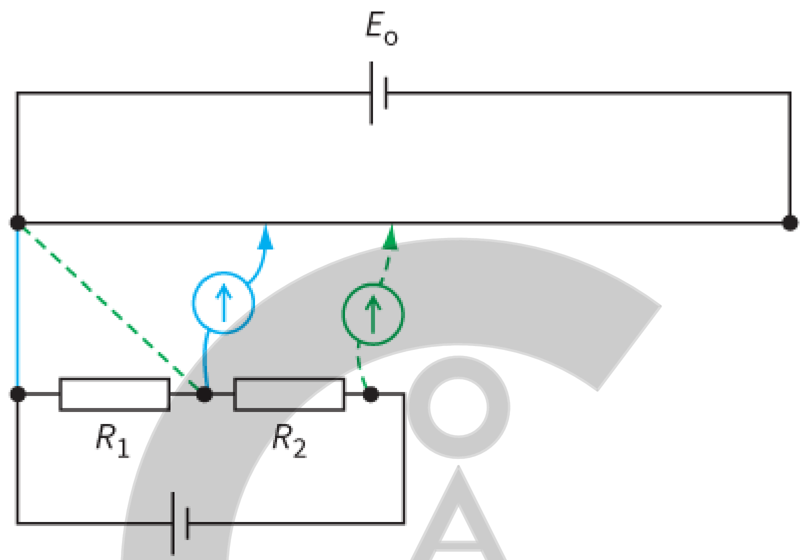


Figure 11.13: Comparing two potential differences using a potentiometer.

Since both resistors have the same current flowing through them, the ratio of the p.d.s is also the ratio of their resistances.

Question

13 To make a potentiometer, a driver cell of e.m.f. 4.0 V is connected across a 1.00 m length of resistance wire.

- **a** What is the potential difference across each 1 cm length of the wire? What length of wire has a p.d. of 1.0 V across it?
- **b** A cell of unknown e.m.f. *E* is connected to the potentiometer and the balance point is found at a distance of 37.0 cm from the end of the wire to which the galvanometer is connected. Estimate the value of *E*. Explain why this can only be an estimate.
- **c** A standard cell of e.m.f. 1.230 V gives a balance length of 31.2 cm. Use this value to obtain a more accurate value for *E*.

REFLECTION

A student sets up a potentiometer circuit to compare the e.m.f.s of two cells. The student is unable to find a balance point.

Discuss with a partner possible reasons for this. Consider using Kirchhoff's Laws as a way of exploring the reasons.

Did you do your work the way other people did theirs? In what ways did you do it differently? In what ways was your work or process similar?

COALI

SUMMARY

A source of e.m.f., such as a battery, has an internal resistance. We can think of the source as having an internal resistance, *r,* in series with an e.m.f., *E*.

The terminal p.d. of a source of e.m.f. is less than the e.m.f. because of the potential difference across the internal resistor:

$$\text{terminal p.d.} = \text{e.m.f.} - \text{p.d across the internal resistor}$$

$$V = E - Ir$$

A potential divider circuit consists of two or more resistors connected in series to a supply. The output voltage V_{out} across the resistor of resistance R_2 is given by:

$$V_{out} = \left(\frac{R_2}{R_1+R_2}\right) \times V_{in}$$

A potentiometer can be used to compare potential differences.

EXAM-STYLE QUESTIONS

1 A resistor of resistance 6.0 Ω and a second resistor of resistance 3.0 Ω are connected in parallel across a battery of e.m.f. 4.5 V and internal resistance 0.50 Ω.

What is the current in the battery? **[1]**

A 0.47 A

B 1.8 A

C 3.0 A

D 11 A

2 This diagram shows a potential divider.

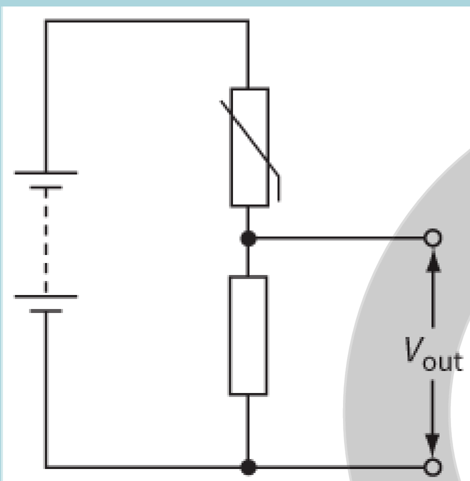


Figure 11.14

What happens when the temperature decreases? **[1]**

A The resistance of the thermistor decreases and V_{out} decreases.

B The resistance of the thermistor decreases and V_{out} increases.

C The resistance of the thermistor increases and V_{out} decreases.

D The resistance of the thermistor increases and V_{out} increases.

3 A single cell of e.m.f. 1.5 V is connected across a 0.30 Ω resistor. The current in the circuit is 2.5 A.

a Calculate the terminal p.d. and explain why it is not equal to the e.m.f. of the cell. **[2]**

b Show that the internal resistance *r* of the cell is 0.30 Ω. **[3]**

c It is suggested that the power dissipated in the external resistor is a maximum when its resistance *R* is equal to the internal resistance *r* of the cell.

i Calculate the power dissipated when $R = r$. **[1]**

ii Show that the power dissipated when $R = 0.50\ \Omega$ and $R = 0.20\ \Omega$ is less than that dissipated when $R = r$, as the statement suggests. **[4]**

[Total: 10]

4 A student is asked to compare the e.m.f.s of a standard cell and a test cell. He sets up the circuit shown using the test cell.

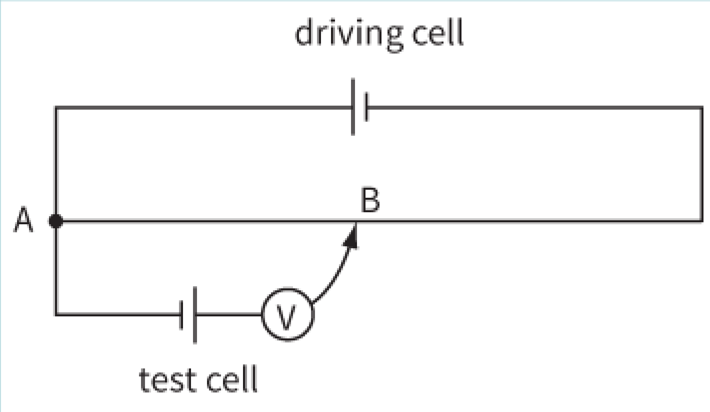


Figure 11.15

a **i** Explain why he is unable to find a balance point and state the change he must make in order to achieve balance. **[2]**

ii State how he would recognise the balance point. **[1]**

b He achieves balance when the distance AB is 22.5 cm. He repeats the experiment with a standard cell of e.m.f. of 1.434 V. The balance point using this cell is at 34.6 cm. Calculate the e.m.f. of the test cell. **[2]**

[Total: 5]

5 **a** Explain what is meant by the **internal resistance** of a cell. **[2]**

b When a cell is connected in series with a resistor of 2.00 Ω there is a current of 0.625 A. If a second resistor of 2.00 Ω is put in series with the first, the current falls to 0.341 A.

Calculate:

i the internal resistance of the cell **[2]**

ii the e.m.f. of the cell. **[1]**

c A car battery needs to supply a current of 200 A to turn over the starter motor. Explain why a battery made of a series of cells of the type described **b** would not be suitable for a car battery. **[2]**

[Total: 7]

6 **a** State what is meant by the term **e.m.f. of a cell**. **[2]**

A student connects a high-resistance voltmeter across the terminals of a battery and observes a reading of 8.94 V. He then connects a 12 Ω resistor across the terminals and finds that the potential difference falls to 8.40 V.

b Explain why the measured voltage falls. **[2]**

c **i** Calculate the current in the circuit. **[2]**

ii Calculate the internal resistance of the cell. **[2]**

iii State any assumptions you made in your calculations. **[1]**

[Total: 9]

7 This diagram shows two circuits that could be used to act as a dimmer switch for a lamp.

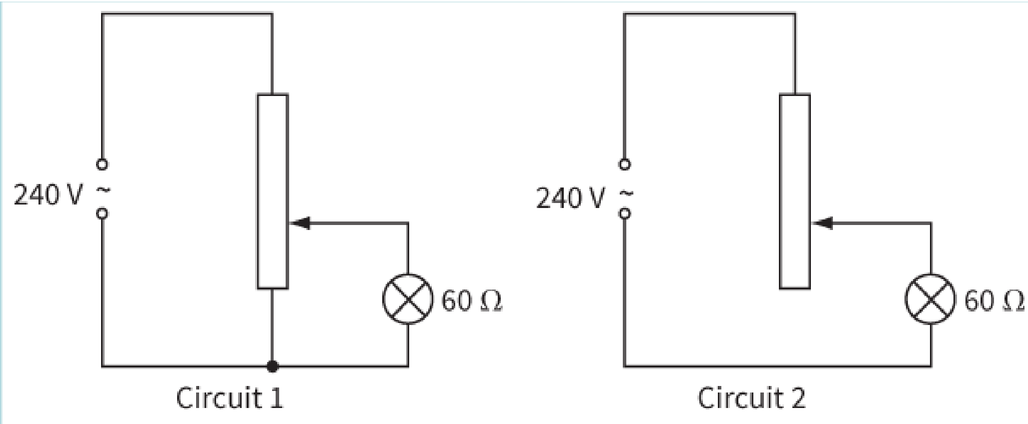


Figure 11.16

a Explain **one** advantage circuit 1 has over circuit 2. **[2]**

b **i** The lamp is rated at 60 W at 240 V. Calculate the resistance of the lamp filament at its normal operating temperature. **[2]**

ii State and explain how the resistance of the filament at room temperature would compare with the value calculated in **i**. **[2]**

[Total: 6]

8 This circuit shows a potential divider. The battery has negligible internal resistance and the voltmeter has infinite resistance.

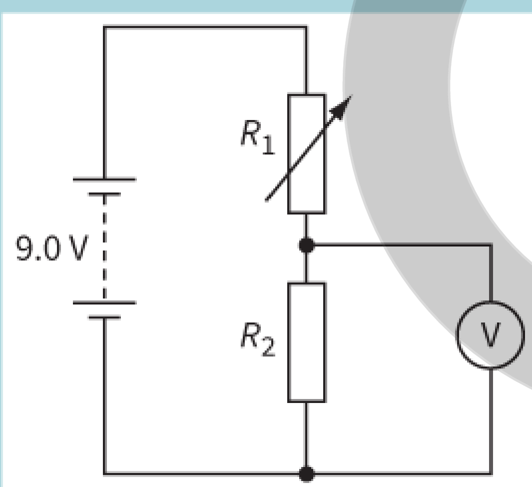


Figure 11.17

a State and explain how the reading on the voltmeter will change when the resistance of the variable resistor is increased. **[2]**

b Resistor R_2 has a resistance of 470 Ω. Calculate the value of the variable resistor when the reading on the voltmeter is 2.0 V. **[2]**

c The voltmeter is now replaced with one of resistance 2 kΩ. Calculate the reading on this voltmeter. **[2]**

[Total: 6]

9 This is a potentiometer circuit.

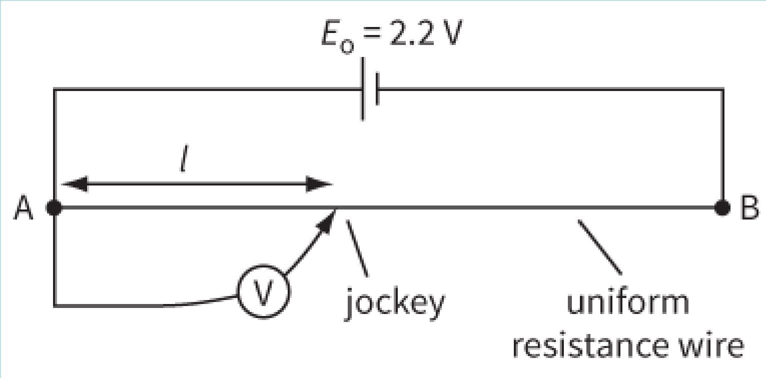


Figure 11.18

a **i** Sketch a graph of reading on the voltmeter against length, *l*, as the jockey is moved from point A to point B. **[2]**

ii State the readings on the voltmeter when the jockey is connected to A and when it is connected to B. (You may assume that the driver cell has negligible internal resistance.) **[1]**

iii Draw a circuit diagram to show how the potentiometer could be used to compare the e.m.f.s of two batteries. **[3]**

b When a pair of 4 Ω resistors are connected in series with a battery, there is a current of 0.60 A current through the battery. When the same two resistors are connected in parallel and then connected across the battery, there is a current of 1.50 A through it. Calculate the e.m.f. and the internal resistance of the battery. **[4]**

[Total: 10]

10 A potentiometer, which consists of a driving cell connected to a resistance wire of length 100 cm, is used to compare the resistances of two resistors.

a Draw a diagram to show the circuits that are used to compare the two resistances. **[2]**

b When resistor R_1 alone is tested the length of resistance wire for balance is 15.4 cm. There is an uncertainty in measuring the beginning of the resistance wire of 0.1 cm, and in establishing the balance point of a further 0.1 cm.

i Determine the uncertainty in the balance length. **[1]**

When R_1 and R_2 are tested in series the balance length is 42.6 cm. There are similar uncertainties in measuring this balance length.

ii Calculate the ratio of $\frac{R_1}{(R_1+R_2)}$ **[1]**

iii Calculate the value of the ration of $\frac{R_1}{R_2}$. **[2]**

iv Calculate the uncertainty in the value of the ratio $\frac{R_1}{R_2}$. **[2]**

[Total: 8]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the concept of internal resistance of a source of e.m.f.	11.1			
solve problems involving internal resistance and e.m.f. and the potential difference across the internal resistance.	11.1			
recognise a potential divider and solve problems using the equation: $V_{out} = \left(\frac{R_2}{R_1+R_2}\right) \times V_{in}$	11.2, 11.3			
use a potentiometer to compare potential differences.	11.4			


OALI

〉Chapter 12

Waves

LEARNING INTENTIONS

In this chapter you will learn how to:

- describe a progressive wave
- describe the motion of transverse and longitudinal waves
- describe waves in terms of their wavelength, amplitude, frequency, speed, phase difference and intensity
- use the time-base and *y*-gain of a cathode-ray oscilloscope (CRO) to determine frequency and amplitude
- use the wave equation $v = f\lambda$
- use the equations $\text{intensity} = \frac{\text{power}}{\text{area}}$ and $\text{intensity} \propto \text{amplitude}^2$
- describe the Doppler effect for sound waves
- use the equation $f_0 = \frac{f_s v}{(v \pm v_s)}$
- describe and understand electromagnetic waves
- recall that wavelengths in the range 400–700 nm in free space are visible to the human eye
- describe and understand polarisation
- use Malus's law to determine the intensity of transmitted light through a polarising filter.

BEFORE YOU START

- Write down definitions for displacement, speed and power.
- What do you know about the electromagnetic spectrum? Can you name any of the waves in this spectrum? Make a list to share with the class.

VIBRATIONS MAKING WAVES

The wind blowing across the surface of the sea produces waves. The surface of the water starts to move up and down, and these vibrations spread outwards – big waves may travel thousands of kilometres across the ocean before they break on a beach (Figure 12.1).

How can you tell from looking at Figure 12.1 that a wave is a form of energy?

Figure 12.1: This photograph shows a wave breaking on the shore and dissipating the energy it has drawn from the wind in its journey across the ocean. The two scientists are 'storm chasers' who are recording the waves produced by a hurricane in the Gulf of Mexico.

12.1 Describing waves

When you pluck the string of a guitar, it vibrates. The vibrations create a wave in the air that we call sound. In fact, all vibrations produce **waves** of one type or another (Figure 12.2). Waves that move through a material (or a vacuum) are called **progressive waves**. A progressive wave transfers energy from one position **to** another.

At the seaside, a wave is what we see on the surface of the sea. The water moves around and a wave travels across the surface. In physics, we extend the idea of a wave to describe many other phenomena, including light, sound and so on. We do this by imagining an ideal wave, as shown in Figure 12.3 – you will never see such a perfect wave on the sea!

Figure 12.2: Radio telescopes detect radio waves from distant stars and galaxies; a rainbow is an effect caused by the reflection and refraction of light waves by water droplets in the atmosphere.

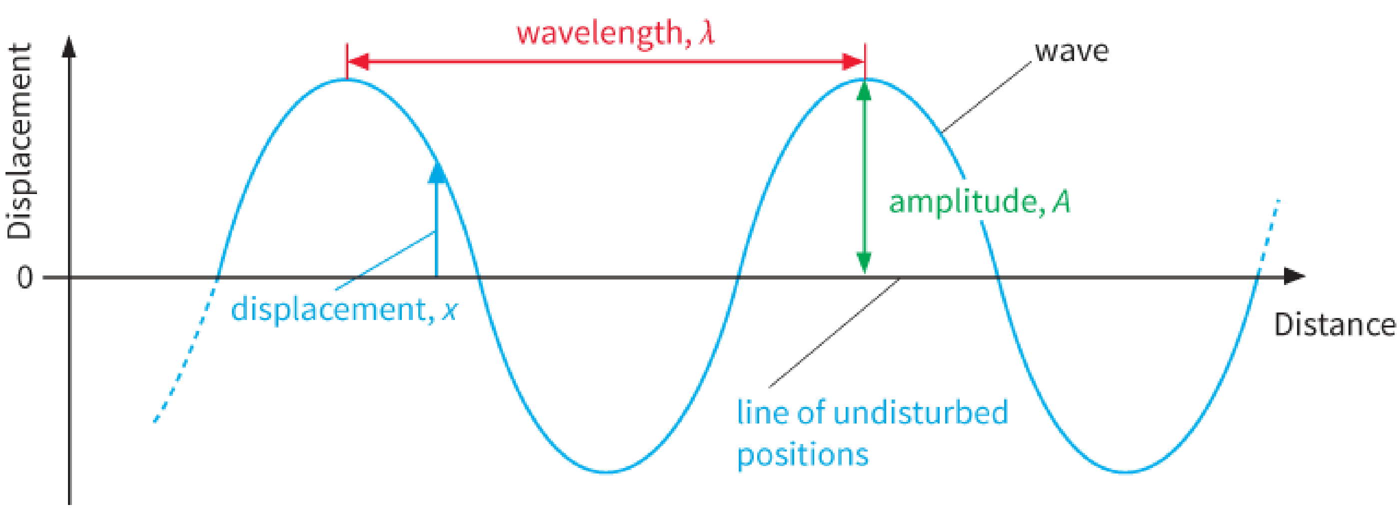


Figure 12.3: A displacement–distance graph illustrating the terms displacement, amplitude and wavelength.

Figure 12.3, or a similar graph of displacement against time, illustrates the following important definitions about waves and wave motion.

- The distance of a point on the wave from its undisturbed position, or equilibrium position, is called the **displacement** x.
- The maximum displacement of any point on the wave from its undisturbed position is called the **amplitude** A. The amplitude of a wave on the sea is measured in units of distance, such as metres.
- The greater the amplitude of the wave, the louder the sound or the rougher the sea.
- The distance between two adjacent points on a wave oscillating in step with each other is called the **wavelength** λ (the Greek letter lambda). This is the same as the distance between two adjacent peaks or troughs. The wavelength of a wave on the sea is measured in units of distance, such as metres.
- The time taken for one complete oscillation of a point in a wave is called the **period** T. It is the time taken for a point to move from one particular position and return to that same position, moving in the same direction. It is measured in units of time, such as seconds.
- The number of oscillations per unit time of a point in a wave is called its **frequency** f. For sound waves, the higher the frequency of a musical note, the higher is its pitch. Frequency is measured in hertz (Hz), where 1 Hz = 1 oscillation per second (1 kHz = 10^3 Hz and 1 MHz = 10^6 Hz).

  The frequency f of a wave is the reciprocal of the period T:

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

Waves are called **mechanical waves** if they need a substance (medium) through which to travel. Sound is one example of such a wave. Other cases are waves on stretched strings, seismic waves and water waves (Figure 12.4).

Figure 12.4: The impact of a droplet on the surface of a liquid creates a vibration, which in turn gives rise to waves on the surface.

PRACTICAL ACTIVITY 12.1

Measuring frequency

You can measure the frequency of sound waves using a cathode-ray oscilloscope (CRO) or oscilloscope for short. Figure 12.6 shows how.

A microphone is connected to the input of the CRO. The microphone converts the sound waves into a varying voltage that has the same frequency as the sound waves. This voltage is displayed on the CRO screen.

It is best to think of a CRO as a voltmeter that is capable of displaying a rapidly varying voltage. To do this, its spot moves across the screen at a steady speed, set by the time-base control. At the same time, the spot moves up and down according to the voltage of the input.

Hence, the display on the screen is a graph of the varying voltage on the (vertical) *y*-axis, with time on the (horizontal) *x*-axis. If we know the horizontal scale, we can determine the period and hence the frequency of the sound wave. Worked example **1** shows how to do this. (In Chapter 14 we will look at one method of measuring the wavelength of sound waves.)

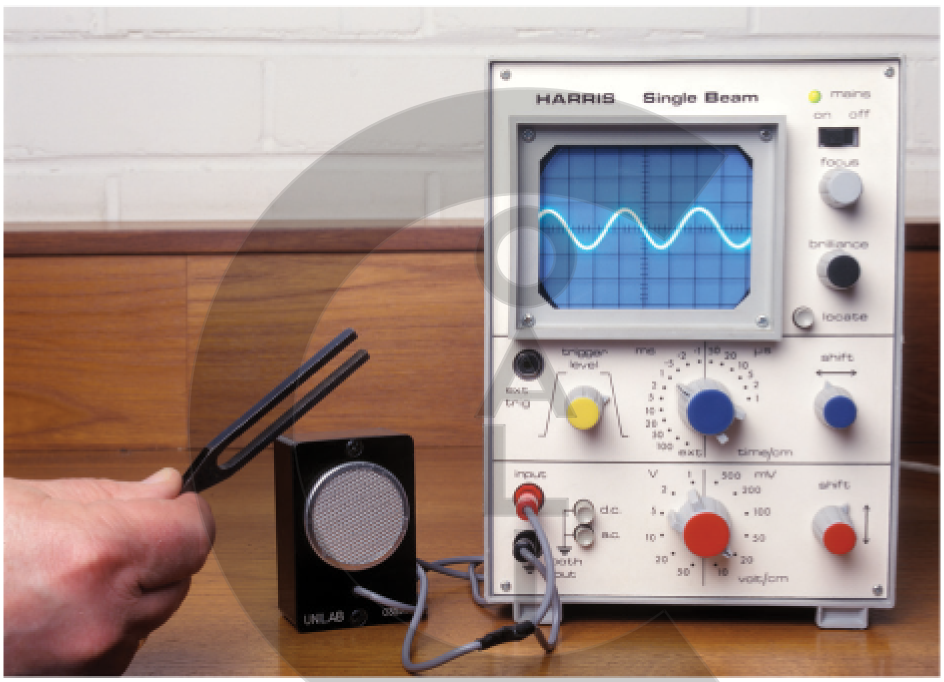


Figure 12.6: Determining the frequency of sound waves from a vibrating tuning fork.

WORKED EXAMPLE

1 Figure 12.7 shows the trace on an oscilloscope screen when sound waves are detected by a microphone. The time-base is set at 1 ms div^{-1}. The *y*-gain is set to 20 mV div^{-1}.

Determine the frequency of the sound waves and the amplitude of the oscilloscope trace.

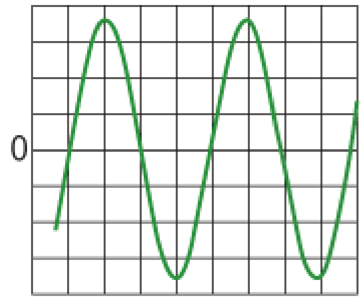


Figure 12.7: A CRO trace – what is the frequency of the sound waves detected by the microphone and the amplitude of the trace?

Step 1 Determine the period of the trace on the screen, in scale divisions. From Figure 12.7, you can see that the period is equivalent to 4.0 scale divisions (div).

period T = 4.0 div

Step 2 Determine the period in seconds (s) using the time-base setting.

period T = 4.0 div × time-base setting = 4.0 div × 1 ms div^{-1} = 4.0 ms

Hint: *Notice how div and div^{-1} cancel out.*

1 ms = 10^{-3} s

Therefore, period T = 4.0 × 10^{-3} s

Step 3 Calculate the frequency f from the period T:

$$f = \frac{1}{T} = \frac{1}{4.0 \times 10^{-3}} = 250 \text{ Hz}$$

So, the sound wave frequency is 250 Hz.

Step 4 Determine the amplitude of the trace on the screen, in scale divisions. From Figure 12.7, you can see that the amplitude is equivalent to 3.5 scale divisions (div). Remember that the amplitude is measured from the 0 volt position.

amplitude of trace = 3.5 div

Step 5 Determine the amplitude in volts (V) using the y-gain setting.

amplitude = 3.5 div × y-setting = 3.5 div × 20 mV div^{-1} = 70 mV

Hint: *Notice how div and div^{-1} cancel out again.*

1 mV = 10^{-3} V

Therefore, amplitude = 70 × 10^{-3} V = 0.070 V

Questions

1 Determine the wavelength and amplitude of each of the two waves shown in Figure 12.5.

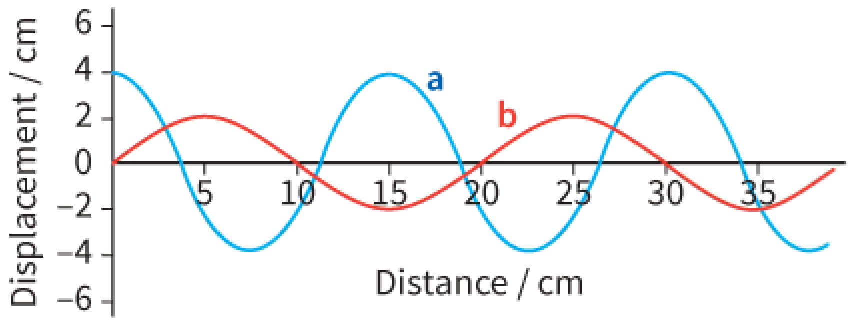


Figure 12.5: Two waves for Question 1.

2 A microphone detects sound waves. The microphone is connected to a CRO. On the CRO screen, two complete cycles occupy five scale divisions along the *x*-axis. The calibrated time-base is set on 0.005 s div^{-1}.

Determine the frequency of the sound waves.

12.2 Longitudinal and transverse waves

There are two distinct types of wave, **longitudinal** and **transverse**. Both can be demonstrated using a toy spring lying along a bench.

Push the end of the spring back and forth; the segments of the spring become compressed and then stretched out, along the length of the spring. Wave pulses run along the spring. These are **longitudinal waves**.

Waggle the end of the spring from side to side. The segments of the spring move from side to side as the wave travels along the spring. These are **transverse waves**.

So, the distinction between longitudinal and transverse waves is as follows.

- In longitudinal waves, the particles of the medium vibrate **parallel** to the direction of the wave velocity.
- In transverse waves, the particles of the medium vibrate at **right angles** to the direction of the wave velocity.

Sound waves are an example of a longitudinal wave. Light and all other electromagnetic waves are transverse waves. Waves in water are quite complex. Particles of the water may move both up and down and from side to side as a water wave travels through the water. You can investigate water waves in a ripple tank. There is more about water waves in Table 12.1 and in Chapter 13.

Representing waves

Figure 12.8 shows how we can represent longitudinal and transverse waves. The longitudinal wave shows how the material through which it is travelling is alternately compressed and expanded. This gives rise to high and low pressure regions, respectively.

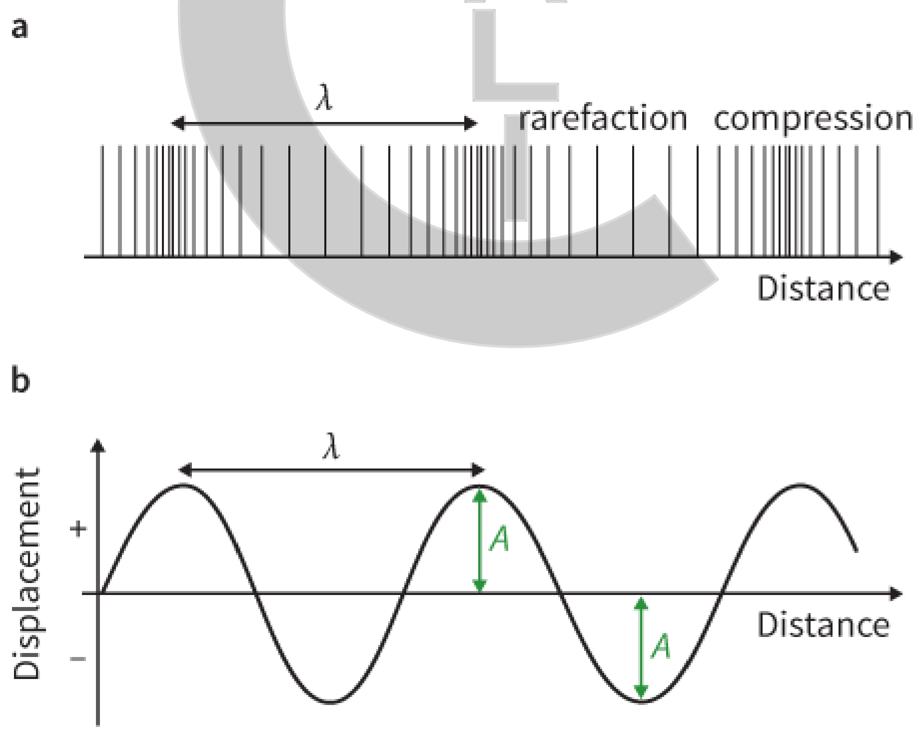


Figure 12.8: **a** Longitudinal waves and **b** transverse waves. *A* = amplitude, λ = wavelength.

However, this can be difficult to draw, so you will often see a longitudinal wave represented as if it were a sine wave. The displacement referred to in the graph is the displacement of the particles in the wave.

We can compare the **compressions** and **rarefactions** (or expansions) of the longitudinal wave with the peaks and troughs of the transverse wave.

Phase and phase difference

All points along a wave have the same pattern of vibration. However, different points do not necessarily vibrate in step with one another. As one point on a wave vibrates, the point next to it vibrates slightly out-of-step with it. We say that they vibrate out of phase with each other – there is a **phase difference** between them. This is the amount by which one oscillation leads or lags behind another.

Two particles oscillating in step have a phase difference of 0°, 360° and so on (or 0 rad, 2π rad and so on).

Two particles oscillating in antiphase have a phase difference of 180°, 270° and so on (or π rad, 3π rad and so on).

Phase difference is measured in degrees or in radians. As you can see from Figure 12.9, two points A and B, with a separation of one whole wavelength λ, vibrate in phase with each other. The phase difference between these two oscillating particles at A and B is 360°. (You can also say it is 0°.) The phase difference between any other two points between A and B can have any value between 0° and 360°. A complete cycle of the wave is thought of as 360°. The separation between points C and D is quarter of a wavelength – the phase difference between these two points is 90°. In general, when the separation between two oscillating particles on a wave is x, then the phase difference ϕ between these particles in degrees can be calculated using the expression:

$$\phi = \frac{x}{\lambda} \times 360°$$

where λ is the wavelength of the wave.

The idea of phase difference is revisited in Chapter 13.

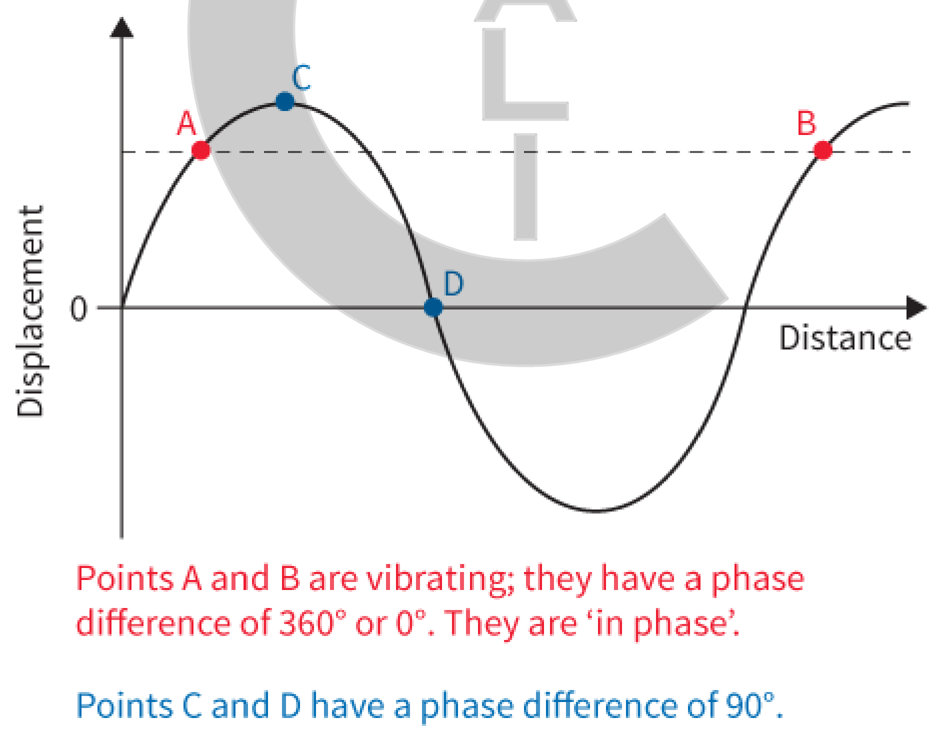


Figure 12.9: Different points along a wave have different phases.

Question

3 Using axes of displacement and distance, sketch two waves A and B such that A has twice the wavelength and half the amplitude of B.

12.3 Wave energy

It is important to realise that, for both types of mechanical wave, the particles that make up the material through which the wave is travelling do not move along – they only oscillate about a fixed point. It is **energy** that is transmitted by the wave. Each particle vibrates; as it does so, it pushes its neighbour, transferring energy to it. Then that particle pushes its neighbour, which pushes its neighbour. In this way, energy is transmitted from one particle to the next, to the next and so on down the line.

Intensity

The term **intensity** has a very precise meaning in physics. The intensity of a wave is defined as the rate of energy transmitted (power) per unit area at right angles to the wave velocity.

$$\text{intensity} = \frac{\text{power}}{\text{area}}$$

Intensity is measured in watts per square metre (W m^{-2}). For example, when the Sun is directly overhead, the intensity of its radiation is about 1.0 kW m^{-2} (1 kilowatt per square metre). This means that energy arrives at the rate of about 1 kW (1000 J s^{-1}) on each square metre of the surface of the Earth. At the top of the atmosphere, the intensity of sunlight is greater, about 1.4 kW m^{-2}.

KEY EQUATION

$$\text{intensity} = \frac{\text{power}}{\text{area}}$$

Question

4 A 100 W lamp emits electromagnetic radiation in all directions. Assuming the lamp to be a point source, calculate the intensity of the radiation:

  a at a distance of 1.0 m from the lamp

  b at a distance of 2.0 m from the lamp.

  Hint: *Think of the area of a sphere at each of the two radii*.

Intensity and amplitude

The intensity of a wave generally decreases as it travels along. There are two reasons for this:

- The wave may ‘spread out’ (as in the example of light spreading out from a lamp in Question 4).
- The wave may be absorbed or scattered (as when light passes through the Earth’s atmosphere).

As a wave spreads out, its amplitude decreases. This suggests that the intensity *I* of a wave is related to its amplitude *A*.

In fact, intensity *I* is directly proportional to the square of the amplitude *A*:

$$\text{intensity} \propto \text{amplitude}^2 \text{ or } I \propto A^2$$

KEY EQUATION

$$\text{intensity} \propto \text{amplitude}^2 \text{ or } I \propto A^2$$

The relationship also implies that, for a particular wave:

$$\frac{\text{intensity}}{\text{amplitude}^2} = \text{constant}$$

So, if one wave has **twice** the amplitude of another, it has **four** times the intensity. This means that the wave is transmitting four times the power per unit area at right angles to the wave velocity.

Question

5 A wave from a source has an amplitude of 5.0 cm and an intensity of 400 W m^{-2}.

a The amplitude of the wave is increased to 10.0 cm. Calculate the intensity now.

b The intensity of the wave is decreased to 100 W m^{-2}. Calculate the amplitude now.

12.4 Wave speed

The speed with which energy is transmitted by a wave is known as the wave speed v. This is measured in m s^{-1}. The wave speed for sound in air at atmospheric pressure of 10^5 Pa and a temperature of 0 °C is about 330 m s^{-1}, while for light in a vacuum it is almost 300 000 000 m s^{-1}.

The wave equation

An important equation connecting the speed v of a wave with its frequency, f and wavelength, λ can be determined as follows. We can find the speed of the wave using:

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

A wave will travel a distance of one whole wavelength, λ in a time equal to one period, T. So:

$$\text{wave speed} = \frac{\text{wavelength}}{\text{period}} \text{ or } v = \frac{\lambda}{T}$$

$$v = \frac{1}{T} \times \lambda$$

However, $f = \frac{1}{T}$ and so:

$$v = f \times \lambda$$

where v is the speed of the wave, f is the frequency and λ is the wavelength.

KEY EQUATION

$$v = f\lambda$$

where v is the speed of the wave, f is the frequency and λ is the wavelength.

A numerical example may help to make this clear. Imagine a wave of frequency 5 Hz and wavelength 3 m going past you. In 1 s, five complete wave cycles, each of length 3 m, go past. So the total length of the waves going past in 1 s is 15 m. The distance travelled by the wave per second is its speed, therefore the speed of the wave is 15 m s^{-1}.

You can see that, for a given speed of wave, the greater the wavelength, the smaller the frequency (and the smaller the wavelength, the greater the frequency). This means, that for a constant wave speed, the wavelength is inversely proportional to the frequency. The speed of sound in air is constant (for a given temperature and pressure). The wavelength of sound can be made smaller by increasing the frequency of the source of sound.

Table 12.1 gives typical values of speed v, frequency f and wavelength λ for some mechanical waves. You can check for yourself that $v = f\lambda$ is valid.

	Water waves in a ripple tank	Sound waves in air	Waves on a toy spring
Speed v / m s^{-1}	about 0.12	330	about 1
Frequency f / Hz	about 6	20 to 20 000 (limits of human hearing)	about 2

	Water waves in a ripple tank	Sound waves in air	Waves on a toy spring
Wavelength λ / m	about 0.2	16.5 to 0.0165	about 0.5

Table 12.1: Data for some mechanical waves that are often investigated in the laboratory.

WORKED EXAMPLE

2 Middle C on a piano tuned to concert pitch should have a frequency of 264 Hz (Figure 12.10). If the speed of sound is 330 m s^{-1}, calculate the wavelength of the sound produced from this note.

Step 1 We rearrange the wave equation to the form:

$$\lambda = \frac{v}{f}$$

Step 2 Substituting the values we get:

$$\lambda = \frac{330}{264} = 1.25 \text{ m}$$

The wavelength, λ is 1.25 m.

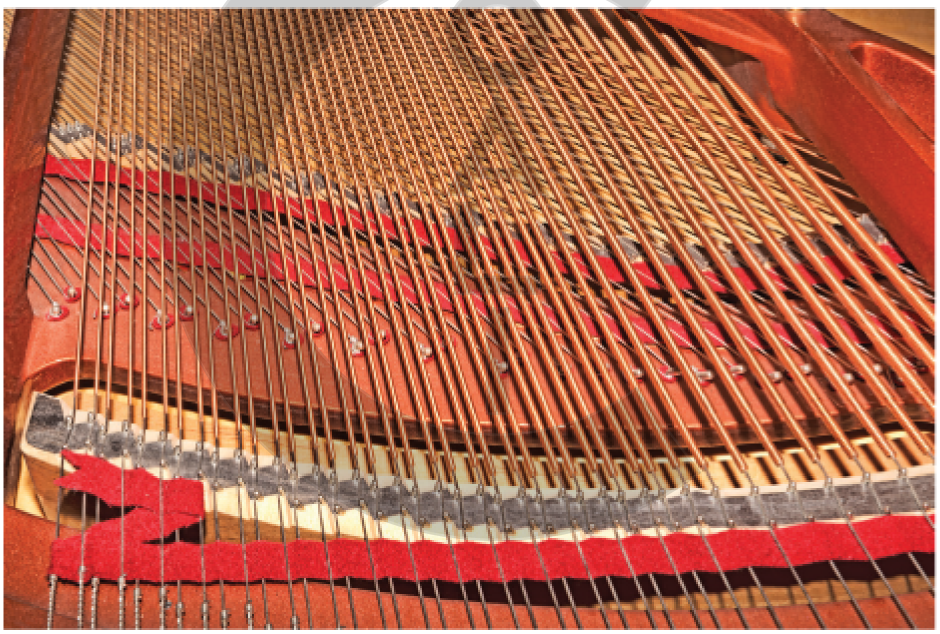


Figure 12.10: Each string in a piano produces a different note.

Questions

6 Sound is a mechanical wave that can be transmitted through a solid.

Calculate the frequency of sound of wavelength 0.25 m that travels through steel at a speed of 5060 m s^{-1}.

7 A cello string vibrates with a frequency of 64 Hz.

Calculate the speed of the transverse waves on the cello string given that the wavelength is 140 cm.

8 An oscillator is used to send a transverse wave along a stretched string. The wavelength of the wave is 5.0 cm when the frequency of the oscillator is 30 Hz.

For this wave, calculate:

a its frequency

b its speed.

9 Copy and complete Table 12.2. (You may assume that the speed of radio waves is 3.00×10^8 m s^{-1}.)

Station	**Wavelength / m**	**Frequency / MHz**
Radio A (FM)		97.6
Radio B (FM)		94.6
Radio B (LW)	1515	
Radio C (MW)	693	

Table 12.2: For Question 9.

12.5 The Doppler effect for sound waves

You may have noticed a change in pitch of the note heard when an emergency vehicle passes you while sounding its siren. The pitch is higher as the vehicle approaches you, and lower as it moves away (recedes). This is an example of the **Doppler effect**; you can hear the same thing if a train passes at speed while sounding its whistle.

Figure 12.11 shows why this change in frequency is observed. It shows a source of sound emitting waves with a constant frequency f_s, together with two observers A and B.

- If the source is stationary (Figure 12.11a), waves arrive at A and B at the same rate, and so both observers hear sounds of the same frequency f_s.
- If the source is moving towards A and away from B (Figure 12.11b), the situation is different. From the diagram, you can see that the waves are squashed together in the direction of A and spread apart in the direction of B.

Observer A will observe, or detect, waves whose wavelength is shortened. More wavelengths per second arrive at A, and so A observes a sound of higher frequency than f_s. Similarly, the waves arriving at B have been stretched out and B will observe a frequency lower than f_s.

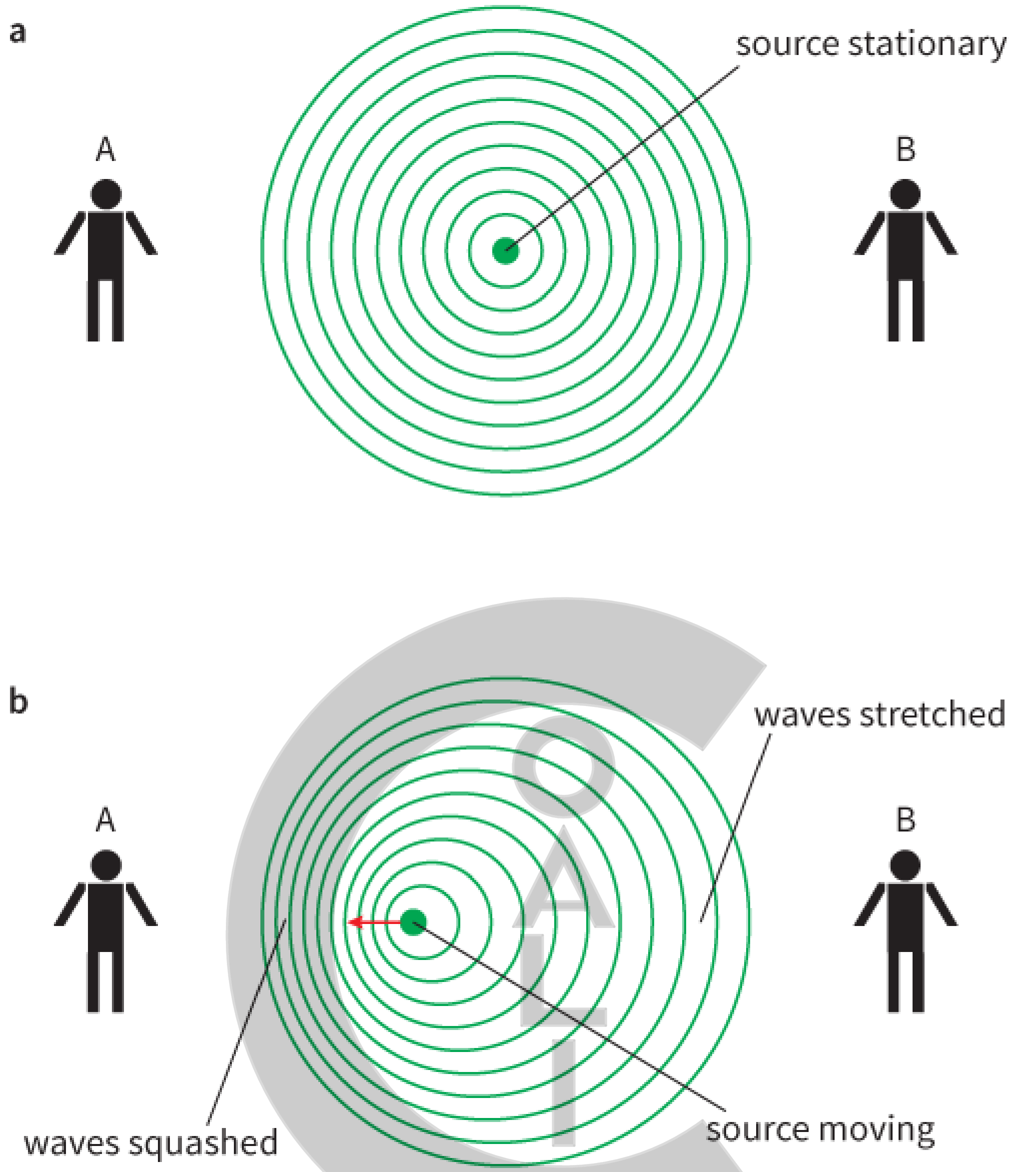


Figure 12.11: Sound waves (green lines) emitted at constant frequency by **a** a stationary source, and **b** a source moving with speed v_s. The separation between adjacent green lines is equal to one wavelength.

An equation for observed frequency

There are two different speeds involved in this situation. The source is moving with speed v_s. The sound waves travel through the air with speed v, which is unaffected by the speed of the source. (Remember, the speed of a wave depends only on the medium it is travelling through.)

The frequency and wavelength observed by an observer will change according to the speed v_s at which the source is moving relative to the stationary observer. Figure 12.12 shows how we can calculate the observed wavelength λ_0 and the observed frequency f_0.

The wave sections shown in Figure 12.12 represent the f_s wavelengths emitted by the source in 1 s. Provided the source is stationary (Figure 12.12a), the length of this section is equal to the wave speed v. The wavelength observed by the observer is simply:

$$\lambda_0 = \frac{v}{f_s}$$

The situation is different when the source is moving away (receding) from the observer (Figure 12.12b). In 1 s, the source moves a distance v_s. Now the section of f_s wavelengths will have a length equal to $v + v_s$.

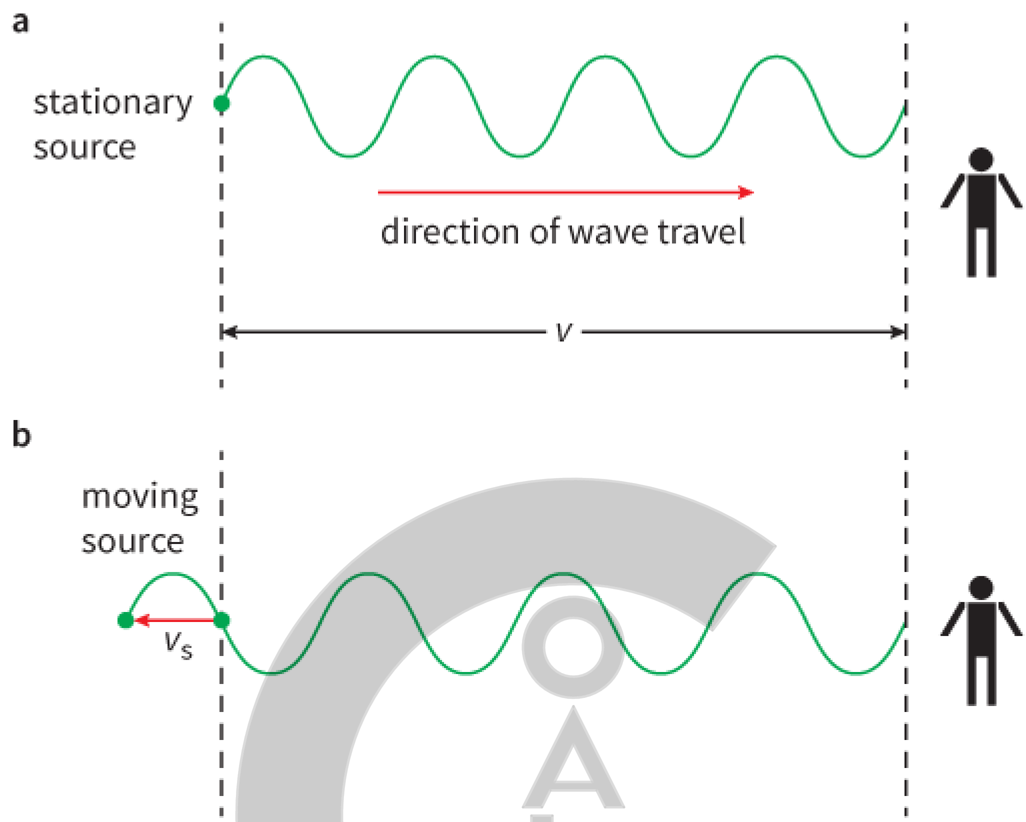


Figure 12.12: Sound waves, emitted at constant frequency by **a** a stationary source, and **b** a source moving with speed v_s *away* from the observer (that is, the person hearing the sound).

The observed wavelength is now given by:

$$\lambda_0 = \frac{(v+v_s)}{f_s}$$

The observed frequency is given by:

$$f_0 = \frac{v}{\lambda_0} = \frac{f_s \times v}{(v+v_s)}$$

where f_0 is the observed frequency, f_s is the frequency of the source, v is the speed of the wave and v_s is the speed of the source relative to the observer.

This shows us how to calculate the observed frequency when the source is moving away from the observer. If the source is moving towards the observer, the section of f_s wavelengths will be compressed into a shorter length equal to $v - v_s$, and the observed frequency will be given by:

$$f_0 = \frac{v}{\lambda_0} = \frac{f_s \times v}{(v-v_s)}$$

We can combine these two equations to give a single equation for the Doppler shift in frequency due to a moving source:

$$\text{observed frequency, } f_0 = \frac{f_s \times v}{(v \pm v_s)}$$

KEY EQUATION

Doppler effect:

$$f_0 = \frac{f_s \times v}{(v \pm v_s)}$$

where the plus sign applies to a receding source and the minus sign to an approaching source. Note these important points:

- The frequency f_s of the source is not affected by the movement of the source.
- The speed v of the waves as they travel through the air (or other medium) is also unaffected by the movement of the source.

Note that a Doppler effect can also be heard when an observer is moving relative to a stationary source, and also when both source and observer are moving. There is more about the Doppler effect and light in Chapter 31.

WORKED EXAMPLE

3 A train with a whistle that emits a note of frequency 800 Hz is approaching a stationary observer at a speed of 60 m s^{-1}.

Calculate the frequency of the note heard by the observer.

speed of sound in air = 330 m s^{-1}

Step 1 Select the appropriate form of the Doppler equation. Here the source is approaching the observer so we choose the minus sign:

$$f_0 = \frac{f_s \times v}{(v - v_s)}$$

Step 2 Substitute values from the question and solve:

$$f_0 = \frac{800 \times 330}{(330 - 60)}$$

$$= \frac{800 \times 330}{270}$$

$$= 978 \text{ Hz} \approx 980 \text{ Hz}$$

So, the observer hears a note whose pitch is raised significantly, because the train is travelling at a speed that is a significant fraction of the speed of sound.

Question

10 A plane's engine emits a note of constant frequency 120 Hz. It is flying away from a stationary observer at a speed of 80 m s^{-1}. Calculate:

a the observed wavelength of the sound received by the observer

b its observed frequency.

(Speed of sound in air = 330 m s^{-1}.)

12.6 Electromagnetic waves

You will have learnt that light is a region of the **electromagnetic spectrum**. You might not think that light has any connection at all with electricity, magnetism and waves – but it does. Physicists studied these topics for centuries before the connections between them became apparent.

An electric current always gives rise to a **magnetic field** (this is known as electromagnetism). A magnetic field is created by any *moving* charged particles such as electrons. Similarly, a changing magnetic field will induce a current in a nearby conductor. These observations led to the unification of the theories of electricity and magnetism by Michael Faraday in the mid-19th century. A vast technology based on the theories of electromagnetism developed rapidly, and continues to expand today (Figure 12.13).

Faraday's studies were extended by James Clerk Maxwell. He produced mathematical equations that predicted that a changing electric or magnetic field would give rise to transverse waves travelling through space. When he calculated the speed of these waves, it turned out to be the known speed of light. He concluded that light is a wave, known as an **electromagnetic wave**, that can travel through space (including a vacuum) as vibrations of electric and magnetic fields.

Faraday had unified electricity and magnetism; now Maxwell had unified electromagnetism and light. In the 20th century, Abdus Salam (Figure 12.14) managed to unify electromagnetic forces with the weak nuclear force, responsible for radioactive decay. Physicists continue to strive to unify the big ideas of physics; you may occasionally hear talk of **a theory of everything**. This would not truly explain *everything*, but it would explain all known forces, as well as the existence of the various fundamental particles of matter.

Figure 12.13: These telecommunications masts are situated 4.5 km above sea level in Ecuador. They transmit microwaves, a form of electromagnetic radiation, across the mountain range of the Andes.

Figure 12.14: Abdus Salam, the Pakistani physicist, won the 1979 Nobel Prize for Physics for his work on unification of the fundamental forces.

12.7 Electromagnetic radiation

By the end of the 19th century, several types of electromagnetic wave had been discovered:

- radio waves – these were discovered by Heinrich Hertz when he was investigating electrical sparks
- infrared and ultraviolet waves–these lie beyond either end of the visible spectrum
- X-rays – these were discovered by Wilhelm Röntgen and were produced when a beam of electrons collided with a metal target such as tungsten
- γ-rays – these were discovered by Henri Becquerel when he was investigating radioactive substances.

We now regard all of these types of radiation as parts of the same electromagnetic spectrum, and we know that they can be produced in a variety of different ways.

The speed of light

James Clerk Maxwell showed that the speed c of electromagnetic waves in a vacuum (free space) was independent of the frequency of the waves. In other words, all types of electromagnetic wave travel at the same speed in a vacuum. In the SI system of units, c has the value:

$$c = 299\,792\,458 \text{ m s}^{-1}$$

The approximate value for the speed of light in a vacuum, which is often used in calculations, is 3.0×10^8 m s^{-1}.

The wavelength λ and the frequency f of the waves are related by the equation:

$$c = f\lambda$$

This is the same as the wave equation: the wave speed $v = c$. When light travels from a vacuum into a material medium such as glass, its speed **decreases** but its frequency **remains the same**, and so we conclude that its wavelength must decrease. We often characterise different forms of electromagnetic by their different wavelengths. But it is better to characterise them by their different *frequencies*. That's because their wavelengths depend on the medium through which they are travelling.

12.8 Orders of magnitude

Table 12.3 shows the approximate ranges of wavelengths in a vacuum of the principal bands that make up the electromagnetic spectrum. This information is shown as a diagram in Figure 12.15.

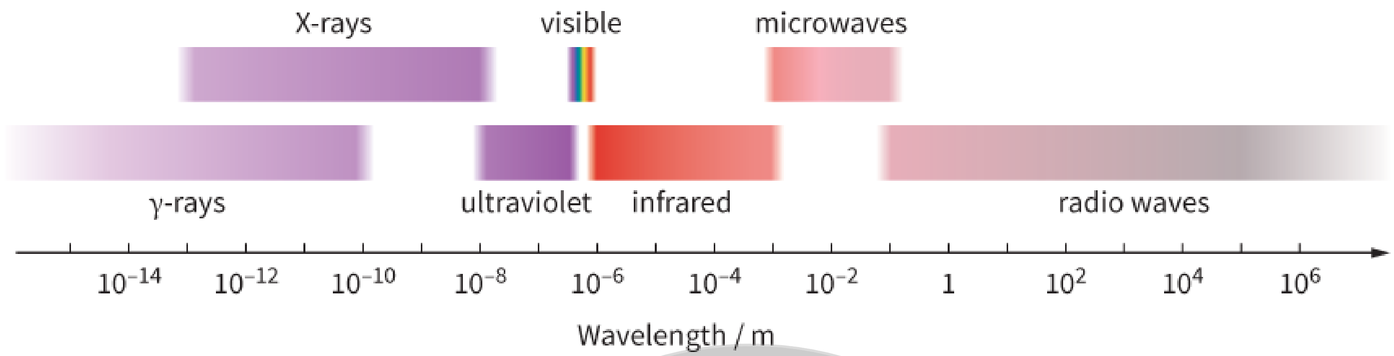


Figure 12.15: Wavelengths of the electromagnetic spectrum. The boundaries between some regions are fuzzy.

Here are some points to note.

- There are no clear divisions between the different ranges or bands in the spectrum. The divisions shown in Table 12.3 are somewhat arbitrary.
- The naming of subdivisions is also arbitrary. For example, microwaves are sometimes regarded as a subdivision of radio waves.
- The wavelength in the range 400 nm to 700 nm in free space (vacuum) are visible to the human eye. Remember, 1 nm = 10^{-9} m
- The ranges of X-rays and γ-rays overlap. The distinction is that X-rays are produced when electrons decelerate rapidly or when they hit a target metal at high speeds. γ-rays are produced by nuclear reactions, such as radioactive decay. There is no difference whatsoever in the radiation between an X-ray and a γ-ray of wavelength, say, 10^{-11} m.

Type of electromagnetic waves	Wavelength range / m
radio waves	$>10^{6}$ to 10^{-1}
microwaves	10^{-1} to 10^{-3}
infrared	10^{-3} to 7×10^{-7}
visible	7×10^{-7} (red) to 4×10^{-7} (violet)
ultraviolet	4×10^{-7} to 10^{-8}
X-rays	10^{-8} to 10^{-13}
γ-rays	10^{-10} to 10^{-16}

Table 12.3: Wavelengths (in a vacuum) of the electromagnetic spectrum.

Questions

11 Copy Table 12.3. Add a third column showing the range of frequencies of each type of radiation.

12 Study Table 12.3 and answer the questions.

- **a** Which type of radiation has the narrowest range of wavelengths?
- **b** Which has the second narrowest range?
- **c** What is the range of wavelengths of microwaves, in millimetres?
- **d** What is the range of wavelengths of visible light, in nanometres?
- **e** What is the frequency range of visible light?

13 For each of the following wavelengths measured in a vacuum, state the type of electromagnetic radiation to which it corresponds.

- **a** 1 km
- **b** 3 cm
- **c** 5000 nm
- **d** 500 nm
- **e** 50 nm
- **f** 10^{-12} m

14 For each of the following frequencies, state the type of electromagnetic wave to which it corresponds.

- **a** 200 kHz
- **b** 100 MHz
- **c** 5×10^{14} Hz
- **d** 10^{18} Hz

12.9 The nature of electromagnetic waves

An electromagnetic wave is a disturbance in the electric and magnetic fields in space. Figure 12.16 shows how we can represent such a wave. In this diagram, the wave is travelling from left to right.

The **electric field** is shown oscillating in the vertical plane. The magnetic field is shown oscillating in the horizontal plane. These are arbitrary choices; the point is that the two fields vary at right angles to each other, and also at right angles to the direction in which the wave is travelling. This shows that electromagnetic waves are transverse waves.

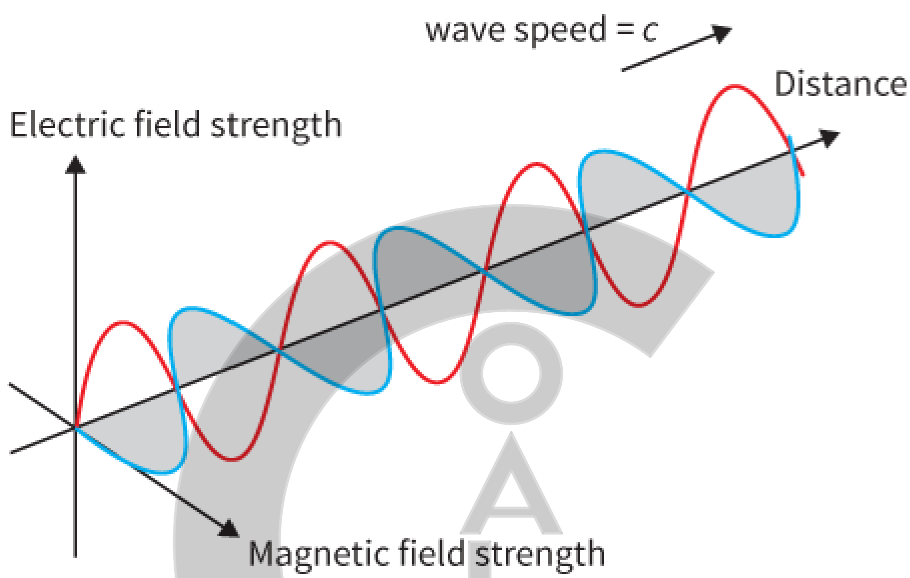


Figure 12.16: An electromagnetic wave is a periodic variation in electric and magnetic fields.

12.10 Polarisation

Polarisation is a wave property associated with transverse waves only.

Imagine you fixed one end of a rope to a post. Grab the other end of the rope and pull it tight so that it is stretched out horizontally. Move the rope repeatedly vertically up and down. This will produce a transverse wave on the rope. The vibrations of the rope are in just one plane – the vertical plane. The vibrations are described as **plane polarised** in the vertical plane. You can produce plane polarised vibrations in the horizontal plane by moving the rope repeatedly from side to side. It would also be fun to keep changing the direction of vibration of the rope – in this case, you will produce an unpolarised wave where the vibrations are in more than one plane.

A plane polarised wave incident at a vertical slit will pass through this slit. When the slit is turned through 90°, the plane polarised wave will be blocked. When an unpolarised wave is incident at a vertical slit, then all vibrations, other than those in the vertical plane, will be blocked (see Figure 12.17). The wave passing through the slit will be a plane polarised wave in the vertical plane.

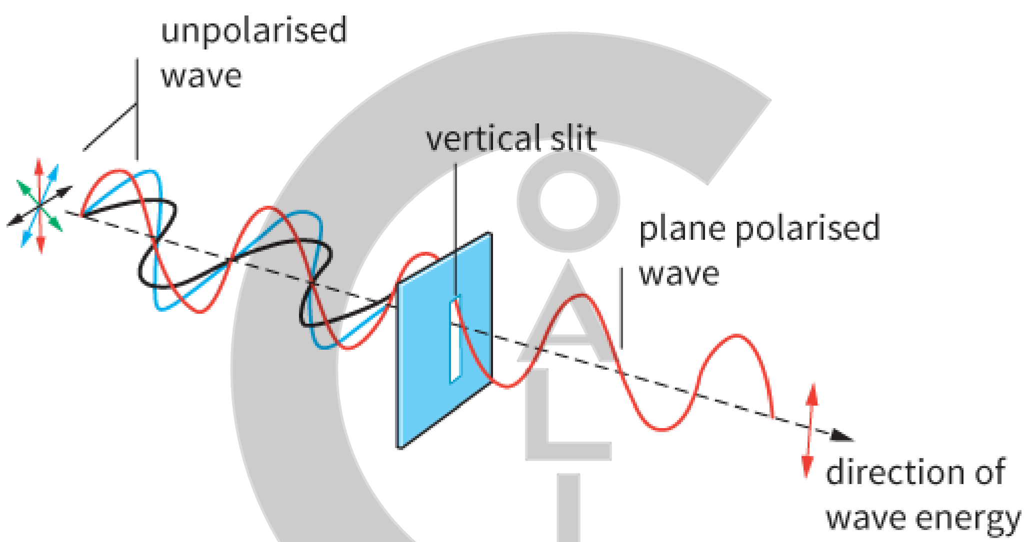


Figure 12.17: The slit helps to produce a plane polarised wave.

Only transverse waves can be plane polarised. So, it should be possible to produce plane polarised light waves. In fact, all types of electromagnetic waves can be plane polarised.

Longitudinal waves vibrate along the direction of wave travel, so no matter what the orientation of the slit, the waves will be able to get through. In short, longitudinal waves, such as sound, cannot be polarised.

Polarised light

Light is a transverse wave. Its transverse nature can be demonstrated by polarising light. As mentioned previously, light consists of oscillating electric and magnetic fields. Light from the Sun, or a filament lamp, is unpolarised. This means it has oscillating electric fields in all planes at right angles to the direction in which it travels. What can we use to plane polarise such light?

We can use transparent polymer material, such as a Polaroid, a type of polarising filter. The Polaroid has long chains of molecules all aligned in one particular direction. Any electric field vibrations along these chains of molecules are absorbed. The energy absorbed is transferred to thermal energy in the Polaroid. Electric field vibrations at right angles to the chains of molecules are transmitted with negligible absorption. Figure 12.18 shows the unpolarised light incident at a Polaroid–the transmitted light is plane polarised.

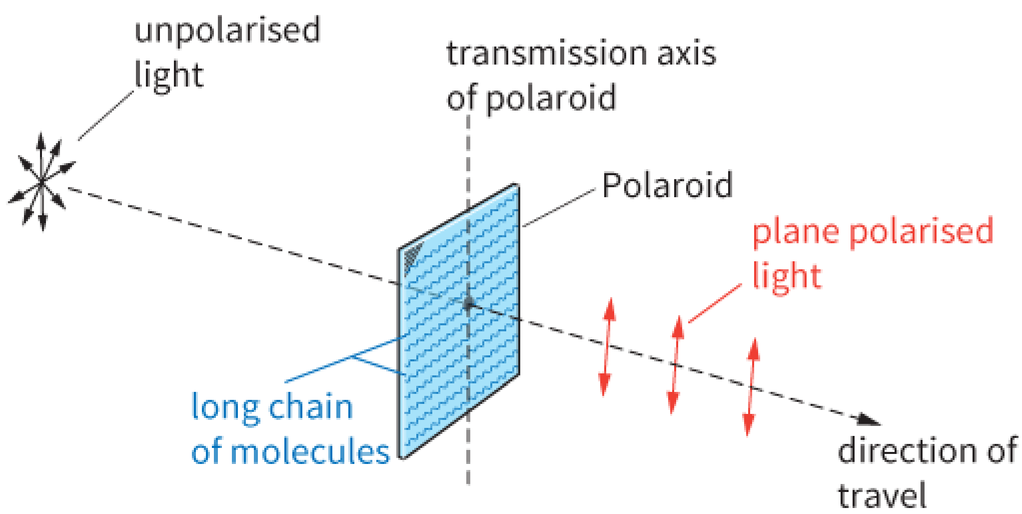


Figure 12.18: A Polaroid has a unique axis of transmission for light.

What would happen when you view unpolarised light using two Polaroids? Figure 12.19 shows plane polarised light produced by the first Polaroid. This plane polarised light is incident at the second Polaroid, whose transmission axis is initially vertical. The second Polaroid is often known as the analyser. The incident light passes straight through. Now rotate the analyser through 90°, so its transmission axis is horizontal. This time, the analyser will absorb all the light. The analyser will appear black. Turning the analyser through a further 90° will let the light through the analyser again. What happens at angles other than 0° and 90° is discussed later.

Here are a few things you can try with a single Polaroid.

- Light reflected from the surface of water, or glass, is partially polarised in a plane parallel to the reflecting surface. Holding a Polaroid with its transmission axis vertical, will reduce the glare of reflected light. This is how your Polaroid sunglasses work. Polarising filters help in photography (Figure 12.20).
- Light from your laptop screen is plane polarised. You can completely cut out the display by viewing the screen through a Polaroid. You can observe the same effect with your LCD calculator display. Twist the Polaroid, and see the display vanish.

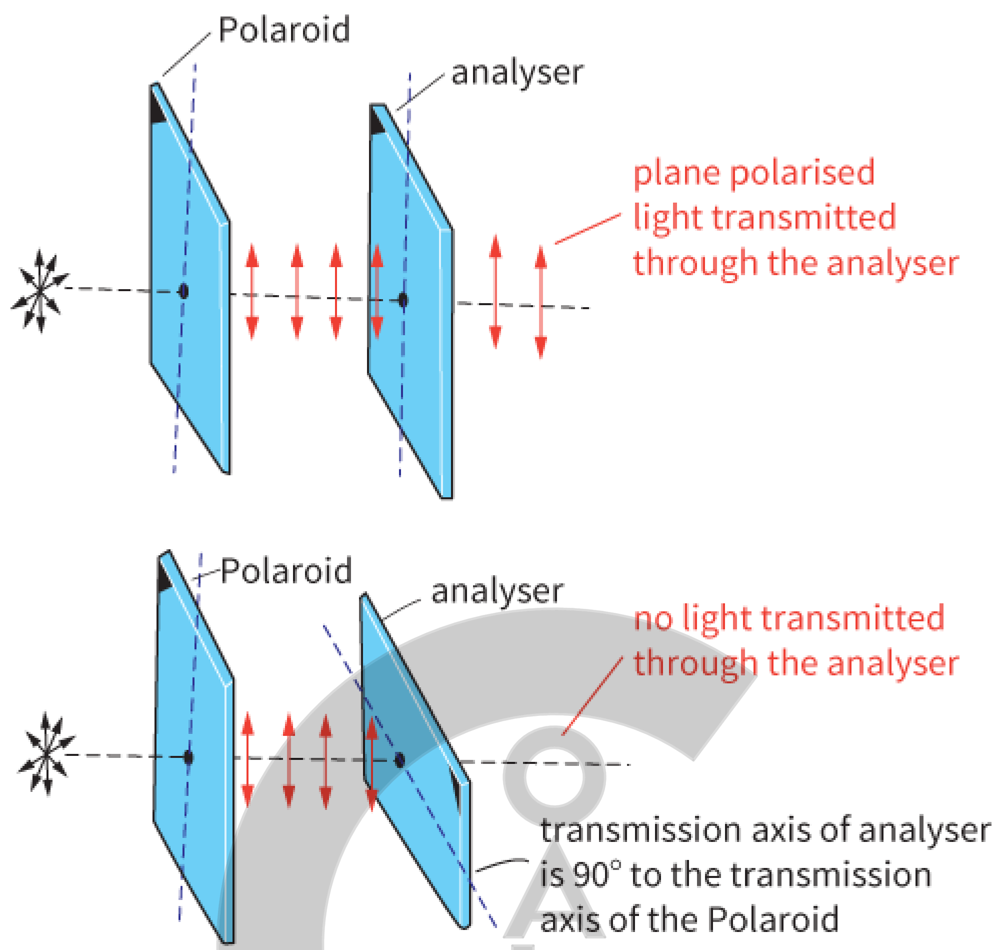


Figure 12.19: The light is blocked by the analyser when its transmission axis is 90° to the plane of the incident light. The dashed lines are transmission axes of the Polaroid and the analyser.

Figure 12.20: Polarising filters are used in photography – there is no glare and you can see the sharks and the boy snorkeling.

Malus's law

Figure 12.21 shows plane polarised light incident at a Polaroid. The transmission axis of this Polaroid is at an angle θ to the plane of the incident light. Now you already know that when $\theta = 0$, then the light will go through the Polaroid, and when $\theta = 90°$, there is no transmitted light. The intensity of the transmitted light depends on the angle θ.

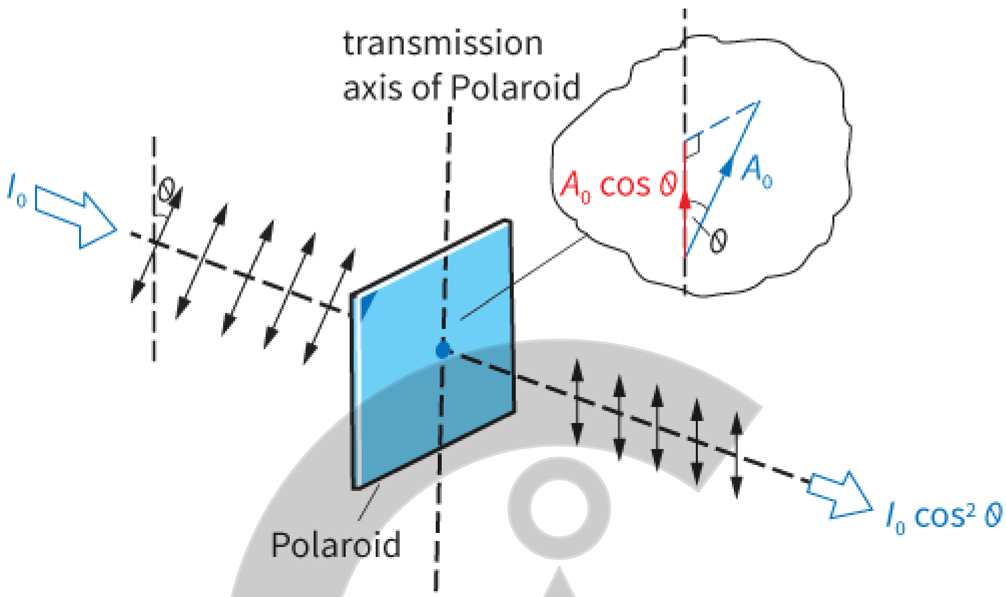


Figure 12.21: The amplitude, and hence the intensity of light, transmitted through the Polaroid depends on the angle θ.

Consider the incident plane polarised light of amplitude A_0. The component of the amplitude transmitted through the Polaroid along its transmission axis is $A_0 \cos\theta$. You know that the intensity of light is directly proportional to the amplitude squared. So, the intensity of light transmitted will be given by the expression:

$$I = I_0 \cos^2\theta$$

where I_0 is the intensity of the incident and I is the transmitted intensity at an angle θ between the transmission axis of the Polaroid and the plane of the incident polarised wave.

The relationship is known as Malus's law.

KEY EQUATION

Malus's law:

$$I = I_0 \cos^2\theta$$

Note that the fraction of the light intensity transmitted is equal to $\cos^2\theta$. This means that a graph of I against θ is a cosine squared graph, see Figure 12.22.

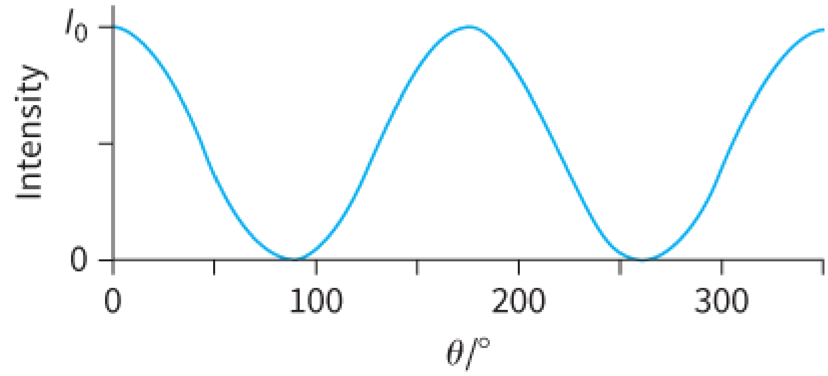


Figure 12.22: Variation of transmitted intensity *I* with angle θ. Notice maximum intensity when θ = 0°, 180° and so on, and zero when θ = 90°, 270° and so on.

Questions

15 Explain what happens to unpolarised light incident at a Polaroid.

16 Plane polarised light of intensity 12 W m^{-2} is incident at a Polaroid.

Calculate the intensity of the transmitted light when the angle between the plane of polarisation of the incident light and the transmission axis of the Polaroid is

a 45°

b 60°.

17 Plane polarised light is incident at a Polaroid.

Calculate the angle θ, which gives transmitted light of intensity 30% that of the incident intensity of light.

REFLECTION

Can you think of any applications of Malus's law?

Make a list of some key words in this chapter. Ask a classmate to make a similar list. Now compare your lists. How good was your list? Did you miss out anything important? What things might you want more help with?

SUMMARY

A progressive wave carries energy from one place to another.
There are two types of progressive waves–longitudinal and transverse. Longitudinal waves have vibrations parallel to the direction in which the wave travels, whereas transverse waves have vibrations at right angles to the direction in which the wave travels.
Displacement is the distance of a point on the wave from its undisturbed position or equilibrium position.
Amplitude is maximum displacement of a wave.
Wavelength is the distance between two adjacent points on a wave oscillating in step with each other.
Period is time taken for one complete oscillation of a point in a wave.
Frequency is the number of oscillations per unit time of a point in a wave.
Phase difference is the fraction of a cycle between oscillating particles, expressed either in degrees or in radians.
Two points on a wave separated by a distance of one wavelength have a phase difference of 0° or 360°.
The frequency *f* of a wave is related to its period *T* by the equation: $$f = \frac{1}{T}$$
The frequency of a sound wave can be measured using a cathode-ray oscilloscope (CRO).
The speed of all waves is given by the wave equation: wave speed = frequency × wavelength $$v = f\lambda$$
The Doppler effect is the change in an observed wave frequency when a source moves with speed v_s. The observed frequency is given by: $$f_0 = \frac{f_s v}{(v \pm v_s)}$$
The intensity of a wave is defined as the wave power transmitted per unit area at right angles to the wave velocity. So: $$\text{intensity} = \frac{\text{power}}{\text{area}}$$
Intensity has units of W m^{-2}.
The intensity *I* of a wave is directly proportional to the square of the amplitude $A (I \propto A^2)$
All electromagnetic waves travel at the same speed of 3.0 × 10^8 m s^{-1} in a vacuum, but have different

wavelengths and frequencies. Electromagnetic waves are transverse waves.
The regions of the electromagnetic spectrum in order of increasing wavelength are: γ-rays, X-rays, ultraviolet, visible, infrared, microwaves and radio waves.
A plane polarised wave has oscillations in just one plane.
Only transverse waves can be plane polarised.
Equation for Malus's law: $$I = I_0 \cos^2 \theta$$ where I is the intensity of the transmitted light through the polarising filter, I_0 is the incident intensity of light and θ is the angle between the transmission axis of the filter and the plane of polarisation of the incident light.

EXAM-STYLE QUESTIONS

1 What is the correct unit for intensity? **[1]**

A J m^2

B J s^{-1}

C W m^2

D W m^{-2}

2 This image shows the screen of an oscilloscope. The time-base of the oscilloscope is set at 500 μs div^{-1}.

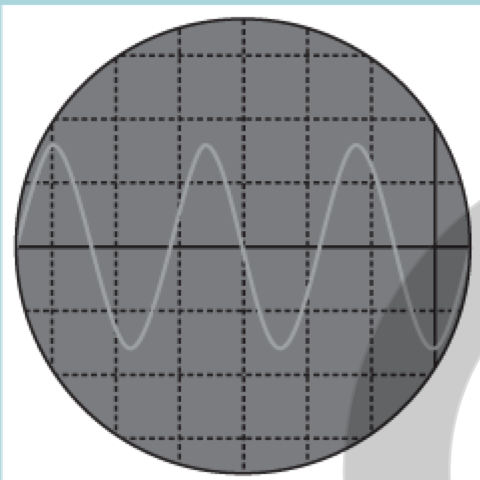


Figure 12.23

Calculate the time period of the signal and hence its frequency. **[3]**

3 **a** State **two** main properties of electromagnetic waves. **[2]**

b State **one** major difference between microwaves and radio waves. **[1]**

c **i** Estimate the wavelength in metres of X-rays. **[1]**

ii Use your answer to **i** to determine the frequency of the X-rays. **[1]**

[Total: 5]

4 A student is sitting on the beach, observing a power boat moving at speed on the sea. The boat has a siren emitting a constant sound of frequency 420 Hz.

The boat moves around in a circular path with a speed of 25 m s^{-1}. The student notices that the pitch of the siren changes with a regular pattern.

a Explain why the pitch of the siren changes, as observed by the student. **[1]**

b Determine the maximum and minimum frequencies that the student will hear. **[4]**

c At which point in the boat's motion will the student hear the most high-pitched note? **[1]**

(Speed of sound in air = 330 m s^{-1}.)

[Total: 6]

5 This diagram shows some air particles as a sound wave passes.

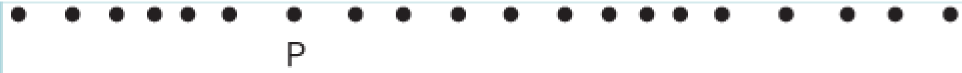


Figure 12.24

a On a copy of the diagram, mark:

i a region of the wave that shows a compression–label it C **[1]**

ii a region of the wave that shows a rarefaction–label it R. **[1]**

b Describe how the particle labelled P moves as the wave passes. **[2]**

c The sound wave has a frequency of 240 Hz. Explain, in terms of the movement of an individual particle, what this means. **[2]**

d The wave speed of the sound is 320 m s^{-1}. Calculate the wavelength of the wave. **[2]**

[Total: 8]

6 In an experiment, a student is determining the speed of sound using the equation $v = f\lambda$. The values of frequency f and wavelength λ are shown below:

$f = 1000 \pm 10$ Hz

$\lambda = 33 \pm 2$ cm

Determine the speed v including the absolute uncertainty. **[5]**

7 This diagram shows a loudspeaker producing a sound and a microphone connected to a cathode-ray oscilloscope (CRO).

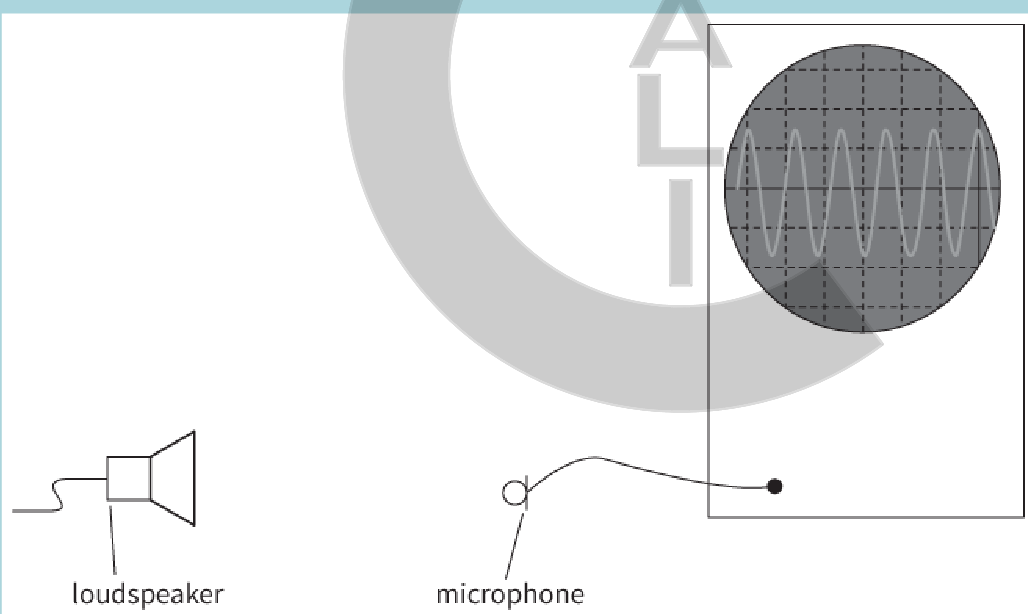


Figure 12.25

a Sound is described as a longitudinal wave. Describe sound waves in terms of the movements of the air particles. **[1]**

b The time-base on the oscilloscope is set at 5 ms div^{-1}. Calculate the frequency of the CRO trace. **[2]**

c The wavelength of the sound is found to be 1.98 m.

Calculate the speed of sound. **[2]**

[Total: 5]

8 The Doppler effect can be used to measure the speed of blood. Ultrasound, which is sound of high frequency, is passed from a transmitter into the body, where it reflects off particles in the blood. The shift in frequency is measured by a stationary detector, placed outside the body and close to the transmitter.

In one patient, particles in the blood are moving at a speed of 30 cm s^{-1} in a direction directly away from the transmitter. The speed of ultrasound in the body is 1500 cm s^{-1}.

This situation is partly modelled by considering the particles to be emitting sound of frequency 4.000 MHz as they move away from the detector. This sound passes to the detector outside the body and the frequency measured by the detector is not 4.000 MHz.

a **i** State whether the frequency received by the stationary detector is higher or lower than the frequency emitted by the moving particles. **[1]**

ii Explain your answer to part **i**. **[3]**

b Calculate the difference between the frequency emitted by the moving particles and the frequency measured by the detector. **[3]**

c Suggest why there is also a frequency difference between the sound received by the particles and the sound emitted by the transmitter. **[1]**

[Total: 8]

9 **a** State what is meant by plane polarised light. **[1]**

b Reflected light from the surface of water is partially plane polarised.

Describe briefly how you could demonstrate this. **[2]**

c Vertically plane polarised light is incident on three polarising filters. The transmission axis of the first Polaroid is vertical. The transmission axis of the second filter is 45° to the vertical and the transmission axis of the last filter is horizontal.

Show that the intensity of light emerging from the final filter is **not** zero. **[4]**

[Total: 7]

SELF-EVALUATION CHECKLIST

After studying this chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand transverse and longitudinal waves	12.2			
define the terms wavelength, amplitude, frequency, wave speed, phase difference and intensity	12.1, 12.2, 12.3			
use a cathode-ray oscilloscope (CRO) to determine frequency and amplitude	12.1			
use the equations: $v = f\lambda$, intensity $= \frac{\text{power}}{\text{area}}$ and intensity $\propto$ amplitude^2	12.3, 12.4			
describe the Doppler effect for sound waves	12.5			
use the Doppler equation $f_0 = \frac{f_s v}{(v \pm v_s)}$ for approaching and receding sound-source	12.5			
understand the properties of electromagnetic waves	12.6			
recall that wavelengths in the range 400–700 nm in free space are visible to the human eye	12.8			
describe and understand polarisation of light	12.10			
use Malus's law: $I = I_0 \cos^2 \theta$	12.10			

〉Chapter 13

Superposition of waves

LEARNING INTENTIONS

In this chapter you will learn how to:

- explain and use the principle of superposition
- explain the meaning of diffraction, interference, path difference and coherence
- understand experiments that demonstrate diffraction
- understand experiments that demonstrate two-source interference
- understand the conditions required if two-source interference fringes are to be observed
- recall and use $\lambda = \frac{ax}{D}$ for double-slit interference using light
- recall and use $d \sin\theta = n\lambda$ for a diffraction grating
- use a diffraction grating to determine the wavelength of light.

BEFORE YOU START

- Can you recall the general properties of waves, including electromagnetic waves? Write down as many properties as you can remember.
- Knowledge of phase difference is vital in understanding how waves combine in space–remind yourself by writing down the phase difference of two particles oscillating in step, and two particles oscillating in antiphase.

VIBRATIONS MAKING WAVES

High-level of noise would not be suitable in some jobs, such as working in a ship's engine room or looking after airplanes landing and lifting off at an airport. The simple solution would be to wear headphones. These will significantly reduce the intensity of the noise reaching the ears. Wearing noise-cancelling headphones will do a better job at protecting the ears. Electronics within such headphones create their own sound that is an exact copy of the incident noise, except it is always in antiphase (phase difference of 180°) with the noise. The addition of these two waves has the effect of reducing the intensity of the sound reaching the ears to almost zero.

Noise-cancelling headphones are useful in some situations, but they are not ideal if you are at a concert!

Can you think of other jobs where such headphones would be useful?

In this chapter, we will study how waves add-up and cancel-out. The principle of superposition of waves is an excellent starting point.

Figure 13.1: The headphones actively cancel out the noise – protecting the ears from damage.

13.1 The principle of superposition of waves

In Chapter 12, we studied the production of waves and the difference between longitudinal and transverse waves. In this chapter, we are going to consider what happens when two or more waves meet at a point in space and combine together (Figure 13.2).

So what happens when two waves arrive together at the same place? We can answer this from our everyday experience. What happens when the beams of light waves from two torches cross over? They pass straight through one another. Similarly, sound waves pass through one another, apparently without affecting each other. This is very different from the behaviour of **particles**. Two marbles meeting in mid-air would ricochet off one another in a very un-wave-like way. If we look carefully at how two sets of waves interact when they meet, we find some surprising results.

Figure 13.2: Ripples produced when drops of water fall into a swimming pool. The ripples overlap to produce a complex pattern of crests and troughs.

When two waves meet, they combine, with the displacements of the two waves adding together. Figure 13.3 shows the displacement–distance graphs for two sinusoidal waves (blue and green) of different wavelengths. It also shows the resultant wave (red), which comes from combining these two. How do we find this resultant displacement shown in red?

Consider position A. Here, the displacement of both waves is zero, and so the resultant displacement must also be zero. At position B, both waves have positive displacement. The resultant displacement is found by adding these together. At position C, the displacement of one wave is positive while the other is negative. The resultant displacement lies between the two displacements. In fact, the resultant displacement is the algebraic sum of the displacements of waves A and B; that is, their sum, taking account of their signs (positive or negative).

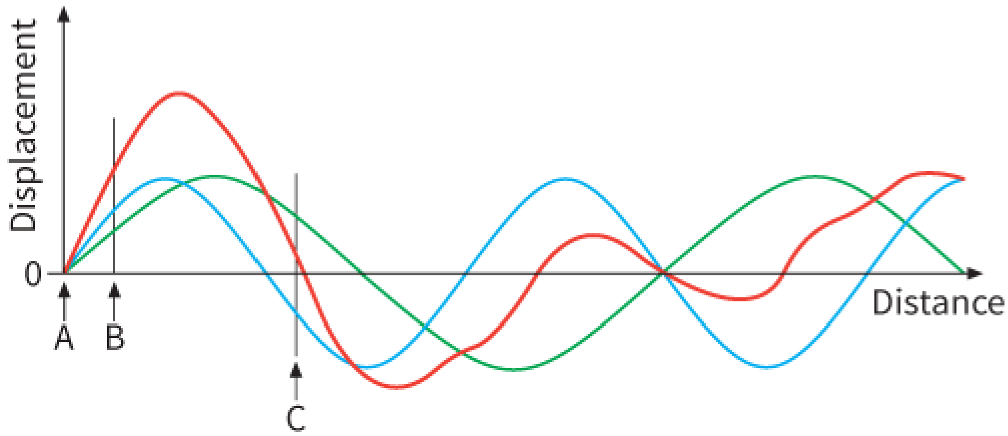


Figure 13.3: Adding two waves by the principle of superposition – the **red line** is the resultant wave.

We can work our way along the distance axis in this way, calculating the resultant of the two waves by algebraically adding them up at intervals. Notice that, for these two waves, the resultant wave is a rather complex wave with dips and bumps along its length.

The idea that we can find the resultant of two waves that meet at a point simply by adding up the displacements at each point is called the **principle of superposition** of waves. This principle can be applied to more than two waves and also to all types of waves. A statement of the principle of superposition is:

When two or more waves meet at a point, the resultant displacement is the algebraic sum of the displacements of the individual waves.

Question

1 On graph paper, draw two 'triangular' waves similar to those shown in Figure 13.4. (These are easier to work with than sinusoidal waves.) One should have wavelength 8.0 cm and amplitude 2.0 cm. The other should have wavelength 16.0 cm and amplitude 3.0 cm.

Use the principle of superposition of waves to determine the resultant displacement at suitable points along the waves, and draw the complete resultant wave.

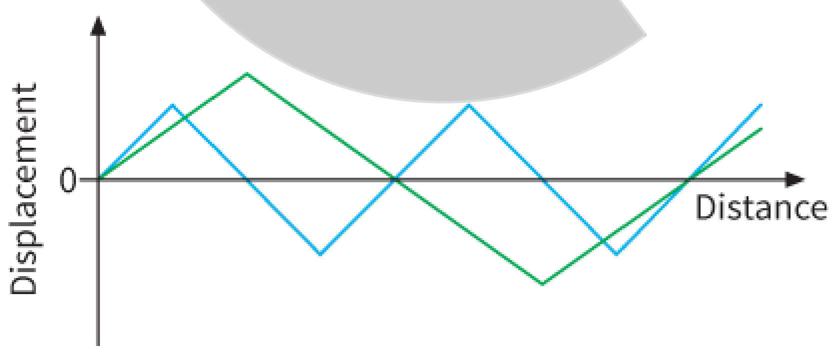


Figure 13.4: Two triangular waves.

13.2 Diffraction of waves

You should be aware that all waves (such as sound and light) can be reflected and refracted. Another wave phenomenon that applies to all waves is that they can be diffracted. **Diffraction** is the spreading of a wave as it passes through a gap or around an edge. It is easy to observe and investigate diffraction effects using water waves, as shown in Practical Activity 13.1.

Diffraction of sound and light

Diffraction effects are greatest when waves pass through a gap with a width roughly equal to their wavelength of the waves. This is useful in explaining why we can observe diffraction readily for some waves, but not for others. For example, sound waves in the audible range have wavelengths from a few centimetres to a few metres (see Table 12.1). So, we might expect to observe diffraction effects for sound in our environment. Sounds, for example, diffract as they pass through doorways. The width of a doorway is comparable to the wavelength of a sound and so a noise in one room spreads out into the next room.

Visible light has much shorter wavelengths (about 5×10^{-7} m). It is not diffracted noticeably by doorways because the width of the gap is a million times larger than the wavelength of light. However, we can observe diffraction of light by passing it through a very narrow slit or a very small hole. When laser light is directed onto a slit whose width is comparable to the wavelength of the incident light, it spreads out into the space beyond to form a smear on the screen (Figure 13.5). An adjustable slit allows you to see the effect of gradually narrowing the gap.

You can see the effects of diffraction for yourself by making a narrow slit with your two thumbs and looking through the slit at a distant light source (Figure 13.8). By gently pressing your thumbs together to narrow the gap between them, you can see the effect of narrowing the slit.

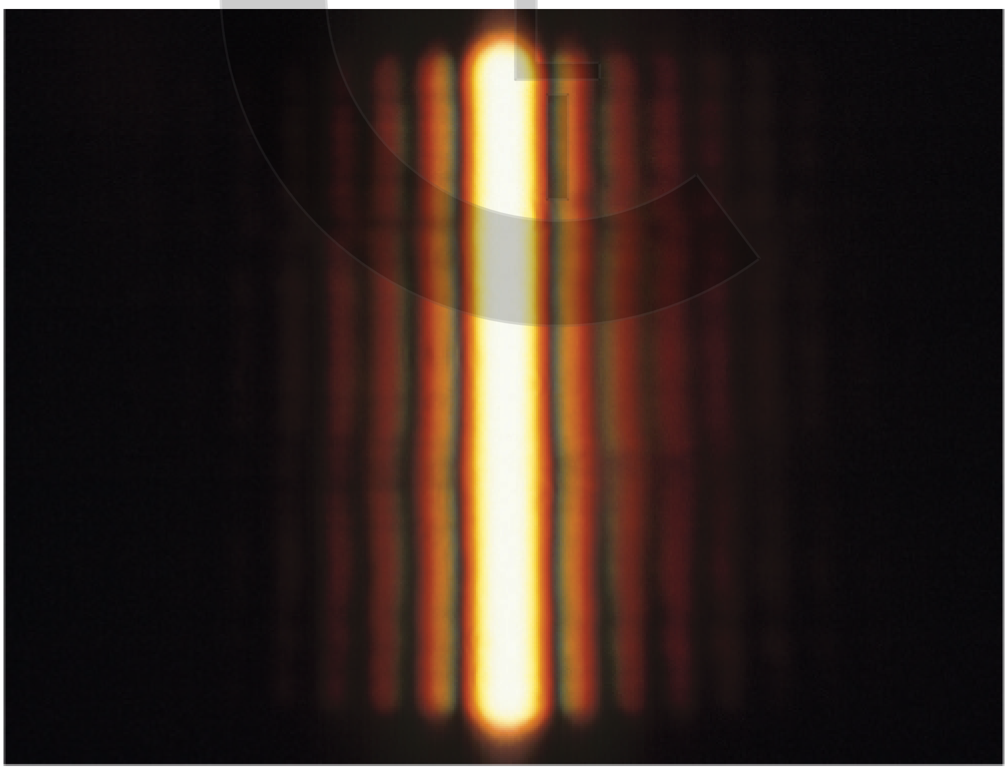


Figure 13.5: Light is diffracted as it passes through a very narrow slit.

PRACTICAL ACTIVITY 13.1

Observing diffraction in a ripple tank

A ripple tank can be used to show diffraction. Plane waves are generated using a vibrating bar, and **move towards** a gap in a barrier (Figure 13.6). Where the ripples strike the barrier, they are reflected back. Where they arrive at the gap, however, they pass through and spread out into the space beyond. It is this spreading out of waves as they travel through a gap (or past the edge of a barrier) that is called diffraction.

The extent to which ripples are diffracted depends on the width of the gap. This is illustrated in Figure 13.6. The lines in this diagram show the wavefronts. It is as if we are looking down on the ripples from above, and drawing lines to represent the tops of the ripples at some instant in time. The separation between adjacent wavefronts is equal to the wavelength λ of the ripples.

When the waves encounter a gap in a barrier, the amount of diffraction depends on the width of the gap. There is hardly any noticeable diffraction when the gap is very much larger than the wavelength. As the gap becomes narrower, the diffraction effect becomes more noticeable. It is greatest when the width of the gap is roughly equal to the wavelength of the ripples.

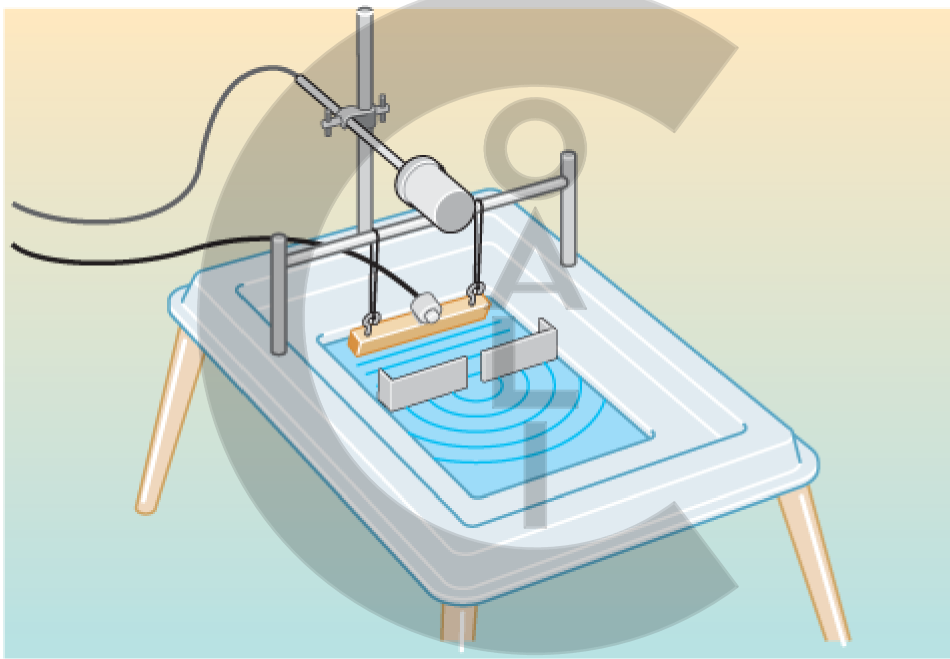


Figure 13.6: Ripples, initially straight, spread out into the space beyond the gap in the barrier.

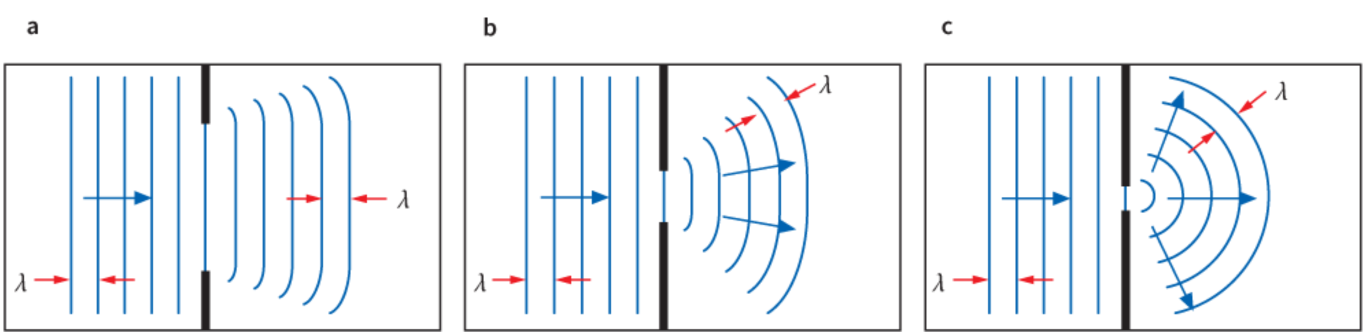


Figure 13.7: The extent to which ripples spread out depends on the relationship between their wavelength and the width of the gap. In **a**, the width of the gap is very much greater than the wavelength and there is hardly any noticeable diffraction. In **b**, the width of the gap is greater than the wavelength and there is

limited diffraction. In **c**, the gap width is approximately equal to the wavelength and the diffraction effect is greatest.

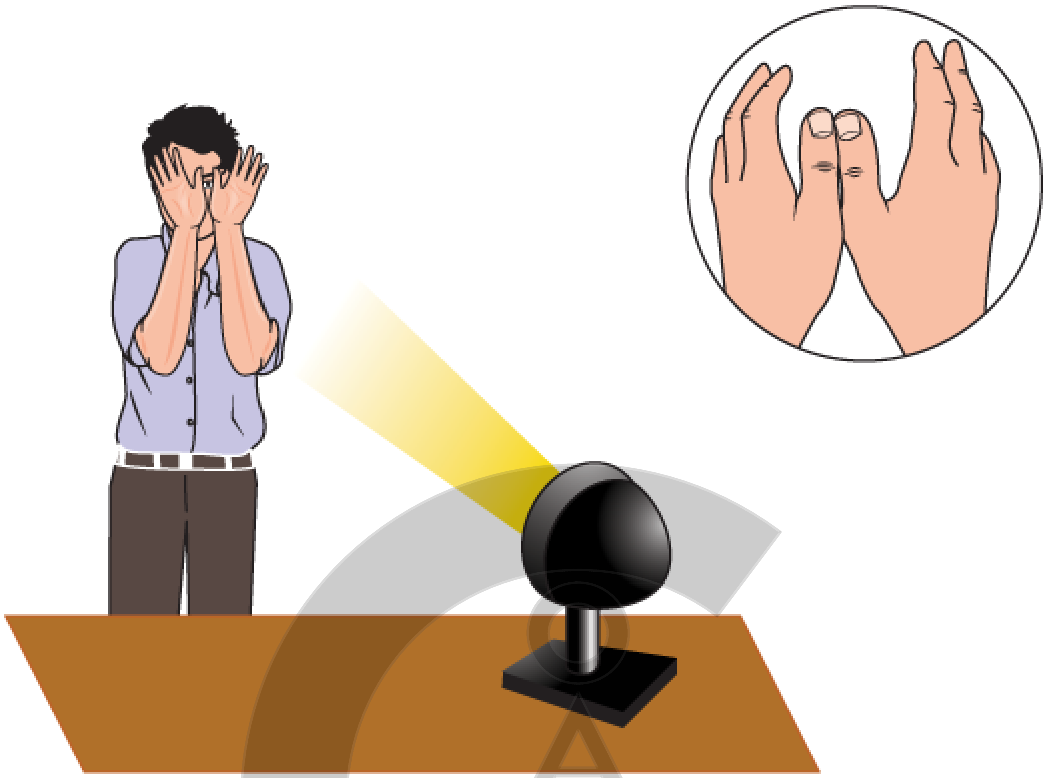


Figure 13.8: You can see the effects of diffraction by looking at a bright source (lamp) through a narrow slit. What happens when you make the slit narrower? What happens to the amount of diffraction when you put different coloured filters in front of the lamp? What does this tell you about the wavelengths of the different colours?

Diffraction of radio waves and microwaves

Radio waves can have wavelengths of the order of a kilometre. These waves are easily diffracted by gaps in the hills and by the tall buildings around our towns and cities. Microwaves, used by the mobile phone network, have wavelengths of about 10 cm. These waves are not easily diffracted (because their wavelengths are much smaller than the dimensions of the gaps) and mostly travel through space in straight lines.

Cars need external radio aerials because radio waves have wavelengths longer than the size of the windows, so they cannot diffract into the car. If you try listening to a radio in a train without an external aerial, you will find that FM signals can be picked up weakly (their wavelength is about 3 m), but AM signals, with longer wavelengths, cannot get in at all.

Question

2 A microwave oven (Figure 13.9) uses microwaves with a wavelength of 12.5 cm. The front door of the oven is made of glass with a metal grid inside; the gaps in the grid are a few millimetres across.

Explain how this design allows us to see the food inside the oven, while the microwaves are not allowed to escape into the kitchen (where they might harm us).

Figure 13.9: A microwave oven has a metal grid in the door to keep microwaves in and let light out

Explaining diffraction

Diffraction is a wave effect that can be explained by the principle of superposition. We have to think about what happens when a plane ripple reaches a gap in a barrier (Figure 13.10). Each point on the surface of the water in the gap is moving up and down. Each of these moving points can be thought of as a source of new ripples spreading out into the space beyond the barrier. Now we have a lot of new ripples, and we can use the principle of superposition to find their resultant effect. Without trying to calculate the effect of an infinite number of ripples, we can say that in some directions the ripples add together while in other directions they cancel out.

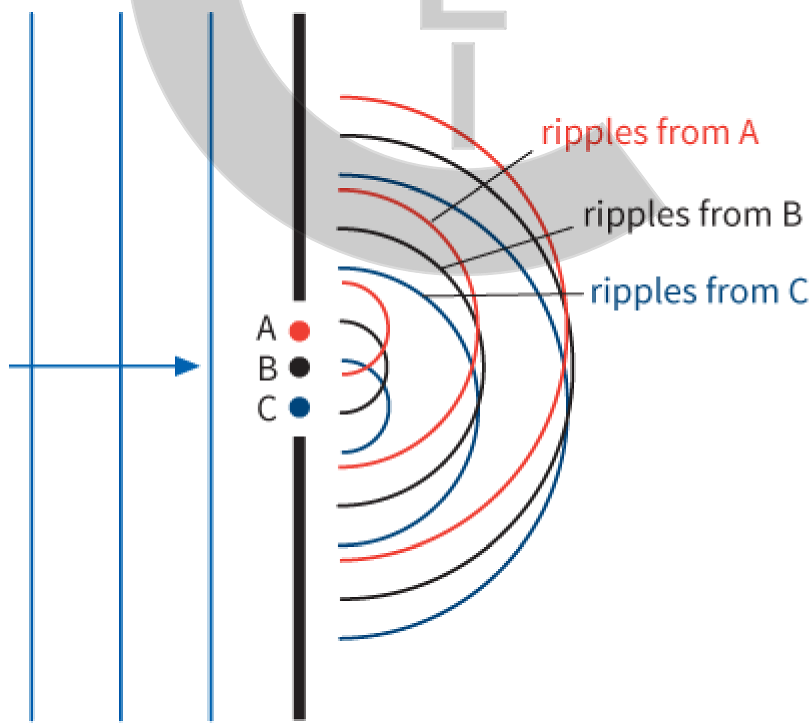


Figure 13.10: Ripples from all points across the gap contribute to the pattern in the space beyond.

COALI

13.3 Interference

Adding waves of different wavelengths and amplitudes results in complex waves – by complex, we really mean not sinusoidal. We can find some interesting effects if we consider what happens when two waves of the same type, and having the same wavelength, overlap at a point. Again, we will use the principle of superposition to explain what we observe.

PRACTICAL ACTIVITY 13.2

Observing interference

Interference of sound waves

A simple experiment shows what happens when two sets of sound waves meet. Two loudspeakers are connected to a single signal generator (Figure 13.11). They each produce sound waves of the same wavelength. Walk around in the space in front of the loudspeakers; you will hear the resultant effect.

You may predict that we would hear a sound twice as loud as that from a single loudspeaker. However, this is not the case. At some points, the sound is louder than for a single loudspeaker. At other points, the sound is much quieter. The space around the two loudspeakers consists of a series of loud and quiet regions. We are observing the phenomenon known as **interference**. This phenomenon results in the formation of points of cancellation and reinforcement where two coherent waves pass through each other.

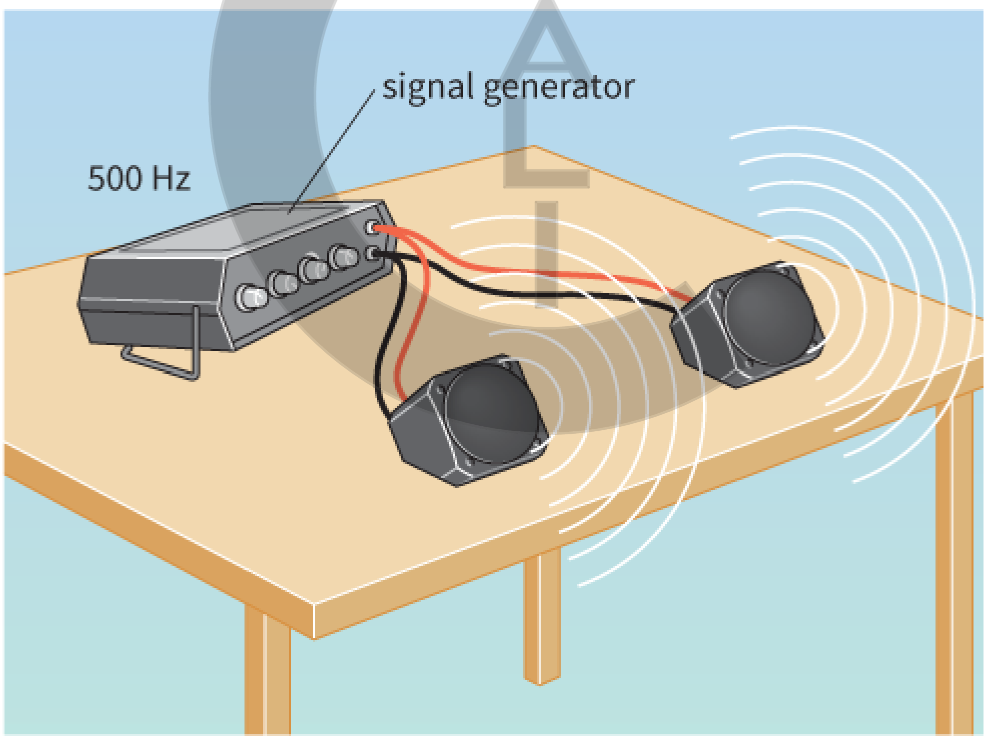


Figure 13.11: The sound waves from two loudspeakers combine to give an interference pattern. This experiment is best done outside so that reflections of sounds (or echoes) do not affect the results.

Interference in a ripple tank

Look at Figure 13.12. The two dippers in the ripple tank should be positioned so that they are just touching the surface of the water. When the bar vibrates, each dipper acts as a point-source of circular ripples spreading outwards. Where these sets of ripples overlap, we observe an interference pattern. Another way to observe interference in a ripple tank is to use plane waves passing through two gaps in a barrier. The water waves are diffracted at the two gaps and then interfere beyond the gaps. Figure 13.13 shows the interference pattern produced by two vibrating dippers in a ripple tank.

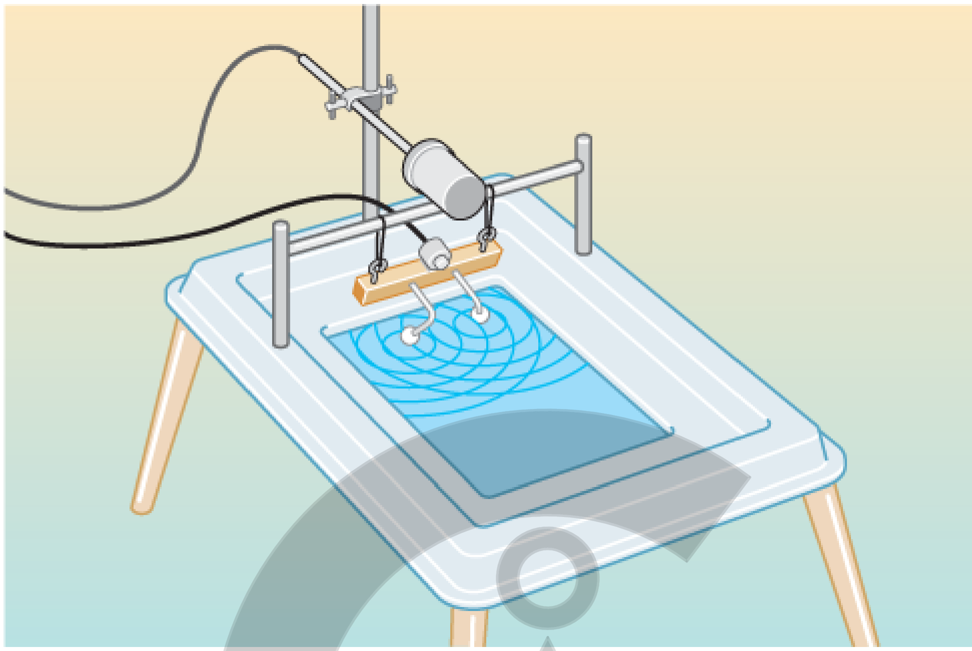


Figure 13.12: A ripple tank can be used to show how two sets of circular ripples combine.

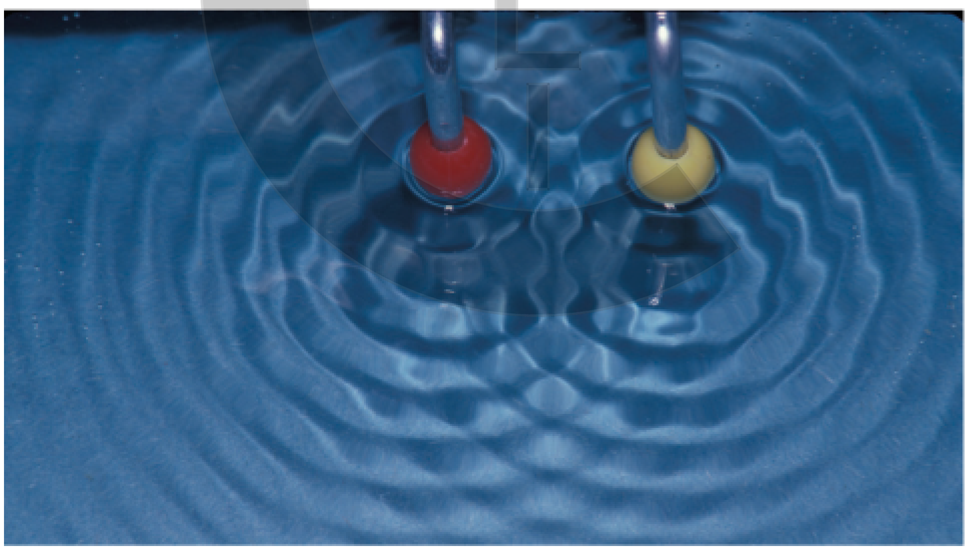


Figure 13.13: Ripples from two point-sources produce an interference pattern.

Explaining interference

Figure 13.14 shows how interference arises. The loudspeakers in Figure 13.11 (Practical Activity 13.2) are emitting waves that are in phase because both are connected to the same signal generator. At each point in front of the loudspeakers, waves are arriving from the two loudspeakers. At some points, the two waves arrive in phase (in step) with one another and with equal amplitude (Figure 13.14a). The principle of superposition predicts that the resultant wave has twice the amplitude of a single wave. We hear a louder sound.

At other points, something different happens. The two waves arrive **completely out of phase** or in antiphase (phase difference is 180°) with one another (Figure 13.14b). There is a cancelling out, and the resultant wave has zero amplitude. At this point, we would expect silence. At other points again, the waves are neither perfectly out of step nor perfectly in step, and the resultant wave has an amplitude less than that at the loudest point.

Where two waves arrive at a point in phase with one another so that they add up, we call this effect **constructive interference**. Where they cancel out, the effect is known as **destructive interference.** Where two waves have different amplitudes but are in phase (Figure 13.14c), constructive interference results in a wave whose amplitude is the sum of the two individual amplitudes.

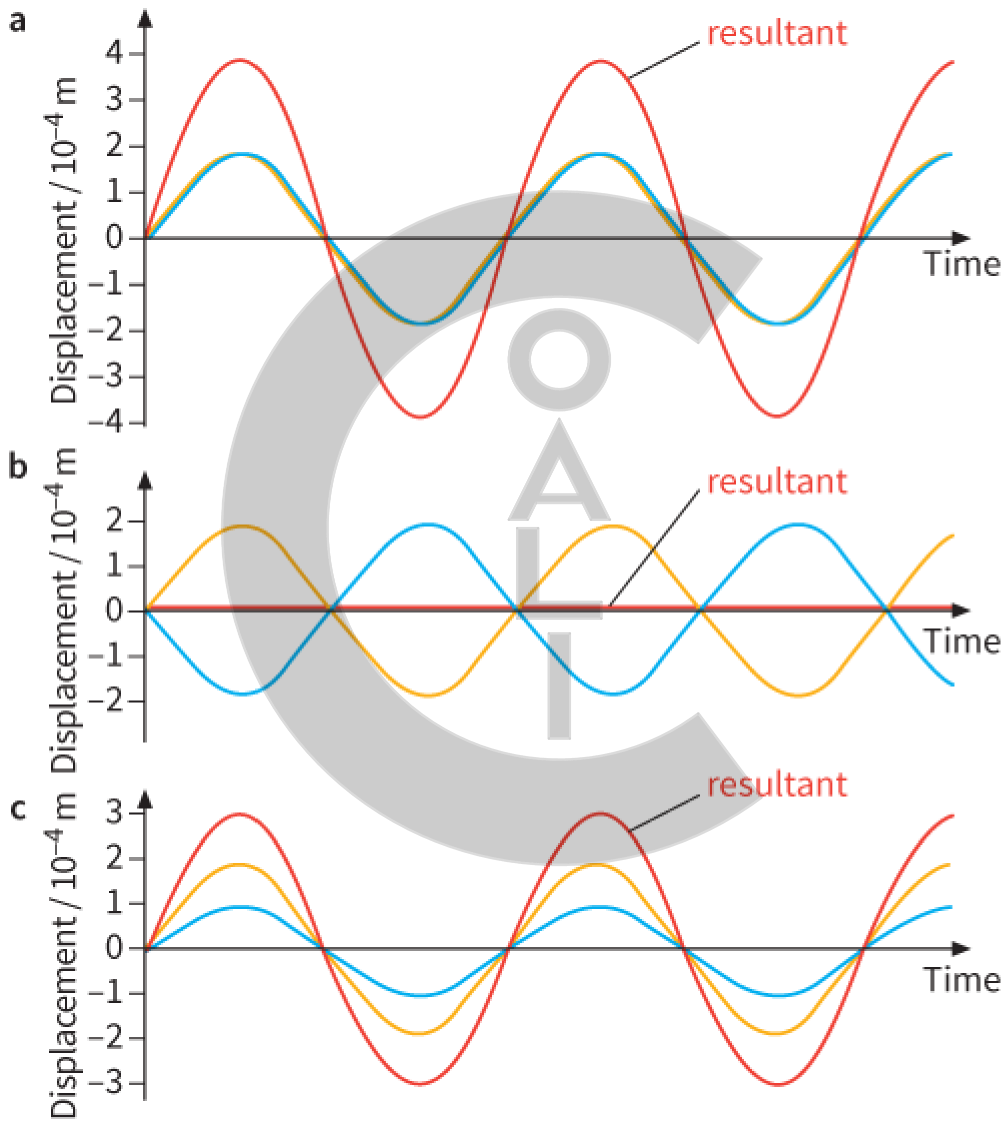


Figure 13.14: Adding waves by the principle of superposition. Blue and orange waves of the same amplitude may give **a** constructive or **b** destructive interference, depending on the phase difference between them. **c** Waves of different amplitudes can also interfere constructively.

Question

3 Explain why the two loudspeakers producing sounds of slightly different frequencies will not produce stable effects of interference.

How can we explain the interference pattern observed in a ripple tank (Practical Activity 13.2)? Look at Figure 13.15 and compare it to Figure 13.13. Figure 13.15 shows two sets of waves setting out from their sources. At a position such as A, ripples from the two sources arrive in phase with one another, and constructive interference occurs. At B, the two sets of ripples arrive in antiphase, and there is destructive interference. Although waves are arriving at B, the surface of the water remains approximately flat.

Whether the waves combine constructively or destructively at a point depends on the path difference of the waves from the two **coherent sources**. The **path difference** is defined as the extra distance travelled by one of the waves compared with the other.

At point A in Figure 13.15, the waves from the red source have travelled three whole wavelengths. The waves from the yellow source have travelled four whole wavelengths. The path difference between the two sets of waves is one wavelength. A path difference of one wavelength is equivalent to a phase difference of zero. This means that the two waves are in phase, so they interfere constructively.

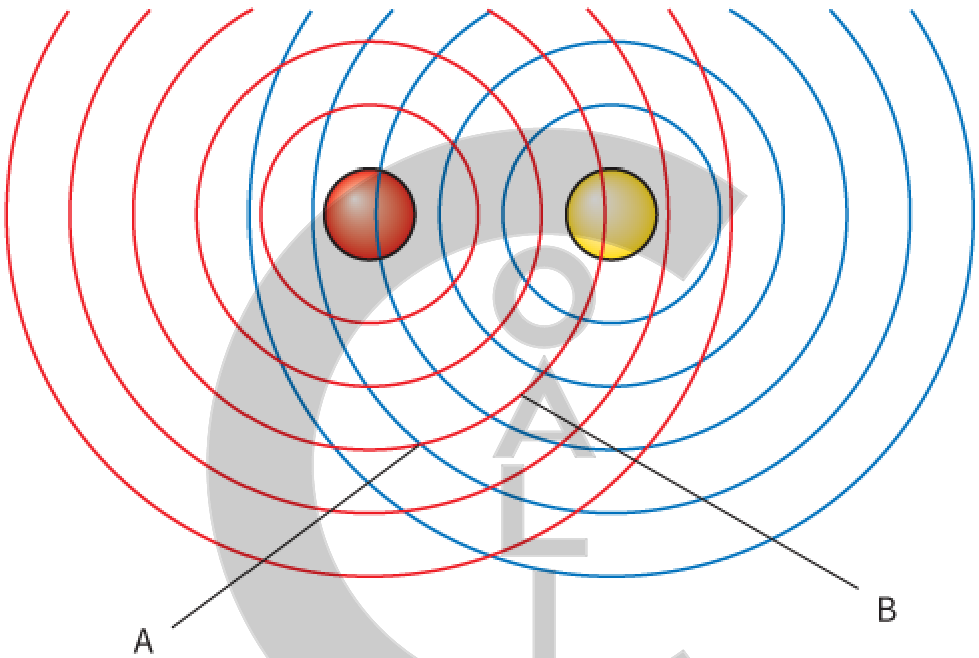


Figure 13.15: The result of interference depends on the path difference between the two waves.

Now think about destructive interference. At point B, the waves from the red source have travelled three wavelengths; the waves from the yellow source have travelled 2.5 wavelengths. The path difference between the two sets of waves is 0.5 wavelengths, which is equivalent to a phase difference of 180°. The waves interfere destructively because they are in antiphase. The conditions for constructive interference and destructive interference, in general, are outlined next. These conditions apply to **all** waves (water waves, light, microwaves, radio waves, sound and so on) that show interference effects. In the equations, n is an integer (any whole number, including zero).

For **constructive interference** the path difference is a whole number of wavelengths:

$$\text{path difference} = 0, \lambda, 2\lambda, 3\lambda, \text{ and so on}$$

or

$$\text{path difference} = n\lambda$$

For **destructive interference** the path difference is an odd number of half wavelengths:

$$\text{path difference} = \tfrac{1}{2}\lambda, 1\tfrac{1}{2}\lambda, 2\tfrac{1}{2}\lambda, \text{ and so on}$$

or

$$\text{path difference} = (n + \tfrac{1}{2})\lambda$$

PRACTICAL ACTIVITY 13.3

Interference of radiation

Interference of light

Here is one way to show the interference effects produced by light. A simple arrangement involves directing the light from a laser at a double-slit (Figure 13.16). The slits are two clear lines on a black slide, separated by a fraction of a millimetre. Where the light falls on the screen, a series of equally spaced dots of light are seen (see Figure 13.21). These bright dots are referred to as **interference maxima** or ‘fringes’, and they are regions where light waves from the two slits are arriving in phase with each other; in other words, there is constructive interference. The dark regions in between are the result of destructive interference.

If you carry out experiments using a laser, you should follow correct safety procedures. In particular, you should wear eye protection and avoid allowing the beam to enter your eye directly.

These bright and dark fringes are the equivalent of the loud and quiet regions that you detected if you investigated the interference pattern of sounds from the two loudspeakers described in Practical Activity 13.2. Bright fringes correspond to loud sound, and dark fringes to quiet sound or silence.

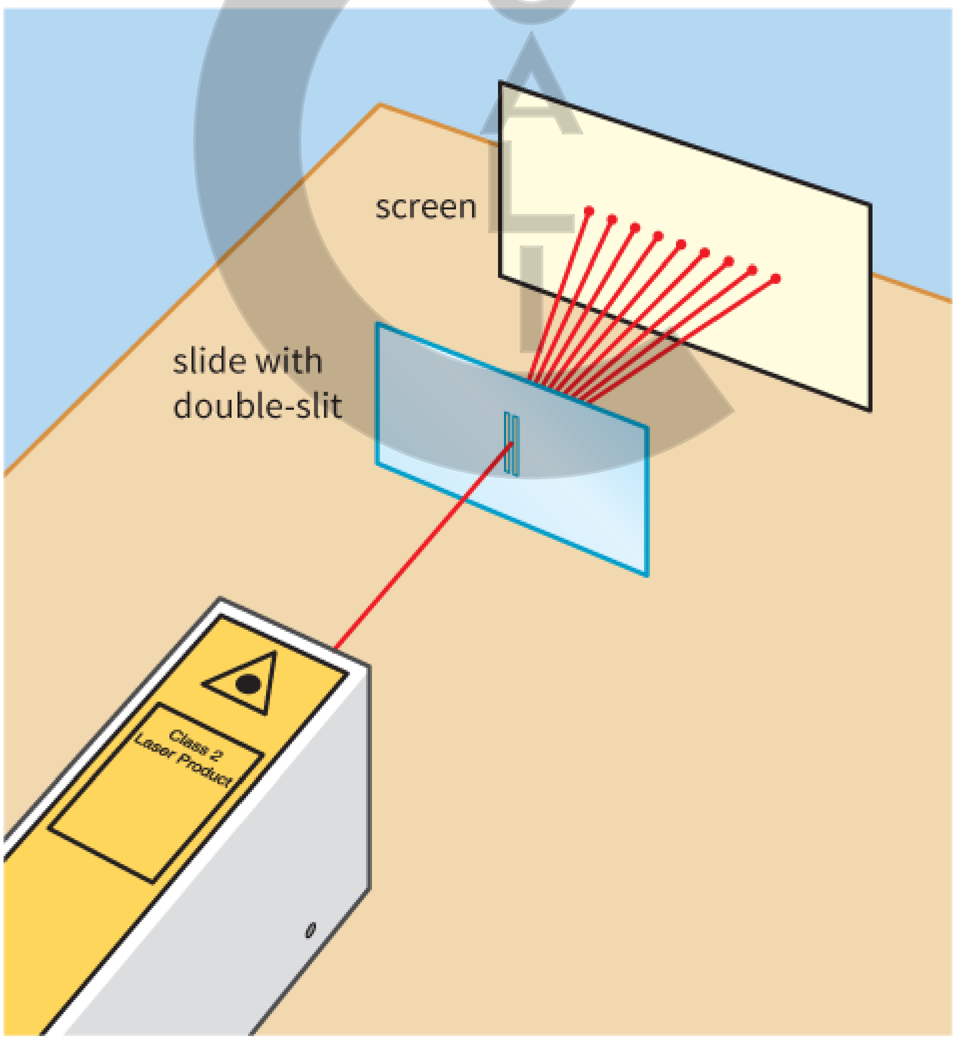


Figure 13.16: Laser light passing through the two slits show interference effects in the space beyond.

You can check that light is indeed reaching the screen from both slits as follows. Mark a point on the screen where there is a dark fringe. Now carefully cover up one of the slits so that light from the laser is only passing through one slit. You should find that the pattern of interference fringes disappears. Instead, a broad band of light appears across the screen. This broad band of light is the diffraction pattern produced by a single slit. The point that was dark is now light. Cover up the other slit instead, and you will see the same effect. You have now shown that light is arriving at the screen from both slits, but at some points (the dark fringes) the two beams of light cancel each other out.

You can achieve similar results with a bright light bulb rather than a laser, but a laser is much more convenient because the light is concentrated into a narrow, more intense beam. This famous experiment is called the Young double-slit experiment (discussed in more detail later in this chapter), although Thomas Young had no laser available to him when he first demonstrated it in 1801.

Interference of microwaves

Using 2.8 cm wavelength microwave equipment (Figure 13.17), you can observe an interference pattern. The microwave transmitter is directed towards the double gap in a metal barrier. The microwaves are diffracted at the two gaps so that they spread out into the region beyond, where they can be detected using the microwave probe (receiver). By moving the probe around, it is possible to detect regions of high intensity (constructive interference) and low intensity (destructive interference). The probe may be connected to a meter, or to an audio amplifier and loudspeaker to give an audible output.

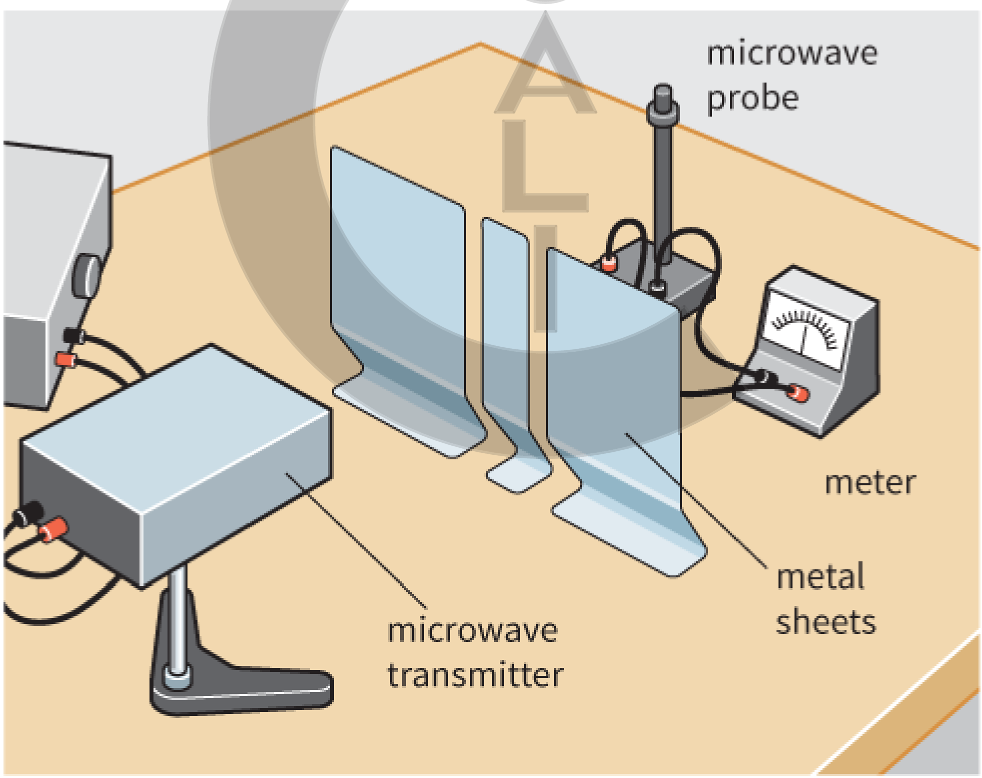


Figure 13.17: Microwaves also show interference effects.

Question

4 Look at the experimental arrangement shown in Figure 13.17. Suppose that the microwave probe is placed at a point of low intensity in the interference pattern.

Suggest what will happen if one of the gaps in the barrier is now blocked.

Coherence

We are surrounded by many types of wave – light, infrared radiation, radio waves, sound and so on. There are waves coming at us from all directions. So why do we not observe interference patterns all the time? Why do we need special equipment in a laboratory to observe these effects?

In fact, we can see interference of light occurring in everyday life. For example, you may have noticed haloes of light around street lamps or the Moon on a foggy night. You may have noticed light and dark bands of light if you look through fabric at a bright source of light. These are all examples of interference effects.

We usually need specially arranged conditions to produce interference effects that we can measure. Think about the demonstration with two loudspeakers. If they were connected to different signal generators with slightly different frequencies, the sound waves might start off in phase with one another, but they would soon go out of phase (Figure 13.18). We would hear loud, then soft, then loud again. The interference pattern would keep shifting around the room – there would be no stable interference pattern of loud and quiet regions.

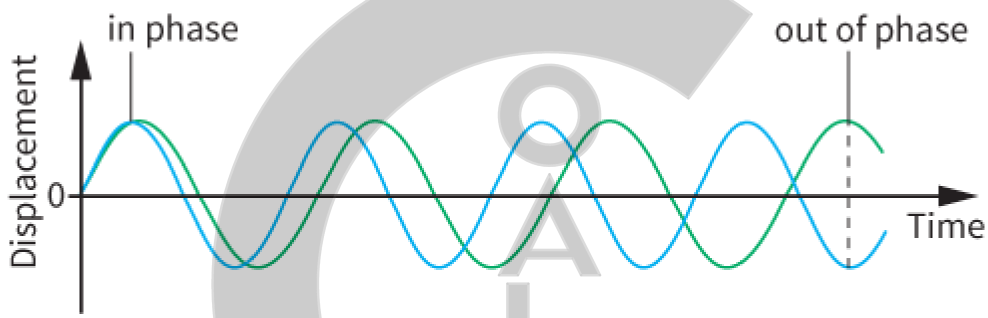


Figure 13.18: Waves of slightly different frequencies (and therefore wavelengths) move in and out of phase with one another.

By connecting the two loudspeakers to the **same** signal generator, we can be sure that the sound waves that they produce are constantly in phase with one another. We say that they act as two **coherent sources** of sound waves (coherent means sticking together). The sound waves from the loudspeakers has **coherence**. Coherent sources emit waves that have a **constant phase difference**. Note that the two waves can only have a constant phase difference if their frequency is the same and remains constant.

Now think about the laser experiment. Could we have used two lasers producing exactly the same frequency and hence wavelength of light? Figure 13.19a represents the light from a laser. We can think of it as being made up of many separate bursts of light. We cannot guarantee that these bursts from two lasers will always be in phase with one another.

This problem is overcome by using a single laser and dividing its light using the two slits (Figure 13.19b). The slits act as two coherent sources of light. They are constantly in phase with one another (or there is a constant phase difference between them).

If they were not coherent sources, the interference pattern would be constantly changing, far too fast for our eyes to detect. We would simply see a uniform band of light, without any definite bright and dark regions. From this you should be able to see that, in order to observe interference, we need two coherent sources of waves.

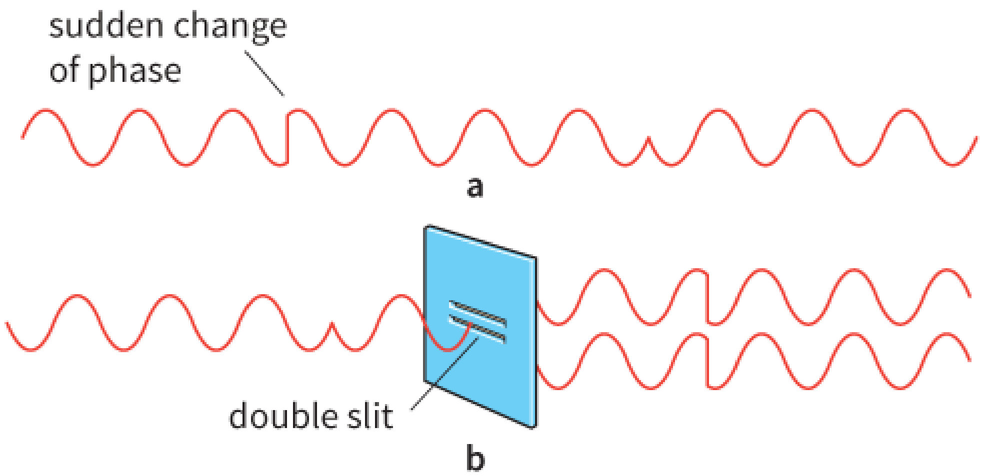


Figure 13.19: Waves must be coherent if they are to produce a clear interference pattern.

Question

5 Draw sketches of displacement against time to illustrate the following:

  a two waves having the same amplitude and in phase with one another

  b two waves having the same amplitude and with a phase difference of 90°

  c two waves initially in phase but with slightly different wavelengths.

13.4 The Young double-slit experiment

Now we will take a close look at a famous experiment that Thomas Young performed in 1801. He used this experiment to show the wave-nature of light. A beam of light is shone on a pair of parallel slits placed at right angles to the beam. Light diffracts and spreads outwards from each slit into the space beyond. The light from the two slits overlaps on a screen. An interference pattern of light and dark bands called 'fringes' is formed on the screen.

Explaining the experiment

In order to observe interference, we need two sets of waves. The sources of the waves must be coherent–the phase difference between the waves emitted at the sources must remain constant. This also means that the waves must have the same wavelength. Today, this is readily achieved by passing a single beam of laser light through the two slits. A laser produces intense coherent light. As the light passes through the slits, it is diffracted so that it spreads out into the space beyond (Figure 13.20). Now we have two overlapping sets of waves, and the pattern of fringes on the screen shows us the result of their interference (Figure 13.21).

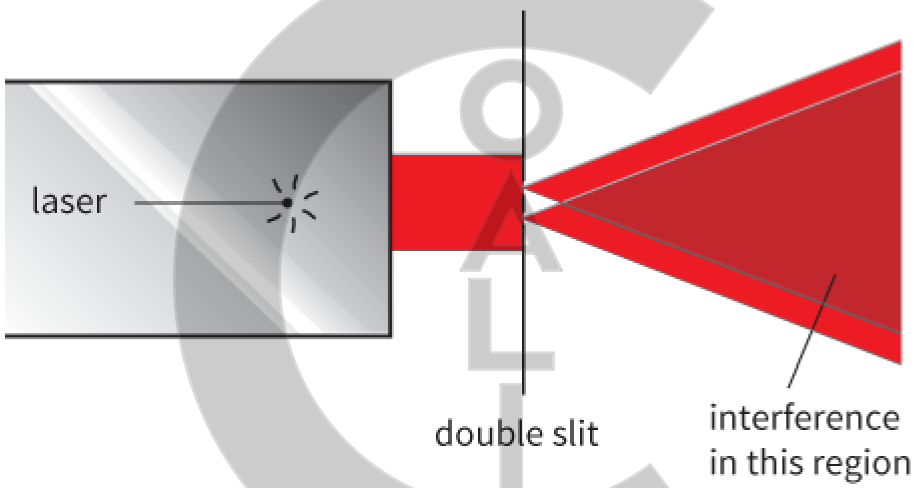


Figure 13.20: Interference occurs where diffracted beams from the two slits overlap. How does this pattern arise? We will consider three points on the screen (Figure 13.22), and explain what we would expect to observe at each.

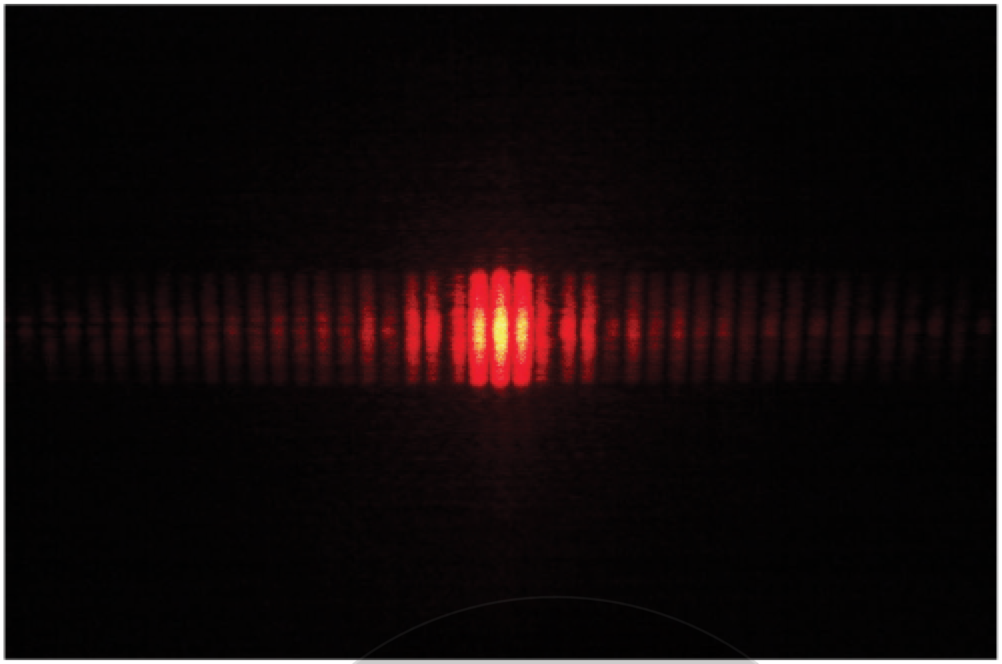


Figure 13.21: Interference fringes obtained using a laser and a double-slit.

Point A

This point is directly opposite the midpoint of the slits. Two rays of light arrive at A, one from slit 1 and the other from slit 2. Point A is equidistant from the two slits, and so the two rays of light have travelled the same distance. The path difference between the two rays of light is zero. If we assume that they were in phase (in step) with each other when they left the slits, then they will be in phase when they arrive at A. Hence they will interfere constructively, and we will observe a bright fringe at A.

Point B

This point is slightly to the side of point A, and is the midpoint of the first dark fringe. Again, two rays of light arrive at B, one from each slit. The light from slit 1 has to travel slightly further than the light from slit 2, and so the two rays are no longer in step. Since point B is at the midpoint of the dark fringe, the two rays must be in antiphase (phase difference of 180°). The path difference between the two rays of light must be half a wavelength, and so the two rays interfere destructively.

Point C

This point is the midpoint of the next bright fringe, with AB = BC. Again, ray 1 has travelled further than ray 2; this time, it has travelled an extra distance equal to a whole wavelength λ. The path difference between the rays of light is now a whole wavelength. The two rays are in phase at the screen. They interfere constructively, and we see a bright fringe.

The complete interference pattern (Figure 13.21) can be explained in this way.

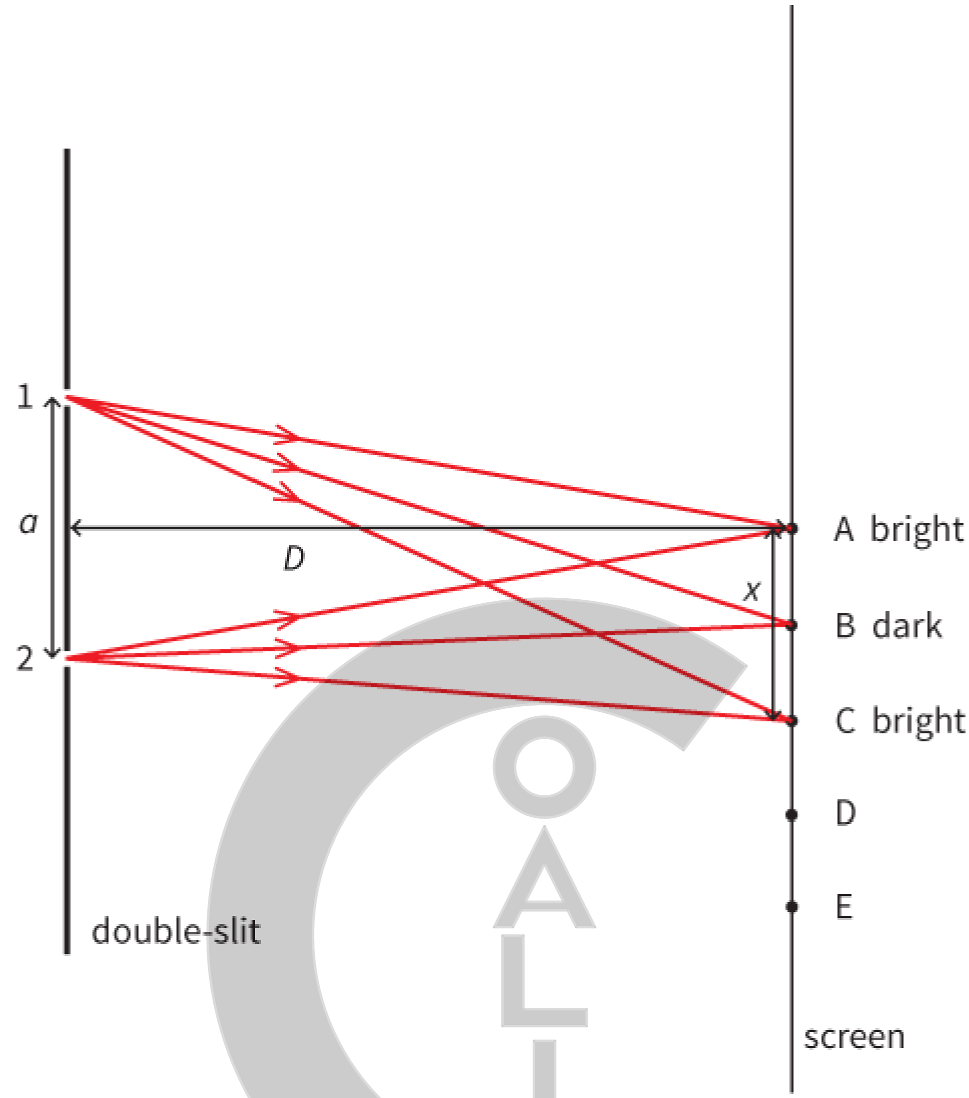


Figure 13.22: The type of interference, and hence whether a bright or a dark fringe is seen on the screen, depends on the path difference between the rays of light arriving at the screen from the double-slit.

Question

6 Consider points D and E on the screen in Figure 13.22, where BC = CD = DE. State and explain what you would expect to observe at D and E.

Determining wavelength λ

The double-slit experiment can be used to determine the wavelength λ of monochromatic light. The following three quantities have to be measured:

- **Slit separation *a*** – This is the distance between the centres of the slits, which is the distance between slits 1 and 2 in Figure 13.22.
- **Fringe separation *x*** – This is the distance between the centres of adjacent bright (or dark) fringes, which is the distance AC in Figure 13.22.
- **Slit-to-screen distance *D*** – This is the distance from the midpoint of the slits to the central fringe on the screen.

Once these three quantities have been determined, the wavelength λ of the light can be found using the relationship:

$$\lambda = \frac{ax}{D}$$

KEY EQUATION

The double-slit equation:

$$\lambda = \frac{ax}{D}$$

where λ is the wavelength of the monochromatic light incident normally at the double-slit. *a* is the separation between the centres of the slits, *x* is the separation between the centres of adjacent bright (or dark) fringes and *D* is distance between the slits and the screen.

WORKED EXAMPLE

1 In a double-slit experiment using light from a helium–neon laser, a student obtained the following results:

width of 10 fringes 10*x* = 1.5 cm

separation of slits *a* = 1.0 mm

slit-to-screen distance *D* = 2.40 m

Calculate the wavelength of the light in nm.

Step 1 Work out the fringe separation *x* in metres (m):

$$\text{fringe sepration, } x = \frac{1.5\times10^{2}}{10}$$
$$= 1.5\times10^{-3}\ \text{m}$$

Step 2 Substitute the values of *a*, *x* and *D* (all in metres) into the equation and then calculate the wavelength λ:

$$\lambda = \frac{ax}{D}$$
$$= \frac{1.0\times10^{-3}\times1.5\times10^{-3}}{2.40}$$
$$= 6.25\times10^{-7}\ \text{m} \approx 6.30\times10^{-7}\ \text{m}$$

1 nm = 10^{-9} m

Therefore:

λ = 630 nm

Question

7 The student in Worked example 1 moved the screen to a distance of 4.8 m from the slits. Determine the fringe separation *x* now.

PRACTICAL ACTIVITY 13.4

Using Young's slits to determine λ

The Young double-slit experiment can be used to determine the wavelength λ of monochromatic light. Here, we look at a number of practical features of the experiment and consider how the percentage uncertainty in the value of λ can be reduced.

One way to carry out the double-slit experiment is shown in Figure 13.23. Here, a white light source is used, rather than a laser. A monochromatic filter allows only one wavelength of light to pass through. A single slit diffracts the light. This diffracted light arrives in phase at the double slit, which ensures that the two parts of the double slit behave as coherent sources of light. The double slit is placed a centimetre or two beyond the single slit, and the fringes are observed on a screen a metre or so away. The experiment has to be carried out in a darkened room, as the intensity of the bright fringes is low – making them hard to see.

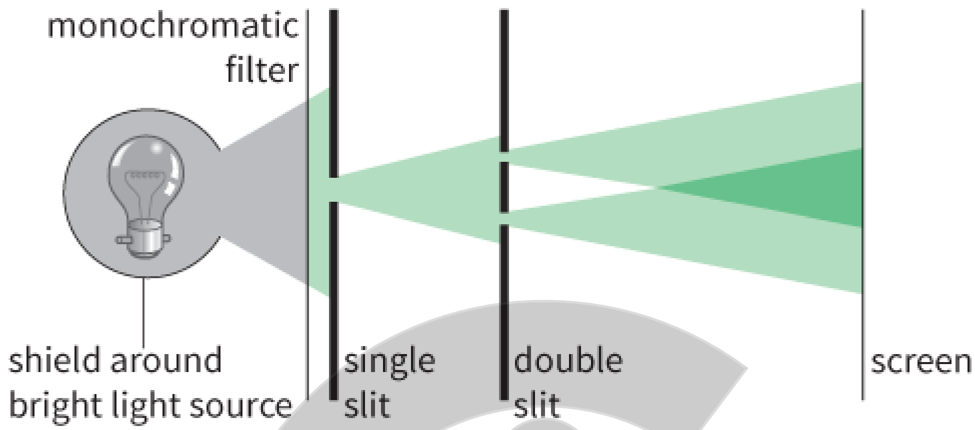


Figure 13.23: Arrangement for seeing fringes using a white light source.

There are three important factors involved in the way the equipment is set up:

- All slits are a fraction of a millimetre in width. Since the wavelength of light is less than a micrometre (10^{-6} m), this gives a small amount of diffraction in the space beyond. If the slits were narrower, the intensity of the light would be too low for visible fringes to be achieved.
- The double slits are about a millimetre apart. If they were much further apart, the fringes would be too close together to be distinguishable.
- The screen is about a metre from the slits. The fringes produced are clearly separated without being too dim.

Measuring *a*, *x* and *D*

Measuring slit separation *a*: a travelling microscope is suitable for measuring *a*. It is difficult to judge the position of the centre of a slit. If the slits are the same width, the separation of their left-hand edges is the same as the separation of their centres.

Measuring fringe width *x*: it is best to measure across several fringes (say, ten) and then to calculate the average separation later. A 30 cm ruler or a travelling microscope can be used.

Measuring the slit-to-screen distance *D*: this can be measured using a metre rule or a tape measure.

Reducing percentage errors

Why use a laser rather than white light? With a laser, the light beam is more concentrated, and the initial single slit is not necessary. The greater intensity of the beam means that the screen can be further from the slits, so that the fringes are further apart. This reduces the percentage uncertainties in measurements of *x* and *D*. Consequently, the overall percentage uncertainty in the calculated value for the wavelength λ will be smaller.

A laser has a second advantage. The light from a laser is monochromatic; that is, it consists of a single wavelength. This makes the fringes very clear, and they are present in large numbers across the screen. With white light, a range of wavelengths is present. Different wavelengths form fringes at different points across the screen, smearing them out so that they are not as clear.

Using white light with no filter results in a central fringe that is white (because all wavelengths are in phase here), but the other fringes show coloured effects, as the different wavelengths interfere constructively at different points. In addition, only a few fringes are visible in the interference pattern.

Questions

8 Use the equation $\lambda = \frac{ax}{D}$ to explain the following observations:

a With the slits closer together, the fringes are further apart.

b Interference fringes for blue light are closer together than for red light.

c In an experiment to measure the wavelength of light, it is desirable to have the screen as far from the slits as possible.

9 Yellow light from a sodium source is used in the double-slit experiment. This yellow light has wavelength 589 nm. The slit separation is 0.20 mm, and the screen is placed 1.20 m from the slits.

Calculate the separation between adjacent bright fringes formed on the screen.

10 In a double-slit experiment, filters were placed in front of a white light source to investigate the effect of changing the wavelength of the light. At first, a red filter was used instead (λ = 600 nm) and the fringe separation was found to be 2.4 mm. A blue filter was then used instead (λ = 450 nm).

Calculate the fringe separation with the blue filter.

13.5 Diffraction gratings

A **transmission** diffraction grating is similar to the slide used in the double-slit experiment, but with many more slits than just two. It consists of a large number of equally spaced lines ruled on a glass or plastic slide. Each line is capable of diffracting the incident light. There may be as many as 10 000 lines per centimetre. When light is shone through this grating, a pattern of interference fringes is seen.

A **reflection** diffraction grating consists of lines made on a reflecting surface so that light is both reflected and diffracted by the grating. The shiny surface of a CD (compact disc), or a DVD (digital versatile disc), is an everyday example of a reflection diffraction grating. Hold a CD in your hand so that you are looking at the reflection of light from a lamp. You will observe coloured bands (Figure 13.24). A CD has thousands of equally spaced lines of microscopic pits on its surface; these carry the digital information. It is the diffraction from these lines that produces the coloured bands of light from the surface of the CD.

Figure 13.24: A CD acts as a reflection diffraction grating. White light is reflected and diffracted at its surface, producing a display of spectral colours.

Observing diffraction with a transmission grating

In Figure 13.25, monochromatic light from a laser is incident normally on a transmission diffraction grating. In the space beyond, interference fringes are formed. These can be observed on a screen, as with the double slit. However, it is usual to measure the angle θ at which they are formed, rather than measuring their separation. With double slits, the fringes are equally spaced and the angles are very small. With a diffraction grating, the angles are much greater and the fringes are not equally spaced.

The bright fringes are also referred to as **maxima**. The central fringe is called the zeroth-order maximum, the next fringe is the first-order maximum, and so on. The pattern is symmetrical, so there are two first-order maxima, two second-order maxima, and so on.

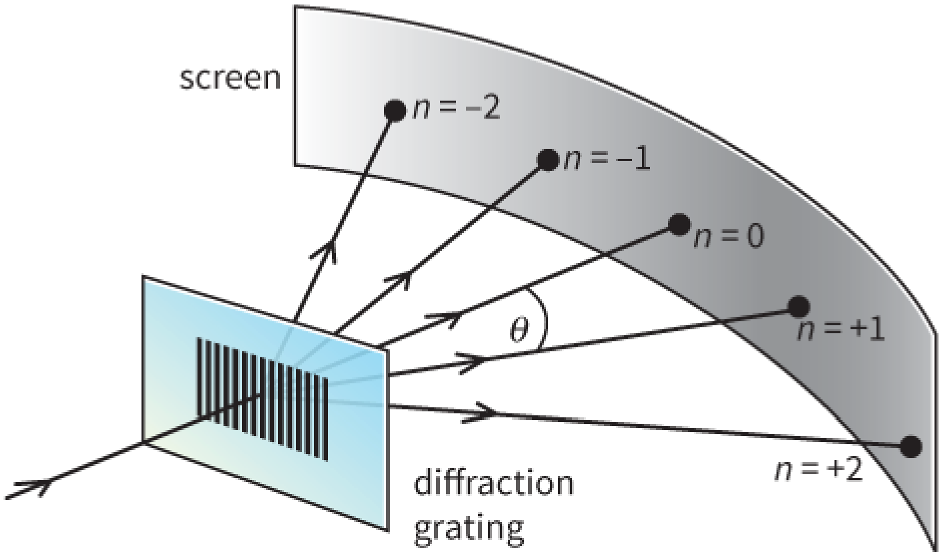


Figure 13.25: A laser beam passing through a diffraction grating produces a symmetrical pattern of maxima on a screen.

Explaining the experiment

The principle is the same as for the double-slit experiment, but here we have light passing through many slits. As it passes through each slit, it diffracts into the space beyond. So now we have many overlapping beams of light, and these interfere with one another.

There is a bright fringe, the zeroth-order maximum, in the straight-through direction ($\theta = 0$). This is because all of the rays here are travelling parallel to one another and in phase, so the interference is constructive (Figure 13.26a).

Imagine if you could look through the diffraction grating at the source of light. Your eye would be focused on the light source, which is far away. All the rays with $\theta = 0$ come together at the back of your eye, where an image is formed. It is here that interference occurs.

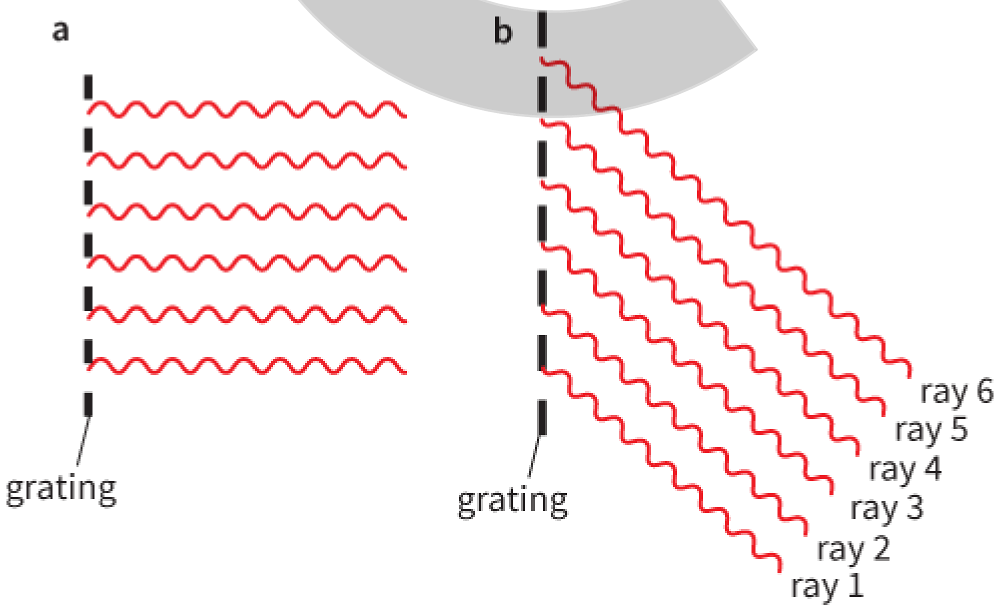


Figure 13.26: a Waves from each slit are in phase in the straight-through direction. **b** In the direction of the first-order maximum, the waves are in phase, but each one has travelled one wavelength further than the one

below it.

The first-order maximum forms in a specific direction as follows. Diffraction occurs at all of the slits. Rays of light emerge from all of the slits to form a bright fringe – all the rays must be in phase. In the direction of the first-order maximum, ray 1 has travelled the smallest distance (Figure 13.26b). Ray 2 has travelled an extra distance equal to one whole wavelength and is therefore in phase with ray 1. The path difference between ray 1 and ray 2 is equal to one wavelength λ. Ray 3 has travelled two extra wavelengths and is in phase with rays 1 and 2. In fact, the rays from all of the slits are in step in this direction, and a bright fringe results.

Question

11 Explain how the second-order maximum arises in terms of path difference.

Determining wavelength λ with a diffraction grating

By measuring the angles at which the maxima occur, we can determine the wavelength λ of the incident monochromatic light. The wavelength λ is related to the angle θ by the equation:

$$d\sin\theta = n\lambda$$

KEY EQUATION

$$d\sin\theta = n\lambda$$

WORKED EXAMPLE

2 Monochromatic light is incident normally on a diffraction grating having 300 lines mm^{-1}. The angle θ between the zeroth- and first-order maxima is measured to be 10.0°.

Calculate the wavelength of the incident light.

Step 1 Calculate the distance between the adjacent lines (grating spacing) *d*. Since there are 3000 lines mm^{-1}, *d* must be:

$$d = \frac{1\text{ mm}}{300} = 3.33\times10^{-3}\text{ mm} = 3.33\times10^{-6}\text{ m}$$

Step 2 Rearrange the equation $d\sin\theta = n\lambda$ and substitute values:

$$\theta = 10.0°, n = 1 \text{ (first order)}$$

$$\lambda = \frac{d\sin\theta}{n} = \frac{3.33\times10^{-6}\times\sin 10.0°}{1} = 5.8\times10^{-7}\text{ m}$$

This is the same as 580 nm. (1 nm = 10^{-9} m.)

where *d* is the separation between adjacent lines of the grating, θ is the angle for the n^{th}-order maximum and λ is the wavelength of the monochromatic light incident normally at the diffraction grating. *n* is known as the **order** of the maximum; *n* can only have integer values 0, 1, 2, 3 and so on. The distance *d* is also known as the grating element or grating spacing.

Worked example 2 shows how you can determine λ.

Questions

12 a For the case described in Worked example 2, with λ = 580 nm, calculate the angle θ for the second-order maximum.

b Repeat the calculation of θ for n = 3, 4, and so on. Determine how many maxima can be seen. Explain your answer.

13 Consider the equation $d\sin\theta = n\lambda$. State and explain how the interference pattern would change when:

a the wavelength of the incident light is increased for the same grating

b the grating is changed for one with more lines per cm for the same incident light.

14 A student is trying to make an accurate measurement of the wavelength of green light from a mercury lamp. The wavelength λ of this light is 546 nm. Using a double-slit of separation 0.50 mm, the student can see 10 clear bright fringes on a screen at a distance of 0.80 m from the slits. The student can measure their overall width to within ±1 mm using a ruler.

The student then tries an alternative experiment using a diffraction grating with 3000 lines cm^{-1}. The angle between the two second-order maxima can be measured to within ±0.1°.

a Determine the width of the 10 fringes that the student can measure in the first experiment.

b Determine the angle of the second-order maximum that the student can measure in the second experiment.

c Based on your answers to parts *a* and *b*, suggest which experiment you think will give the more accurate value of λ.

Diffracting white light

A diffraction grating can be used to split white light up into its component colours. This splitting of light is known as **dispersion**, shown in Figure 13.27. A beam of white light is shone onto the grating. A zeroth-order, white maximum is observed at θ = 0°, because all waves of each colour are in phase in this direction.

On either side, a series of spectra appear, with violet closest to the centre and red furthest away. We can see why different wavelengths have their maxima at different angles if we rearrange the equation $d\sin\theta = n\lambda$ to give:

$$\sin\theta = \frac{n\lambda}{d}$$

From this it follows that the greater the wavelength λ, the greater the value of $\sin\theta$ and hence the greater the angle θ. Red light is at the long wavelength end of the visible spectrum, and so it appears at the greatest angle.

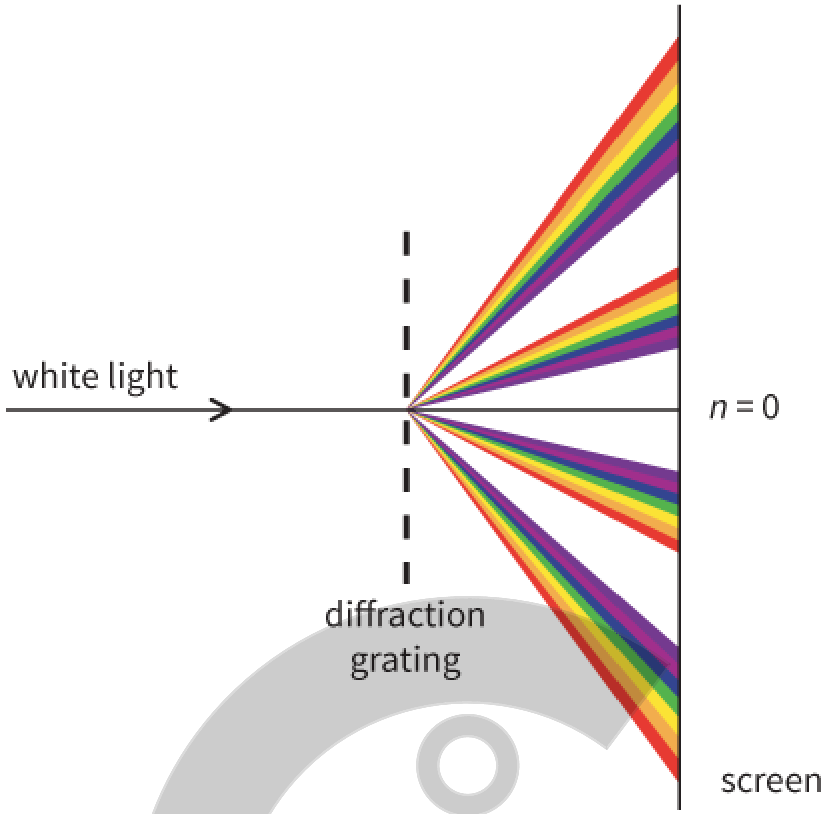


Figure 13.27: A diffraction grating is a simple way of separating white light into its constituent wavelengths.

PRACTICAL ACTIVITY 13.5

Diffraction gratings versus double-slit

It is worth comparing the use of a diffraction grating to determine wavelength with the Young two-slit experiment.

- With a diffraction grating, the maxima are very **sharp**.
- With a diffraction grating, the maxima are also very **bright**. This is because rather than contributions from only two slits, there are contributions from a thousand or more slits.
- With double-slit, there may be a large uncertainty in the measurement of the slit separation **a**. The fringes are close together, so their separation may also be measured imprecisely.
- With a diffraction grating the maxima are widely separated, the angle θ can be measured to a high degree of precision. So, an experiment with a diffraction grating can be expected to give a value for the wavelength to a much higher degree of precision than a simple double-slit arrangement.

Question

15 White light is incident normally on a diffraction grating with a slit-separation d of 2.00×10^{-6} m. The visible spectrum has wavelengths between 400 nm and 700 nm.

a Calculate the angle between the red and violet ends of the first-order spectrum.

b Explain why the second- and third-order spectra overlap.

REFLECTION

Make a short list of everyday items that would diffract sound, then do the same for light.

Summarise two experiments for your fellow learners for determining the wavelength of visible light. What did you learn about yourself as you worked on this summary?

SUMMARY

The principle of superposition states that when two or more waves meet at a point, the resultant displacement is the algebraic sum of the displacements of the individual waves.

When waves pass through a slit, they may be diffracted so that they spread out into the space beyond. The diffraction effect is greatest when the wavelength of the waves is similar to the width of the gap.

Interference is the superposition of two or more waves from coherent sources.

Two sources are coherent when they emit waves that have a **constant phase difference**. (This can only happen if the waves have the same frequency or wavelength.)

Path difference is the extra distance travelled by one of the waves compared with the other.

For **constructive interference**, the path difference is a whole number of wavelengths $(0, \lambda, 2\lambda, 3\lambda, \text{ and so on})$ or simply path difference = $n\lambda$.

For constructive interference, the waves are always in phase (phase difference = 0°).

For **destructive interference**, the path difference is an odd number of half wavelengths $(\frac{1}{2}\lambda, 1\frac{1}{2}\lambda, 2\frac{1}{2}\lambda \text{ and so on})$ or simply $\text{path difference} = (n + \frac{1}{2})\lambda$.

For destructive interference, the waves are completely out of phase (e.g. phase difference = 180°).

When light passes through a double-slit, it is diffracted at each slit and an interference pattern of equally spaced light and dark fringes is observed. This can be used to determine the wavelength of light using the equation:

$$\lambda = \frac{ax}{D}$$

This equation can be used for all waves, including sound and microwaves.

A diffraction grating diffracts light at its many slits or lines. The diffracted light interferes in the space beyond the grating.

The equation for a diffraction grating is:

$$d \sin\theta = n\lambda$$

where d is the distance between adjacent lines of the grating θ is the angle between the zeroth order and n^{th}-order maximum, and λ is the wavelength of the light incident normally at the grating.

EXAM-STYLE QUESTIONS

1 Rays of light from two coherent sources produces constructive interference.

Which of the following **cannot** be the phase difference between these two rays? **[1]**

A 0°

B 270°

C 360°

D 720°

2 **a** Copy the waves shown in the diagram onto a sheet of graph paper and use the principle of superposition to sketch the resultant wave. **[2]**

b **Compare** the wavelength of the resultant wave with that of the component waves. **[1]**

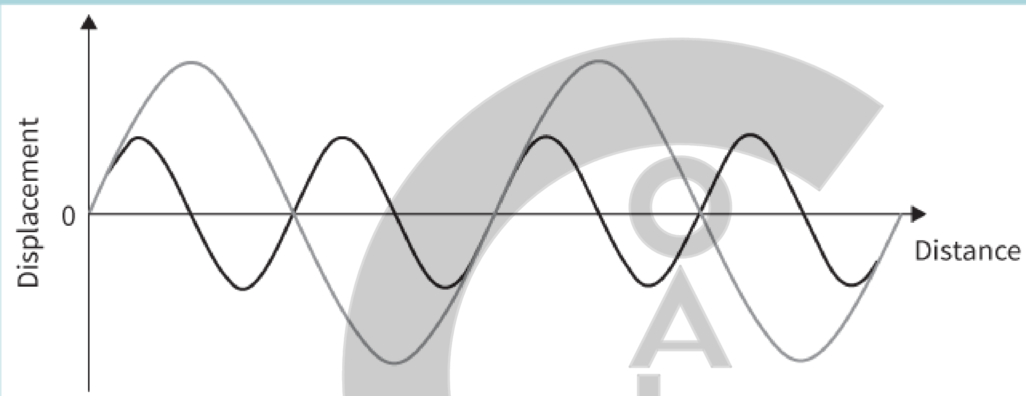


Figure 13.28

[Total: 3]

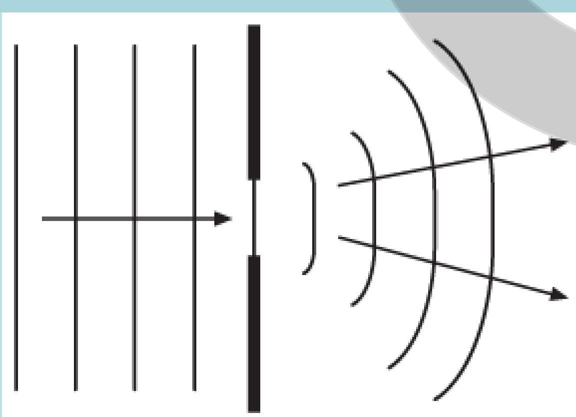


Figure 13.29

3 State how the diffracted pattern will change when:

a the wavelength of the incident wave is increased **[1]**

b the wavelength of the incident wave is decreased. **[1]**

[Total: 2]

4 Explain why, in remote mountainous regions, such as the Hindu Kush, radio signals **[2]**

from terrestrial transmitters can be received, but television reception can only be received from satellite transmissions.

5 A constant frequency signal from a signal generator is fed to two loudspeakers placed 1.5 m apart. A student, who is 8.0 m away from the loudspeakers, walks across in a line parallel to the line between the loudspeakers. The student measures the distance between successive spots of loudness to be 1.2 m.

Calculate:

a the wavelength of the sound **[2]**

b the frequency of the sound (assume the speed of sound is 330 m s^{-1}) **[2]**

[Total: 4]

6 Two signal generators feed signals with slightly different frequencies to two separate loudspeakers. Suggest why a sound of continuously rising and falling loudness is heard. **[3]**

7 One of the spectral lines from a hydrogen discharge lamp has wavelength 656 nm. This light is incident normally at a diffraction grating with 5000 lines cm^{-1}.

Calculate the angles for the first- and second-order maxima for this light. **[5]**

8 **a** Explain what is meant by the term superposition. **[2]**

b In a double-slit experiment, yellow light of wavelength 590 nm from a sodium discharge tube is used. A student sets up a screen 1.8 m from the double-slit. The distance between 12 bright fringes is measured to be 16.8 mm.

Calculate the separation of the slits. **[3]**

c Describe the effect of:

i using slits of narrower width, but with the same separation **[2]**

ii using slits with a smaller separation, but of the same width. **[2]**

[Total: 9]

9 **a** A laser light is described as producing light that is both highly coherent and highly monochromatic.

Explain what is meant by the terms **coherent** and monochromatic. **[2]**

b This diagram shows the experimental setup (left) used to analyse the spectrum of a sodium discharge lamp with a diffraction grating with 500 lines mm^{-1}, and the spectral lines observed (right) in the developed photographic film.

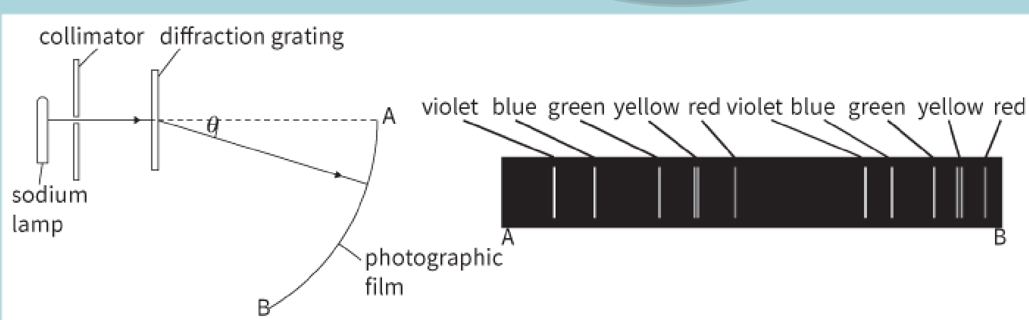


Figure 13.30

i Explain why two spectra are observed. **[2]**

ii Describe **two** differences between these two spectra. **[2]**

iii The green maximum near end A is at an angle θ of 19.5°.

Calculate the wavelength of the green light. **[3]**

iv Calculate the angle produced by the second green line. [2]

[Total: 11]

10 a Explain what is meant by **destructive interference**. [2]

b A student sets up an experiment to investigate the interference pattern formed by microwaves of wavelength 1.5 cm. The apparatus is set up as in Figure 13.17. The distance between the centres of the two slits is 12.5 cm. The detector is centrally placed 1.2 m from the metal plates where it detects a maximum. The student moves the detector 450 cm across the bench parallel to the plates.

Calculate how many maxima the detector will be moved through. [3]

c Calculate the frequency of these microwaves. [2]

[Total: 7]

11 a Explain what is meant by the **diffraction** of a wave. [2]

b This diagram shows waves, in a ripple tank, spreading out from two slits.

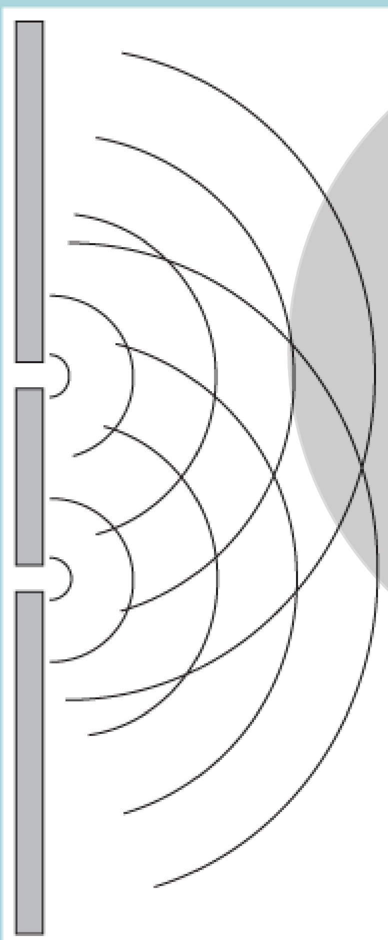


Figure 13.31

Copy this diagram. On your diagram, sketch:

i a line showing points along the central maximum–label this line **0** [1]

ii a line showing the points along first maximum–label this line **1** [1]

iii a line showing points along one of the first minima–label this line **min**. [1]

c The centres of the slits are 12 cm apart. At a distance of 60 cm from the barrier, the first maxima are 18 cm either side of the central maximum. [3]

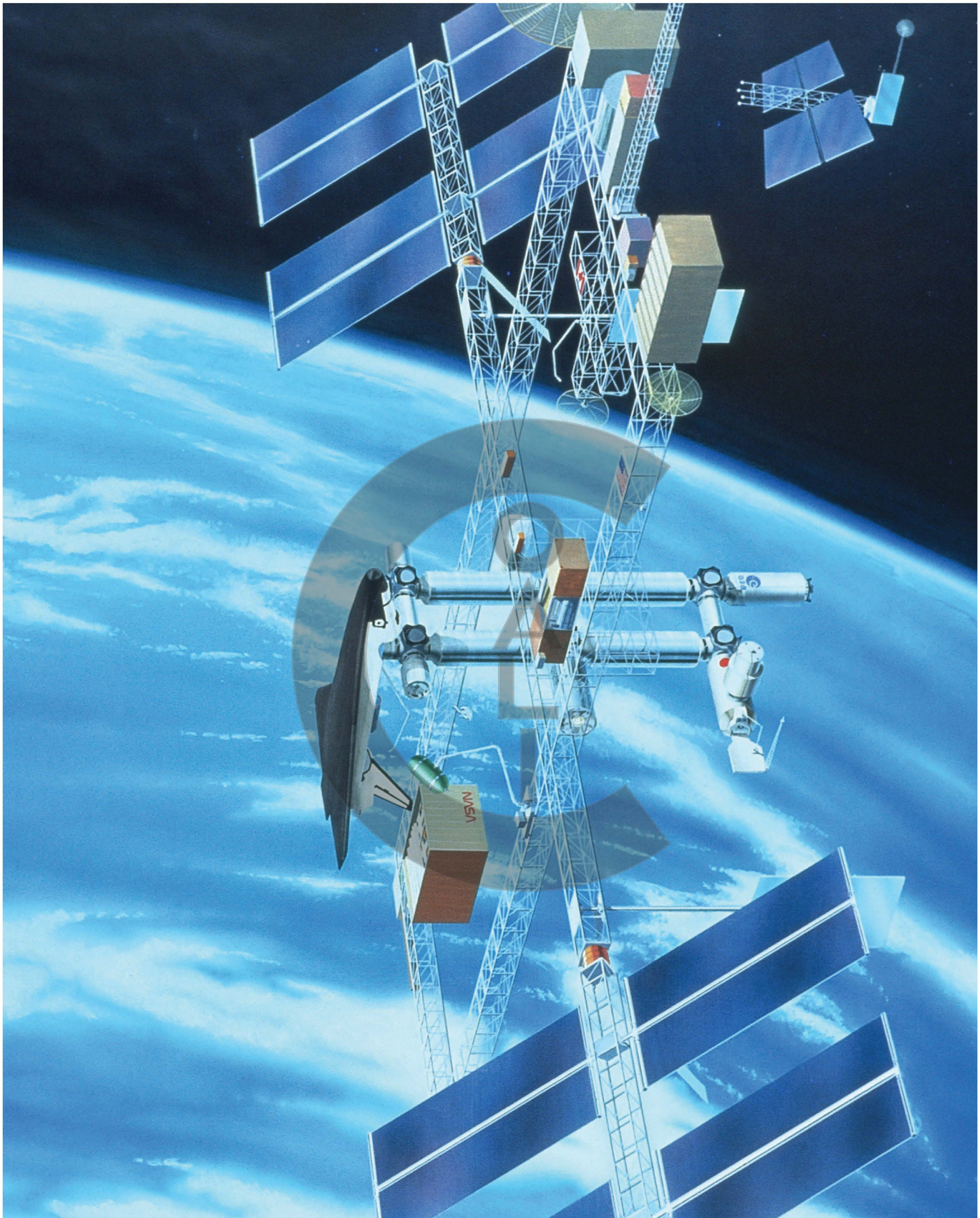
Calculate the wavelength of the waves. (You may assume that the double-slit equation developed for light is applicable to ripples.)

[Total: 8]

SELF-EVALUATION CHECKLIST

After studying this chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the principle of superposition	13.1			
understand diffraction, interference, path difference and coherence	13.2, 13.3			
understand the conditions for constructive and destructive interference	13.3			
understand experiments involving two coherent sources	13.3			
recall and use $\lambda = \frac{ax}{D}$ for double-slit interference using light	13.4			
recall and use $d \sin\theta = n\lambda$ for a diffraction grating	13.5			
use a diffraction grating to determine the wavelength of light.	13.5			



〉Chapter 14

Stationary waves

LEARNING INTENTIONS

In this chapter you will learn how to:

- explain the formation of stationary waves using graphical methods
- understand experiments to demonstrate stationary waves using microwaves, stretched strings and air columns
- identify nodes and antinodes
- determine the wavelength of sound using stationary waves.

BEFORE YOU START

- Write down the wave equation and use it to estimate the wavelength of ripples on the surface of a pond.
- Write down a few notes about the principle of superposition of waves. This will help you to understand how stationary (standing) waves are formed.

THE BRIDGE THAT BROKE

Figure 14.1a shows the Normandy Bridge under construction in France. When designing bridges, engineers must take into account the possibility of the wind causing a build-up of stationary waves, which may lead the bridge to oscillate violently. Famously, this happened in October 1940 to the Tacoma Narrows Bridge in Washington State, USA. High winds caused the bridge to vibrate with increasing amplitude until it fell apart (Figure 14.1b).

Did you know that the Tacoma Narrows Bridge fell apart because its natural frequency of oscillation matched the thumping frequency of the swirling wind? Do a web search for a videoclip of this momentous event.

Figure 14.1: **a** A suspension bridge under construction. **b** One that failed – the Tacoma Narrows Bridge.

14.1 From moving to stationary

The waves we have considered so far in Chapters 12 and 13 have been **progressive waves**; they start from a source and travel outwards, transferring energy from one place to another. A second important class of waves is **stationary waves** (**standing waves**). These can be observed as follows. Use a long spring or a plastic toy spring. A long rope or piece of rubber tubing will also do. Lay it on the floor and fix one end firmly. Move the other end from side to side so that transverse waves travel along the length of the spring and reflect off the fixed end (Figure 14.2). If you adjust the frequency of the shaking, you should be able to achieve a stable pattern like one of those shown in Figure 14.3. Alter the frequency in order to achieve one of the other patterns.

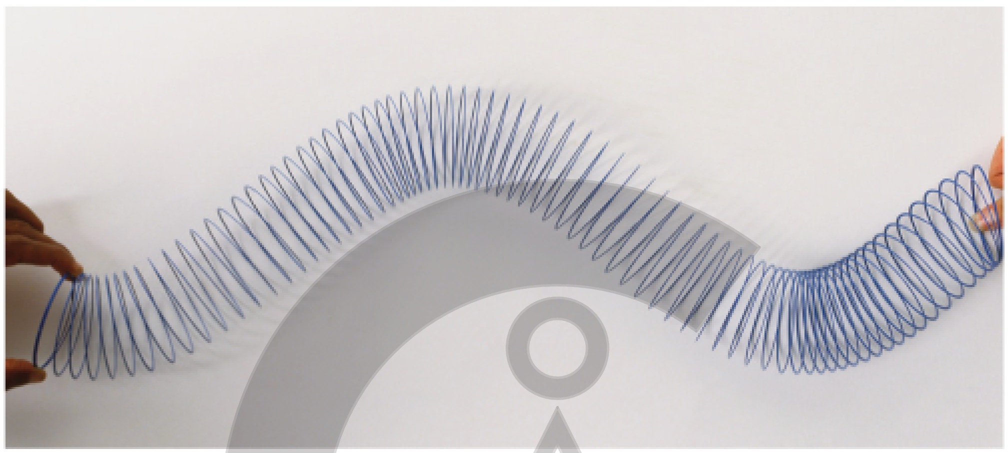


Figure 14.2: A toy spring is used to generate a stationary wave pattern.

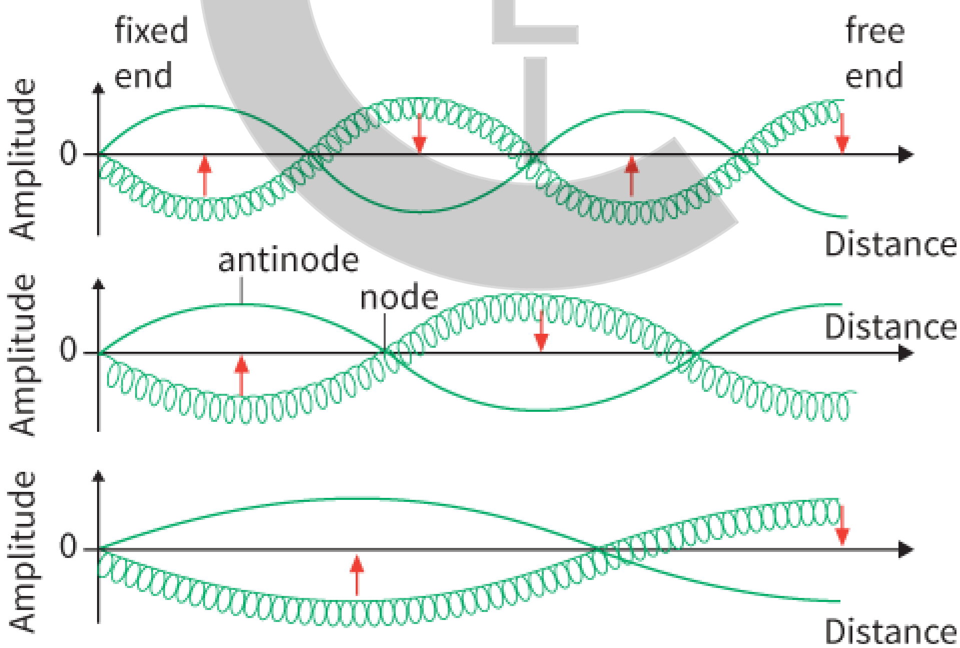


Figure 14.3: Different stationary wave patterns are possible, depending on the frequency of vibration.

You should notice that you have to move the end of the spring with just the right frequency to get one of these interesting patterns. The pattern disappears when the frequency of the shaking of the free end of the spring is slightly increased or decreased.

COALI

14.2 Nodes and antinodes

What you have observed is a stationary wave on the long spring. There are points along the spring that remain (almost) motionless while points on either side are oscillating with the greatest amplitude. The points that do not move are called the **nodes** and the points where the spring oscillates with maximum amplitude are called the **antinodes**. At the same time, it is clear that the wave profile is not travelling along the length of the spring. Hence, we call it a stationary wave or a standing wave.

We normally represent a stationary wave by drawing the shape of the spring in its two extreme positions (Figure 14.4). The spring appears as a series of loops, separated by nodes. In this diagram, point A is moving downwards. At the same time, point B in the next loop is moving upwards. The phase difference between points A and B is 180°. Hence, the sections of spring in adjacent loops are always moving in antiphase; they are half a cycle out of phase with one another.

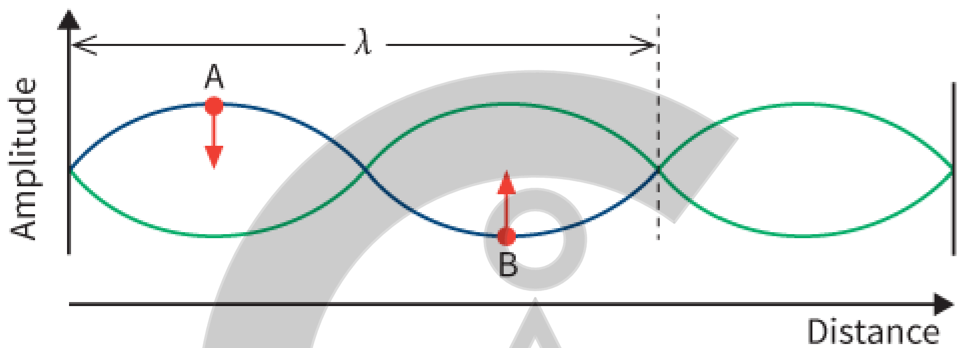


Figure 14.4: The fixed ends of a long spring must be nodes in the stationary wave pattern.

14.3 Formation of stationary waves

Imagine a string stretched between two fixed points, for example, a guitar string. Pulling the middle of the string and then releasing it produces a stationary wave. There is a node at each of the fixed ends and an antinode in the middle. Releasing the string produces two progressive waves travelling in opposite directions. These are reflected at the fixed ends. The reflected waves combine to produce the stationary wave.

Figure 14.3 shows how a stationary wave can be set up using a long spring. A stationary wave is formed whenever two progressive waves of the same amplitude and wavelength, travelling in **opposite** directions, superpose. Figure 14.5 uses a displacement–distance graph (*s*–*x*) to illustrate the formation of a stationary wave along a long spring (or a stretched length of string):

- At time $t = 0$, the progressive waves travelling to the left and right are in phase. The waves combine **constructively**, giving an amplitude twice that of each wave.
- After a time equal to one-quarter of a period $\left(t = \frac{T}{4}\right)$ each wave has travelled a distance of one quarter of a wavelength to the left or right. Consequently, the two waves are in antiphase (phase difference = 180°). The waves combine **destructively**, giving zero displacement.
- After a time equal to one-half of a period $\left(t = \frac{T}{2}\right)$ the two waves are back in phase again. They once again combine **constructively**.
- After a time equal to three-quarters of a period $\left(t = \frac{3T}{4}\right)$ the waves are in antiphase again. They combine **destructively**, with the resultant wave showing zero displacement.
- After a time equal to one whole period ($t = T$), the waves combine **constructively**. The profile of the spring is as it was at $t = 0$.

This cycle repeats itself, with the long spring showing nodes and antinodes along its length. The separation between adjacent nodes or antinodes tells us about the progressive waves that produce the stationary wave.

A closer inspection of the graphs in Figure 14.5 shows that the separation between adjacent nodes or antinodes is related to the wavelength λ of the progressive wave. The important conclusions are:

- separation between two adjacent nodes (or between two adjacent antinodes) = $\frac{\lambda}{2}$
- separation between adjacent node and antinode = $\frac{\lambda}{4}$

The wavelength λ of any progressive wave can be determined from the separation between neighbouring nodes or antinodes of the resulting stationary wave pattern.

(This separation is $\frac{\lambda}{2}$.) This can then be used to determine either the speed *v* of the progressive wave or its frequency *f* by using the wave equation:

$$v = f\lambda$$

KEY EQUATION

$$v = f\lambda$$

The wave equation, where *v* is the speed of the wave, *f* is the frequency and λ is the wavelength.

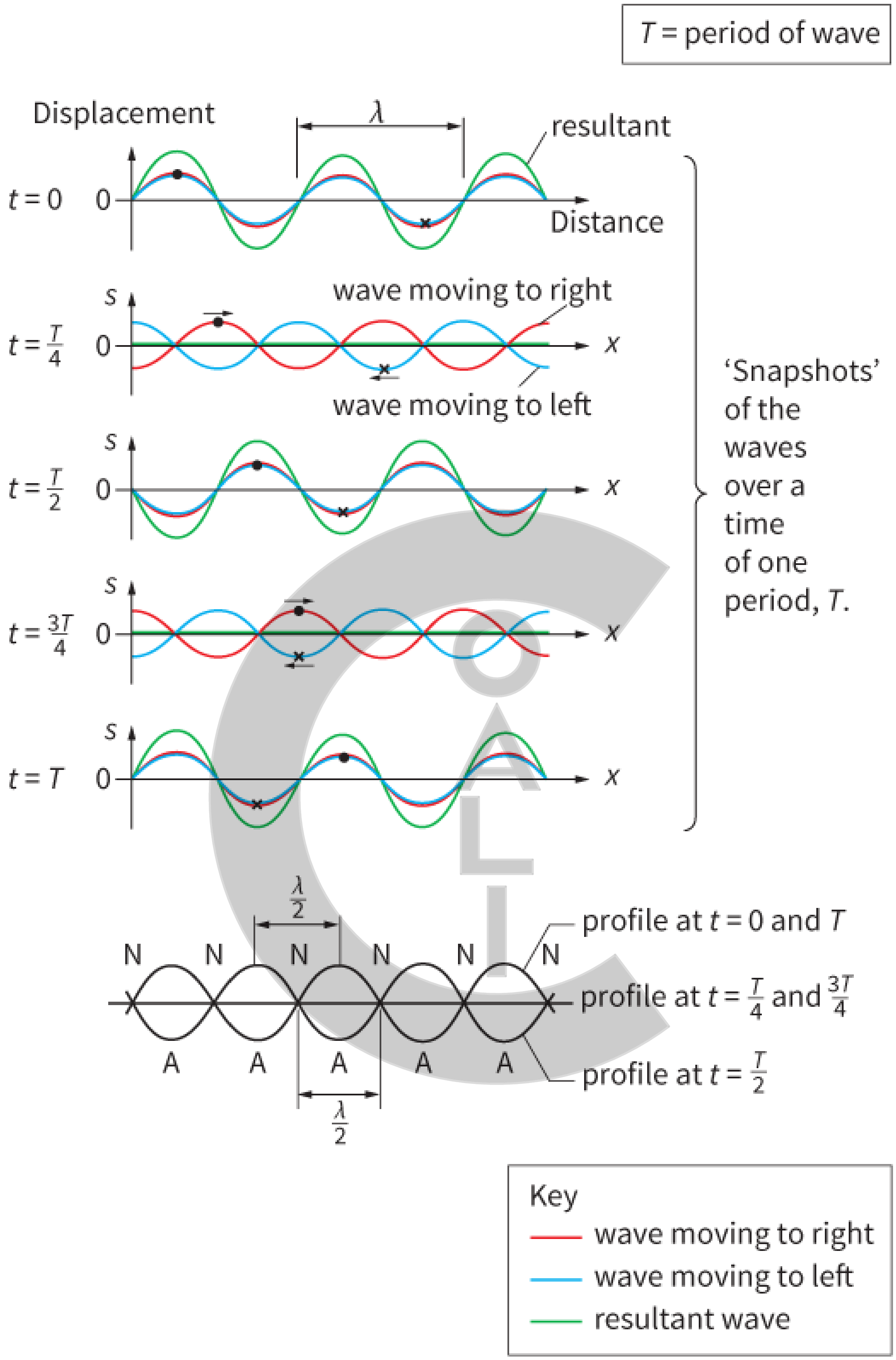


Figure 14.5: The blue-coloured wave is moving to the left and the red-coloured wave to the right. The *principle of superposition* of waves is used to determine the resultant displacement. The profile of the long spring is shown in green.

It is worth noting that a stationary wave does not travel and therefore has no speed. It does not transfer energy between two points like a progressive wave. Table 14.1 shows some of the key features of a progressive wave and its stationary wave.

	Progressive wave	Stationary wave
wavelength	λ	λ
frequency	*f*	*f*
speed	*v*	zero

Table 14.1: A summary of progressive and stationary waves.

Question

1 A stationary (standing) wave is set up on a vibrating spring. Adjacent nodes are separated by 25 cm. Calculate:

  a the wavelength of the progressive wave

  b the distance from a node to an adjacent antinode.

PRACTICAL ACTIVITY 14.1

Observing stationary waves

Here we look at experimental arrangements for observing stationary waves, for mechanical waves on strings, microwaves and sound waves in air columns.

Stretched strings: Melde's experiment

A string is attached at one end to a vibration generator, driven by a signal generator (Figure 14.6). The other end hangs over a pulley and weights maintain the tension in the string. When the signal generator is switched on, the string vibrates with small amplitude. Larger amplitude stationary waves can be produced by adjusting the frequency.

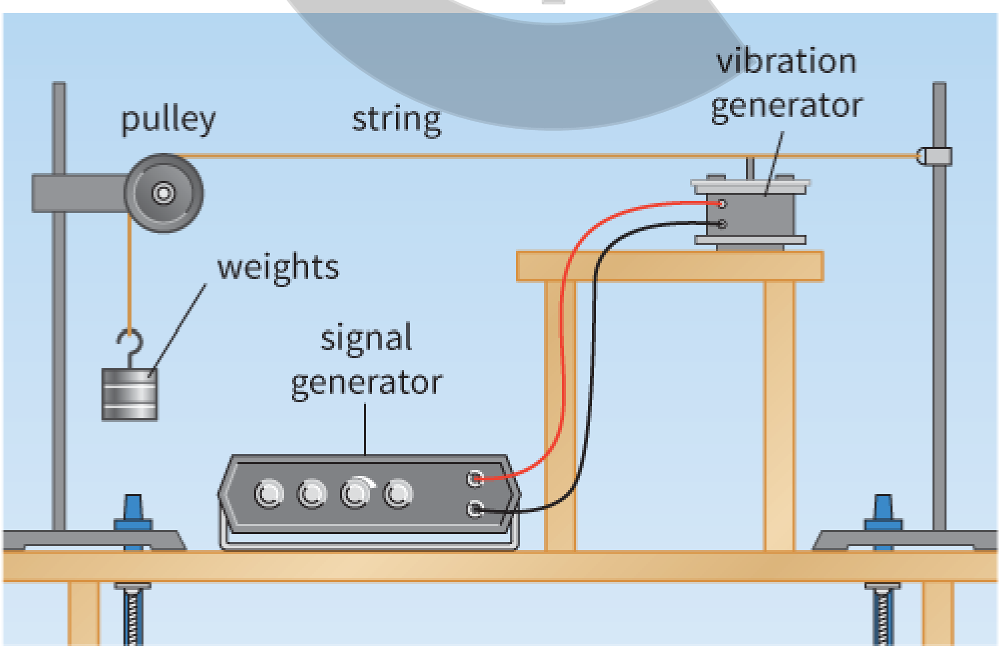


Figure 14.6: Melde's experiment for investigating stationary waves on a string.

The pulley end of the string cannot vibrate; this is a node. Similarly, the end attached to the vibrator can only move a small amount, and this is also a node. As the frequency is increased, it is possible to observe one loop (one antinode), two loops, three loops and more. Figure 14.7 shows a vibrating string where the frequency of the vibrator has been set to produce two loops.

A flashing stroboscope is useful to reveal the motion of the string at these frequencies, which look blurred to the eye. The frequency of vibration is set so that there are two loops along the string; the frequency of the stroboscope is set so that it almost matches that of the vibrations. Now we can see the string moving 'in slow motion', and it is easy to see the opposite movements of the two adjacent loops.

Figure 14.7: When a stationary wave is established, one half of the string moves upwards as the other half moves downwards. In this photograph, the string is moving too fast to observe the effect.

This experiment is known as **Melde's experiment**, and it can be extended to investigate the effect of changing the length of the string, the tension in the string and the thickness of the string.

Microwaves

Start by directing the microwave transmitter at a metal plate, which reflects the microwaves back towards the source (Figure 14.8). Move the probe receiver around in the space between the transmitter and the reflector and you will observe positions of high and low intensity. This is because a stationary wave is set up between the transmitter and the sheet; the positions of high and low intensity are the antinodes and nodes, respectively.

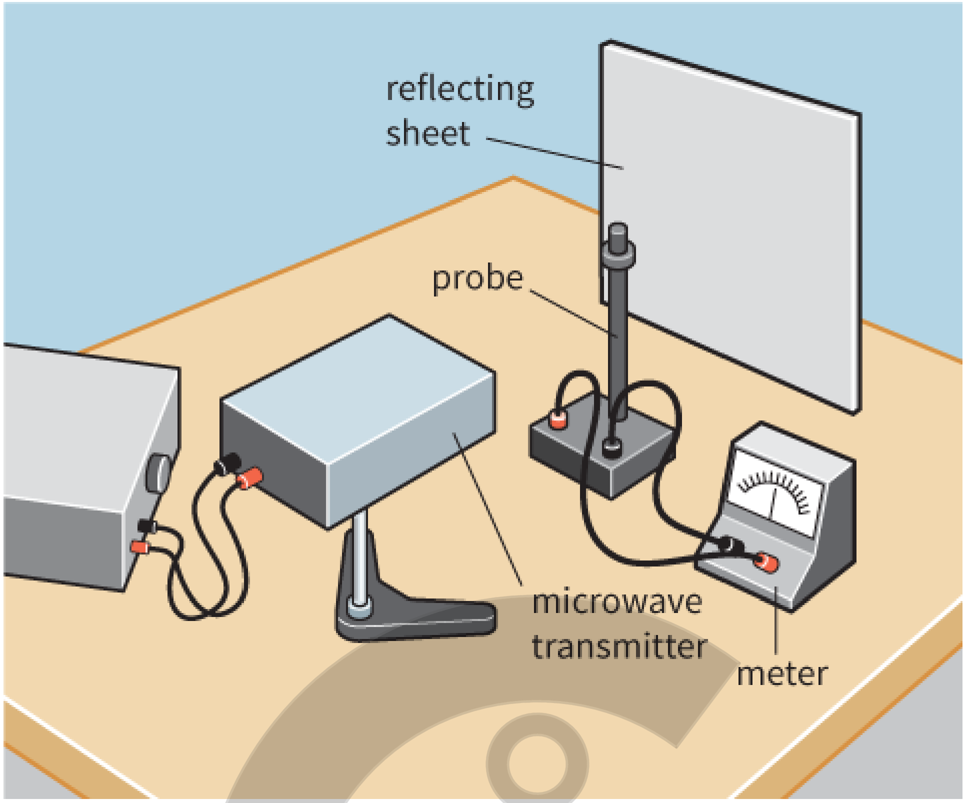


Figure 14.8: A stationary wave is created when microwaves are reflected from the metal sheet.

If the probe is moved along the direct line from the transmitter to the plate, the wavelength of the microwaves can be determined from the distance between the nodes. Knowing that microwaves travel at the speed of light c (3.0×10^8 m s^{-1}), we can then determine their frequency f using the wave equation:

$$c = f\lambda$$

An air column closed at one end

A glass tube (open at both ends) is clamped so that one end dips into a cylinder of water. By adjusting its height in the clamp, you can change the length of the column of air in the tube (Figure 14.9). When you hold a vibrating tuning fork above the open end, the air column may be forced to vibrate and the note of the tuning fork sounds much louder. This is an example of a phenomenon called **resonance**. The experiment described here is known as the resonance tube.

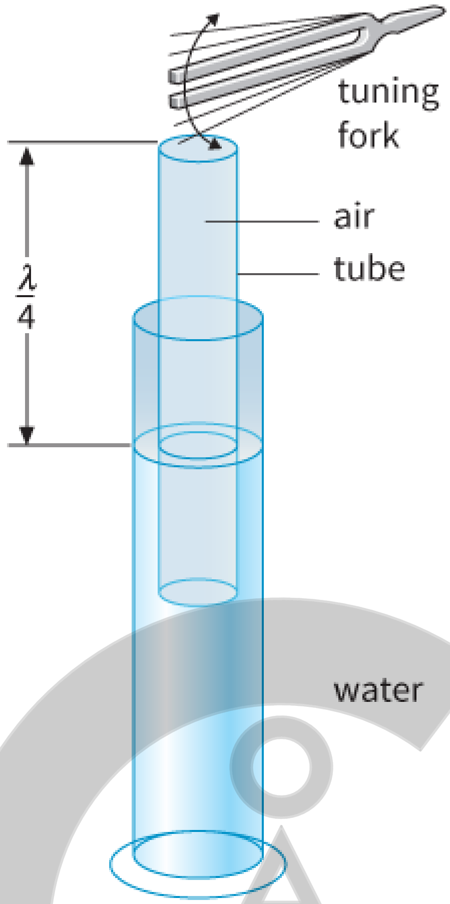


Figure 14.9: A stationary wave is created in the air in the tube when the length of the air column is adjusted to the correct length.

For resonance to occur, the length of the air column must be just right. The air at the bottom of the tube is unable to vibrate, so this point must be a node. The air at the open end of the tube can vibrate most freely, so this is an antinode. Hence, the length of the air column must be one-quarter of a wavelength (Figure 14.10a). (Alternatively, the length of the air column could be set to equal three-quarters of a wavelength – see Figure 14.10b.)

Take care! The representation of stationary sound waves can be misleading. Remember that a sound wave is a longitudinal wave, but the diagram we draw is more like a transverse wave. Figure 14.11a shows how we normally represent a stationary sound wave, while Figure 14.11b shows the direction of vibration of the particles along the wave.

Open-ended air columns

The air in a tube that is open at both ends will vibrate in a similar way to that in a closed column. Take an open-ended tube and blow gently across the top. You should hear a note whose pitch depends on the length of the tube. Now cover the bottom of the tube with the palm of your hand and repeat the process. The pitch of the note now produced will be about an octave lower than the previous note, which means that the frequency is approximately half of the original frequency.

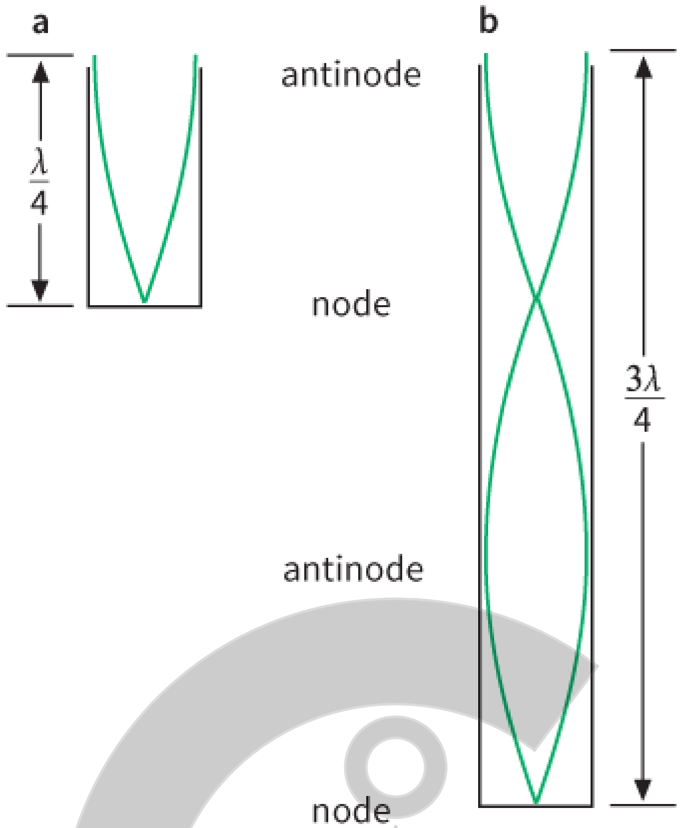


Figure 14.10: Stationary wave patterns for air in a tube with one end closed.

It is rather surprising that a stationary wave can be set up in an open column of air in this way. What is going on? Figure 14.12 compares the situation for open and closed tubes. An open-ended tube has two open ends, so there must be an antinode at each end. There is a node at the midpoint.

For a tube of length *l* you can see that in the closed tube the stationary wave formed is one-quarter of a wavelength, so the wavelength is 4*l*, whereas in the open tube it is half a wavelength, giving a wavelength of 2*l*. Closing one end of the tube thus doubles the wavelength of the note and so the frequency halves.

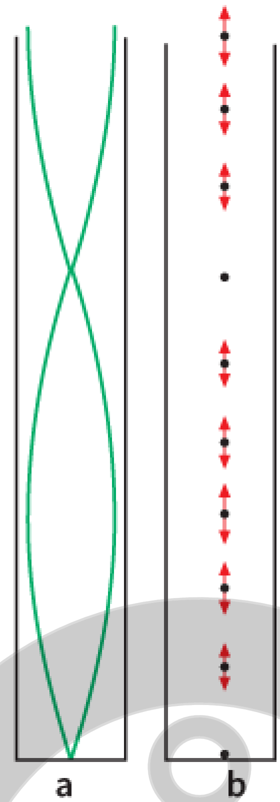


Figure 14.11: **a** The standard representation of a stationary sound wave may suggest that it is a transverse wave. **b** A sound wave is really a longitudinal wave, so that the particles vibrate as shown.

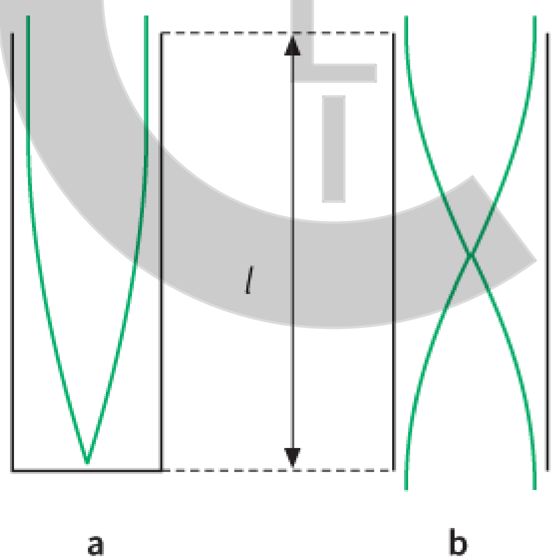


Figure 14.12: Stationary wave patterns for sound waves in **a** a closed tube, and **b** an open tube.

Questions

2 Look at the stationary (standing) wave on the string in Figure 14.7. The length of the vibrating section of the string is 60 cm.

a Determine the wavelength of the progressive wave and the separation of the two neighbouring antinodes.

The frequency of vibration is increased until a stationary wave with three antinodes appears on the string.

b Sketch a stationary wave pattern to illustrate the appearance of the string.

c Calculate the wavelength of the progressive wave on this string.

3 a Sketch a stationary wave pattern for the microwave experiment in Practical Activity 14.1. Clearly show whether there is a node or an antinode at the reflecting sheet.

b The separation of two adjacent points of high intensity is found to be 14 mm. Calculate the wavelength and frequency of the microwaves.

4 Explain how two sets of identical but oppositely travelling waves are established in the microwave and air column experiments described in Practical Activity 14.1.

Stationary waves and musical instruments (extension)

The production of different notes by musical instruments often depends on the creation of stationary waves (Figure 14.13). For a stringed instrument, such as a guitar, the two ends of a string are fixed, so nodes must be established at these points. When the string is plucked half-way along its length, it vibrates with an antinode at its midpoint. This is known as the **fundamental mode of vibration** of the string. The fundamental frequency is the **minimum frequency** of a stationary wave for a given system or arrangement.

Figure 14.13: When a guitar string is plucked, the vibrations of the strings continue for some time afterwards. Here, you can clearly see a node close to the end of each string.

Similarly, the air column inside a wind instrument is caused to vibrate by blowing, and the note that is heard depends on a stationary wave being established. By changing the length of the air column, as in a trombone, the note can be changed. Alternatively, holes can be uncovered so that the air can vibrate more freely, giving a different pattern of nodes and antinodes.

In practice, the sounds that are produced are made up of several different stationary waves having different patterns of nodes and antinodes. For example, a guitar string may vibrate with two antinodes along its length. This gives a note having twice the frequency of the fundamental, and is described as a harmonic of the fundamental. The musician's skill is in stimulating the string or air column to produce a desired mixture of frequencies.

The frequency of a harmonic is always a multiple of the fundamental frequency. The diagrams show some of the modes of vibration of a fixed length of string (Figure 14.14) and an air column in a tube of a given length that is closed at one end (Figure 14.15).

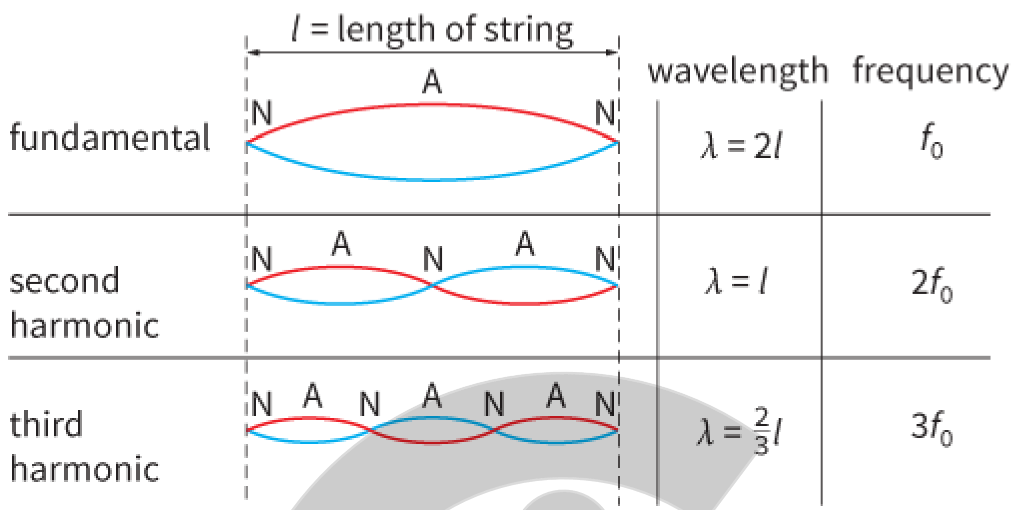


Figure 14.14: Some of the possible stationary waves for a fixed string of length *l*. The frequency of the harmonics is a multiple of the fundamental frequency f_0.

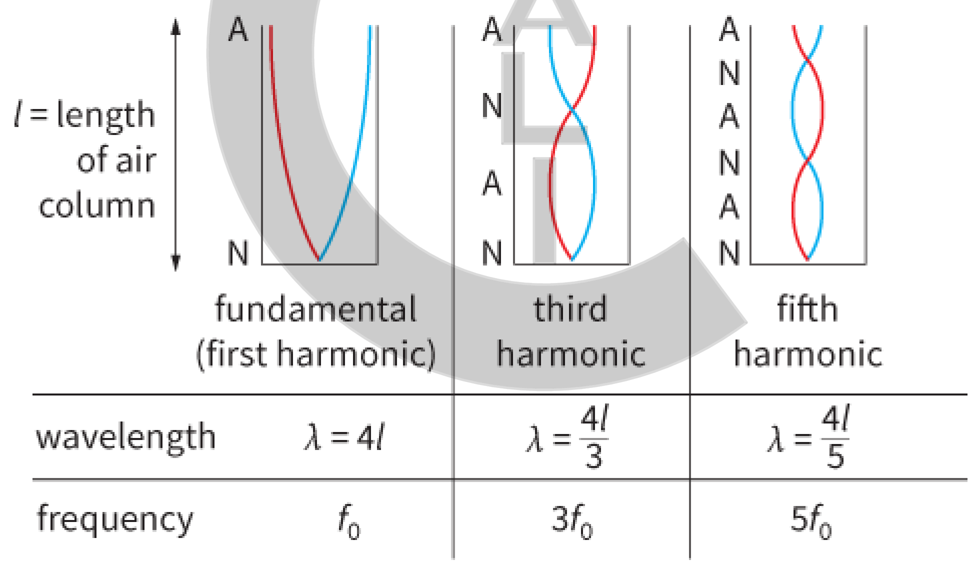


Figure 14.15: Some of the possible stationary waves for an air column, closed at one end. The frequency of each harmonic is an odd multiple of the fundamental frequency f_0.

14.4 Determining the wavelength and speed of sound

Since we know that adjacent nodes (or antinodes) of a stationary wave are separated by half a wavelength, we can use this fact to determine the wavelength λ of a progressive wave. If we also know the frequency f of the waves, we can find their speed v using the wave equation $v = f\lambda$.

One approach uses Kundt's dust tube (Figure 14.16). A loudspeaker sends sound waves along the inside of a tube. The sound is reflected at the closed end. When a stationary wave is established, the dust (fine powder) at the antinodes vibrates violently. It tends to accumulate at the nodes, where the movement of the air is zero. Hence, the positions of the nodes and antinodes can be clearly seen.

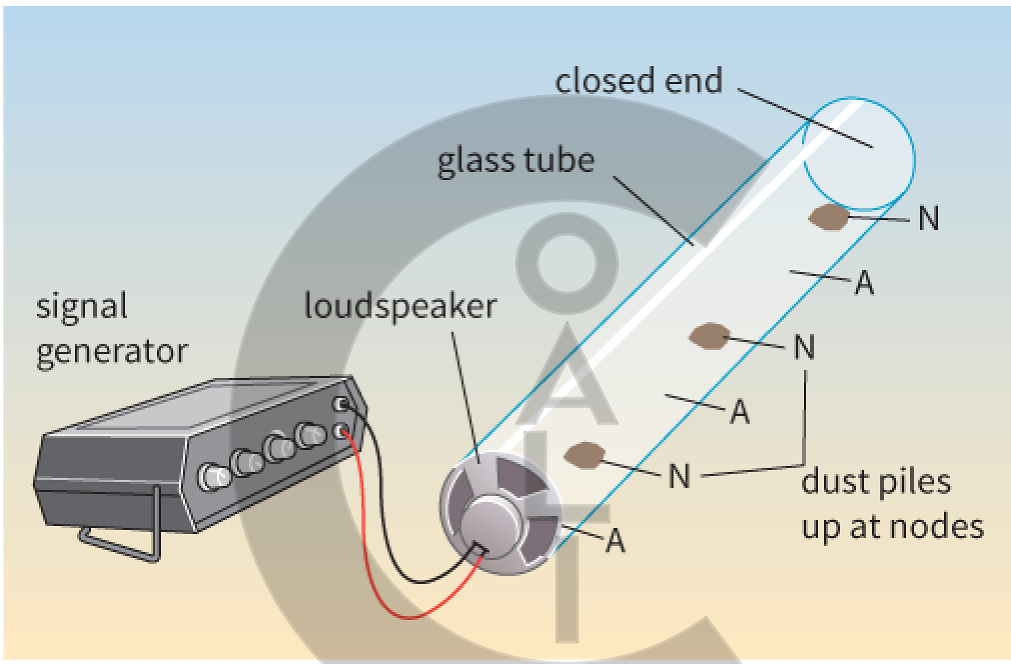


Figure 14.16: Kundt's dust tube can be used to determine the speed of sound.

PRACTICAL ACTIVITY 14.2

Using stationary sound waves to determine λ and v

This method is shown in Figure 14.17; it is the same arrangement as used for microwaves (Practical Activity 14.1). The loudspeaker produces sound waves, and these are reflected from the vertical board. The microphone detects the stationary sound wave in the space between the speaker and the board, and its output is displayed on the oscilloscope. It is simplest to turn off the time-base of the oscilloscope, so that the spot no longer moves across the screen. The spot moves up and down the screen, and the height of the vertical trace gives a measure of the intensity of the sound.

By moving the microphone along the line between the speaker and the board, it is easy to detect nodes and antinodes. For maximum accuracy, we do not measure the separation of adjacent nodes; it is better to measure the distance across several nodes.

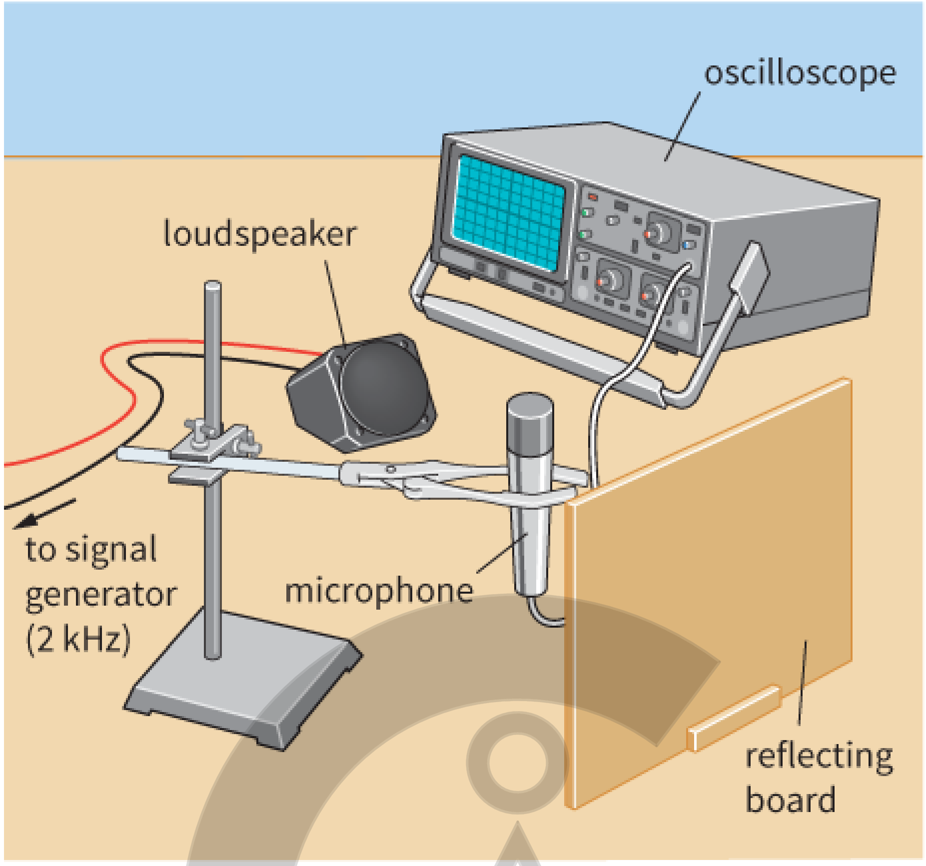


Figure 14.17: A stationary sound wave is established between the loudspeaker and the board.

Questions

5 **a** For the arrangement shown in Figure 14.17, suggest why it is easier to determine accurately the position of a node rather than an antinode.

b Explain why it is better to measure the distance across several nodes.

6 For sound waves of frequency 2500 Hz, it is found that two nodes are separated by 20 cm, with three antinodes between them.

a Determine the wavelength of these sound waves.

b Use the wave equation $v = f\lambda$ to determine the speed of sound in air.

REFLECTION

Explain to your classmates the difference between progressive sound waves and stationary sound waves.

Sketch four possible stationary wave patterns in a tube closed at just one end. Show these to your fellow learners. What grade would you give yourself for the patterns? Why?

SUMMARY

Stationary waves are formed when two identical progressive waves travelling in opposite directions meet and superpose. This usually happens when one wave is a reflection of the other.
A stationary wave has a characteristic pattern of nodes and antinodes.
A node is a point where the amplitude is always zero.
An antinode is a point of maximum amplitude.
Adjacent nodes (or adjacent antinodes) are separated by a distance equal to half a wavelength of the progressive wave.
We can use the wave equation $v = f\lambda$ to determine the speed v or the frequency f of a progressive wave. The wavelength λ is found using the nodes or antinodes of the stationary wave pattern.

EXAM-STYLE QUESTIONS

1 Which statement is *not* correct about stationary waves? **[1]**

A A stationary wave always has transverse oscillations.

B A stationary wave must have at least one node.

C The separation between two adjacent nodes is $\frac{\lambda}{2}$, where λ is the wavelength of the progressive wave.

D The superposition of two progressive waves travelling in opposite directions will produce a stationary wave.

2 A string is fixed between points X and Y.

A stationary wave pattern is formed on the stretched string.

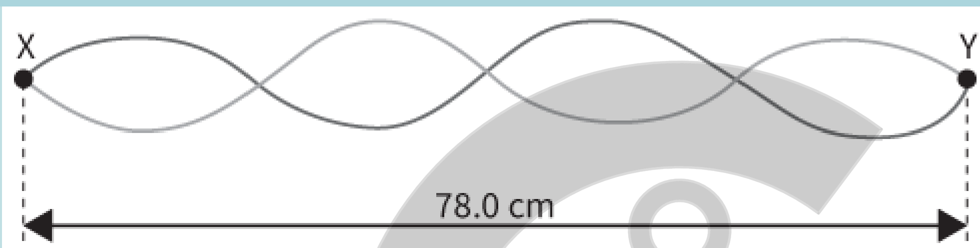


Figure 14.18

The distance between X and Y is 78.0 cm. The string vibrates at a frequency of 120 Hz.

What is the speed of the progressive wave on the string? **[1]**

A $11.7\ \text{m s}^{-1}$

B $23.4\ \text{m s}^{-1}$

C $46.8\ \text{m s}^{-1}$

D $93.6\ \text{m s}^{-1}$

3 This diagram shows a stationary wave on a string.

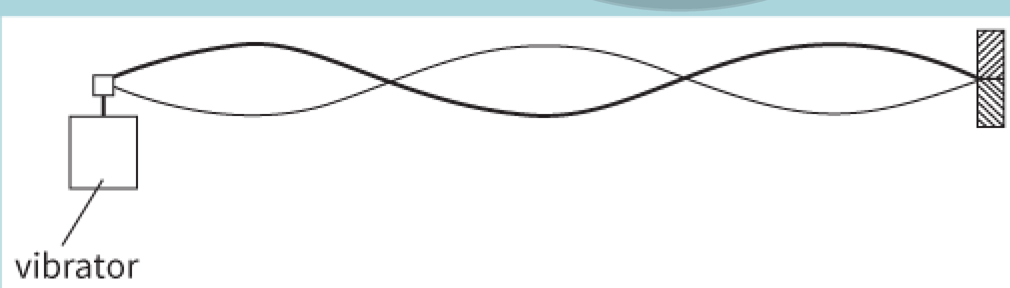


Figure 14.19

a On a copy of the diagram, label one **node** (N) and one **antinode** (A). **[1]**

b Mark on your diagram the wavelength of the progressive wave and label it λ. **[1]**

c The frequency of the vibrator is doubled. Describe the changes in the stationary wave pattern. **[1]**

[Total: 3]

4 A tuning fork that produces a note of 256 Hz is placed above a tube that is nearly filled with water. The water level is lowered until resonance is first heard.

a Explain what is meant by the term **resonance**. **[1]**

b The length of the column of air above the water when resonance is first heard is 31.2 cm.

Calculate the speed of the sound wave. **[2]**

[Total: 3]

5 a State **two** similarities and **two** differences between progressive waves and stationary waves. **[4]**

b This diagram shows an experiment to measure the speed of a sound in a string. The frequency of the vibrator is adjusted until the stationary wave shown is formed.

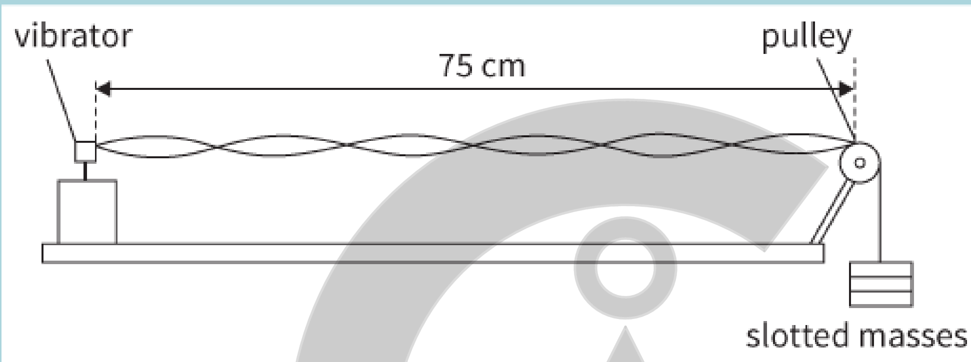


Figure 14.20

i On a copy of the diagram, mark a node (label it N) and an antinode (label it A). **[2]**

ii The frequency of the vibrator is 120 Hz. Calculate the speed at which a progressive wave would travel along the string. **[3]**

c The experiment is now repeated with the load on the string halved. In order to get a similar stationary wave the frequency has to be decreased to 30 Hz. Explain, in terms of the speed of the wave in the string, why the frequency must be adjusted. **[2]**

[Total: 11]

6 This diagram shows a stationary wave, of frequency 400 Hz, produced by a loudspeaker in a closed tube.

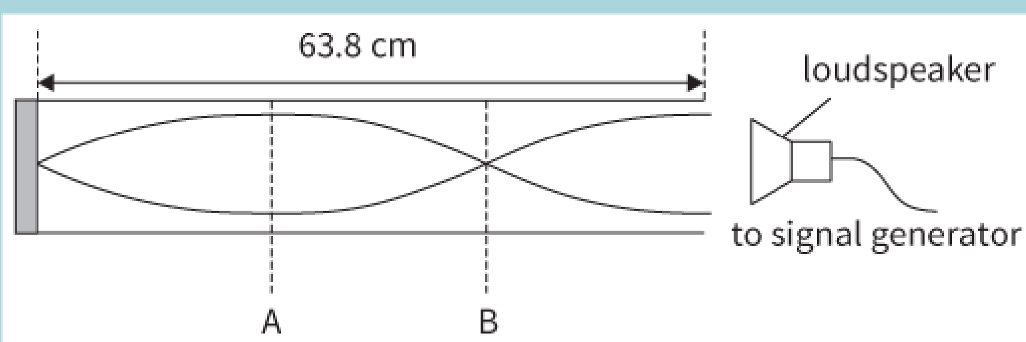


Figure 14.21

a Describe the movement of the air particles at:

i A [2]

ii B [1]

b The length the tube is 63.8 cm.

Calculate the speed of the sound. [3]

[Total: 6]

7 a Explain what is meant by:

i a **coherent** source of waves. [2]

ii **phase difference**. [2]

b A student, experimenting with microwaves, sets up the arrangement shown in this diagram.

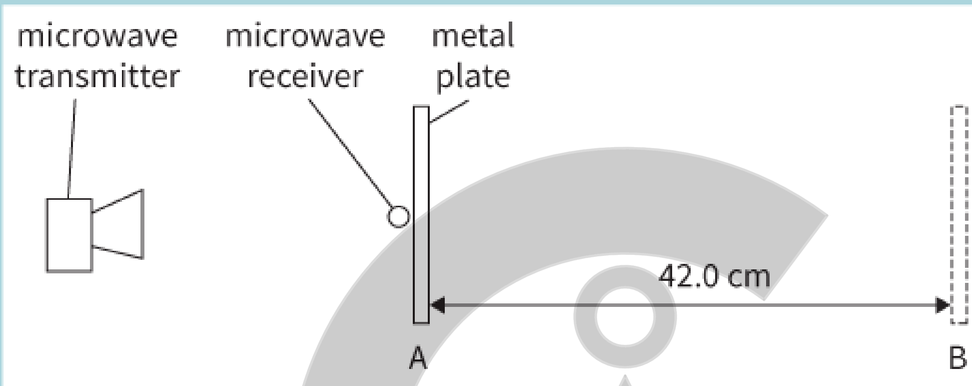


Figure 14.22

With the metal plate at position A there is a very small signal. He slowly moves the plate back, leaving the receiver in the same position. As he does so, he finds that the intensity initially rises until it becomes a maximum, then falls back to a minimum. This cycle repeats a total of five times until the plate reaches position B, where once again there is a minimum.

i Explain why a series of maxima and minima are heard. [2]

ii Determine the frequency of the microwaves. [5]

c Explain why there was a minimum when the plate was at position A, next to the detector. [2]

[Total: 13]

8 This diagram shows an experiment to measure the speed of sound in air.

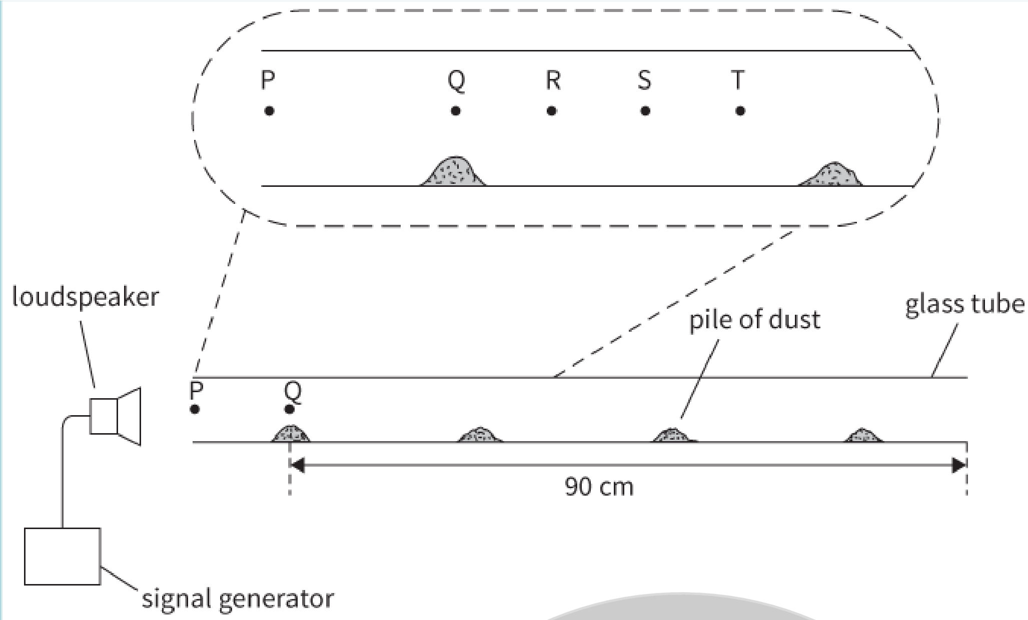


Figure 14.23

A small amount of dust is scattered along the tube. The loudspeaker is switched on. When the frequency is set at 512 Hz the dust collects in small piles as shown in the diagram.

a Determine the wavelength of the sound wave and calculate the speed of sound in the air in the tube. **[3]**

b On a copy of the diagram, show the movement of the air particles at positions P, Q, R, S and T. **[3]**

c Mark two points on your diagram where the movements of the air particles are 180° out of phase with each other. Label them A and B. **[1]**

[Total: 7]

9 The speed *v* of a transverse wave on a stretched wire is given by the expression

$v \propto \sqrt{T}$

where *T* is the tension in the wire.

A length of wire is stretched between two fixed point. The tension in the wire is *T*. The wire is gently plucked from the middle. A stationary wave, of fundamental frequency 210 Hz, is produced.

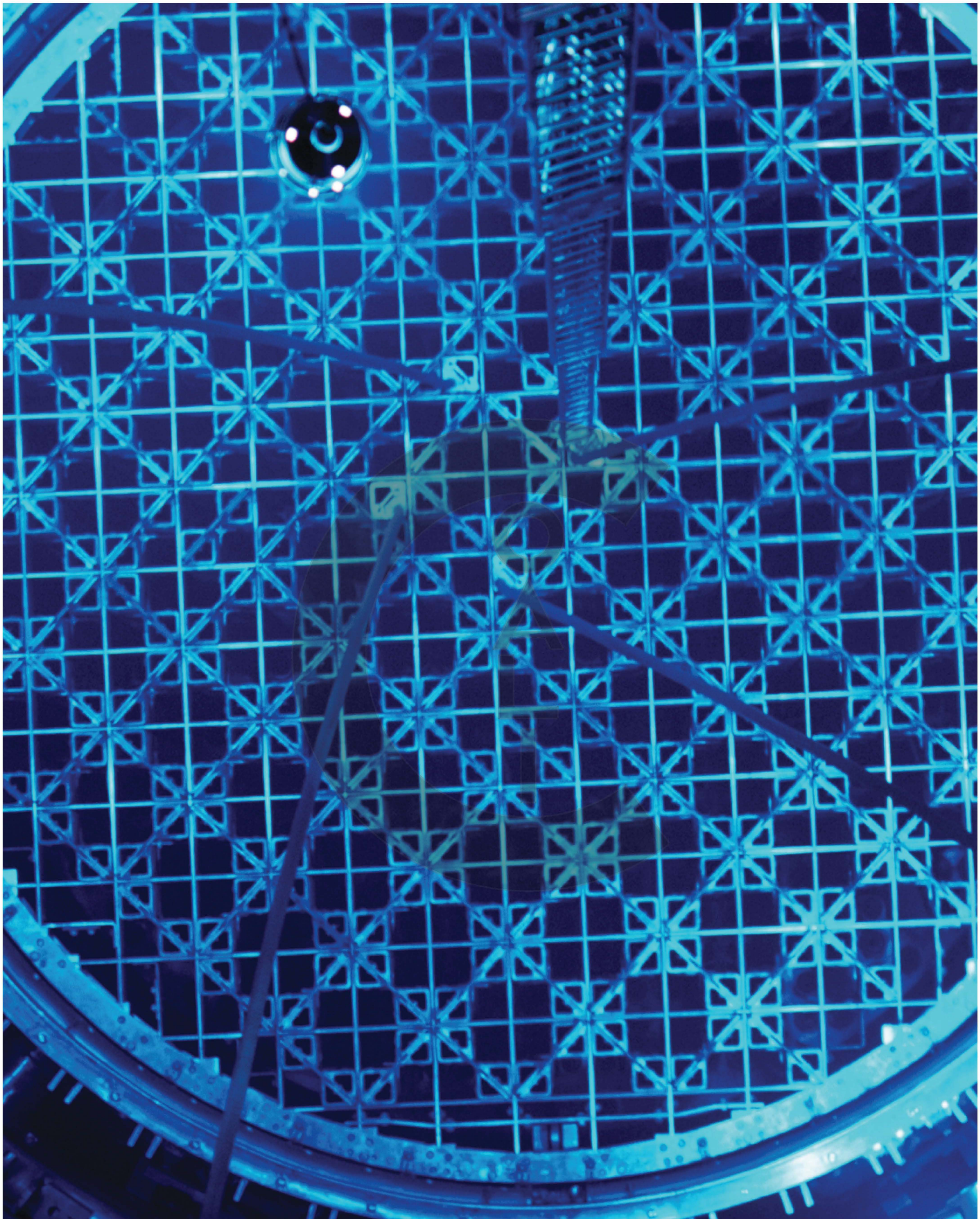
The tension in the wire is now increased to 1.4*T*. The percentage uncertainty in new tension is 8.0%. The length of the wire is unchanged.

Calculate the new value for the fundamental frequency when the wire is plucked in the middle. Your answer must include the absolute uncertainty written to an appropriate number of significant figures. **[4]**

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
explain the formation of stationary waves using graphical methods	14.1, 14.3			
describe experiments that demonstrate stationary waves using microwaves, stretched strings and air columns	14.3			
state what is meant by nodes and antinodes	14.2			
recall the separation between neighbouring nodes (or antinodes) in terms of the wavelength of the progressive wave	14.3			
determine the wavelength of sound using stationary waves.	14.4			



〉 Chapter 15

Atomic structure and particle physics

LEARNING INTENTIONS

In this chapter you will learn how to:

- describe the nuclear model of the atom and the evidence for it
- show an understanding of the nature and properties of α-, β- and γ-radiations
- understand that in α and β decay a nuclide changes into a different nuclide
- recognise that there are two classes of sub-atomic particles – leptons and hadrons
- recognise that leptons are fundamental particles
- appreciate that electrons and neutrinos are leptons
- recognise that hadrons are not fundamental particles
- understand that hadrons are made up of particles called quarks.

BEFORE YOU START

- Try drawing the structure of the atom.
- Suggest why, in the late 19th century, physicists felt that atoms were not the basic building blocks of matter and that the atoms themselves had an internal structure. Discuss your ideas with your fellow students.

RADIOACTIVITY AT WORK

Radioactive substances have many uses, for example, in engineering and medicine.

In the 1950s, many shoe shops had an X-ray machine where you put your feet into an opening and you could view the bones in your feet on a fluorescent screen – quite exciting for a young child! These have long since disappeared. Why do you think they are not used anymore?

Radioactive substances must be handled with great care to ensure that no-one becomes contaminated and so exposed to the radiation that comes from these substances (Figure 15.1).

Do you know how modern-day workers who are likely to be exposed to radiation (such as radiographers in a hospital) are protected from radiation? Are the short-term and long-term protections different?

In this chapter, we will look at the structure of the atom, and then the nature of radioactive substances and the different types of radiation they produce.

Figure 15.1: A worker at a nuclear power station is checked for any radioactive material on his body.

15.1 Looking inside the atom

The idea that matter is composed of very small particles called atoms was first suggested by the Ancient Greeks about 2000 years ago. However, it was not until the middle of the 19th century that any ideas about the **inside** of the atom were proposed.

It was the English scientist J.J. Thomson who suggested that the atom is a neutral particle made of a positive charge with lumps of negative charge (electrons) in it. He could not determine the charge and the mass of the negative particles separately, but it was clear that a new particle, probably much smaller than the hydrogen atom, had been discovered. Since atoms are neutral and physicists had discovered a negatively charged part of an atom, it meant that there were both positive and negative charges in an atom. We now call this the **plum pudding model of the atom** (positive pudding with negative plums!).

Other experiments show that the electron has a mass of approximately 9.11×10^{-31} kg (m_e) and a charge of -1.60×10^{-19} C ($-e$). Today, we use the idea of the electron to explain all sorts of phenomena, including electrostatics, current electricity and electronics.

15.2 Alpha-particle scattering and the nucleus

Early in the 20th century, many physicists were investigating the recently discovered phenomenon of radioactivity, the process whereby unstable nuclei emit radiation. One kind of radiation they found consisted of what they called α-particles (alpha-particles).

These α-particles were known to have a similar mass to the smaller atoms (such as hydrogen, helium and lithium) and had relatively high kinetic energies. Hence, they were useful in experiments designed to discover the composition of atoms.

Figure 15.2: Ernest Rutherford (on the right) in the Cavendish Laboratory, Cambridge, England. He had a loud voice that could disturb sensitive apparatus and so the notice was a joke aimed at him.

In 1906, while experimenting with the passage of α-particles through a thin mica sheet, Ernest Rutherford (Figure 15.2) noticed that most of the α-particles passed straight through. (Mica is a natural mineral that can be split into very thin sheets.) This suggested to him that there might be a large amount of empty space in the atom, and by 1909 he had developed what we now call the **nuclear model of the atom**.

In 1911, Rutherford carried out a further series of experiments with Hans Geiger and Ernest Marsden at the University of Manchester using gold foil in place of the mica. They directed parallel beams of α-particles at a piece of gold foil only 10^{-6} m thick. Most of the α-particles went straight through. Some were deflected slightly, but about 1 in 20 000 were deflected through an angle of more than 90°, so that they appeared to bounce back off the foil. This helped to confirm Rutherford in his thinking about the atom – that it was mostly empty space, with most of the mass and all of the positive charge concentrated in a tiny region at the centre. This central **nucleus** only affected the α-particles when they came close to it.

Later, Rutherford wrote: ‘It was quite the most incredible event that has happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.’ In fact, he was not quite as surprised as this suggests, because the results confirmed ideas he had used in designing the experiment.

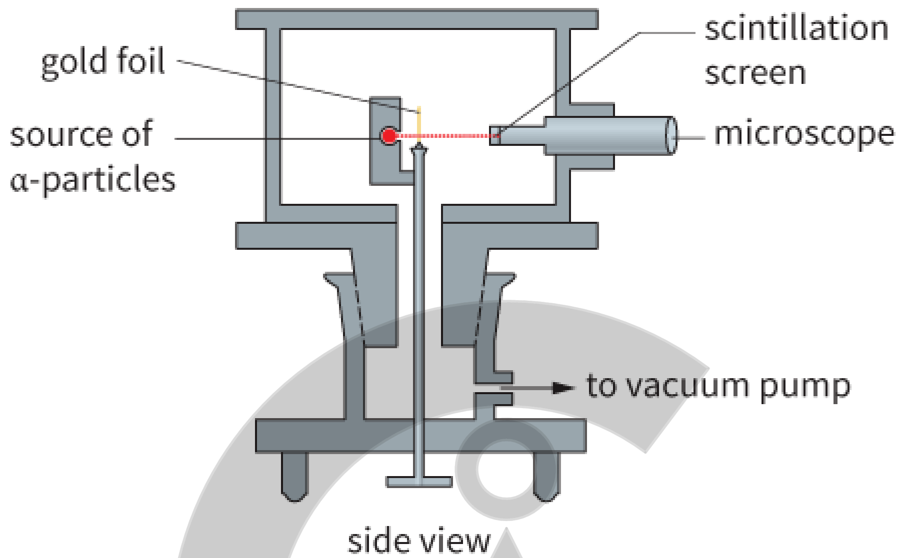


Figure 15.3: The apparatus used for the α-scattering experiment. The microscope can be moved round to detect scattered radiation at different angles.

Figure 15.3 shows the apparatus used in the α-scattering experiment. Notice the following points:

- The α-particle source was encased in metal with a small aperture, allowing a fine beam of α-particles to emerge.
- Air in the apparatus was pumped out to leave a vacuum; α-radiation is absorbed by a few centimetres of air.
- One reason for choosing gold was that it can be made into a very thin sheet or foil. Rutherford’s foil was only a few hundreds of atoms thick.
- The α-particles were detected when they struck a solid ‘scintillating’ material. Each α-particle gave a tiny flash of light and these were counted by the experimenters (Geiger and Marsden).
- The detector could be moved round to detect α-particles scattered through different angles.

Geiger and Marsden had the difficult task of observing and counting the tiny flashes of light produced by individual α-particles striking the scintillation screen. They had to spend several minutes in the darkened laboratory to allow the pupils of their eyes to become dilated so that they could see the faint flashes. Each experimenter could only stare into the detector for about a minute before the strain was too much and they had to change places.

Explaining α-scattering

How can we explain the back-scattering of α-particles by the gold atoms?

If the atom was as Thomson pictured it, with negatively charged electrons scattered through a ‘pudding’ of positive charge, an individual α-particle would pass through it like a bullet, hardly being deflected at all. This is because the α-particles are more massive than electrons–they might push an electron out of the atom, but their own path would be scarcely affected.

However, if the mass and positive charge of the atom were concentrated at one point in the atom, as Rutherford suggested, an α-particle striking this part would be striking something more massive than itself and with a greater charge. A head-on collision would send the α-particle backwards.

The paths of an α-particle near a nucleus are shown in Figure 15.4.

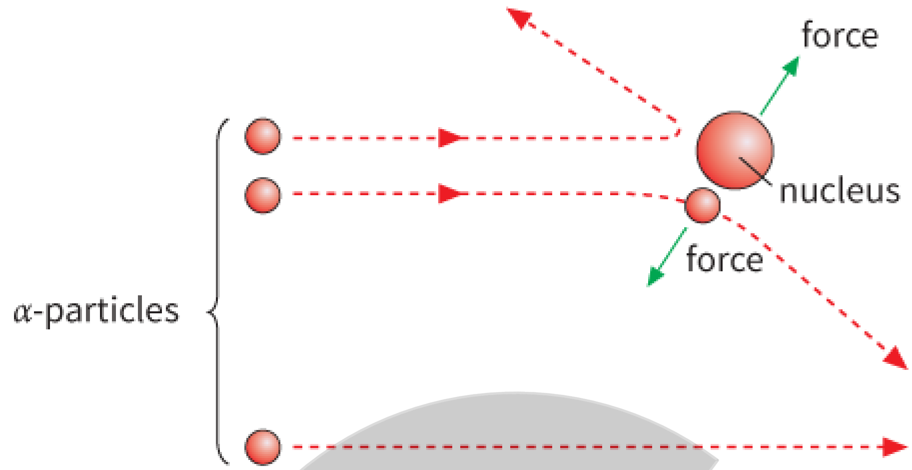


Figure 15.4: Possible paths of an α-particle near a nucleus. The nucleus and the α-particle both experience electrostatic repulsion.

Rutherford reasoned that the large deflection of the α-particle must be due to a very small charged nucleus. From his experiments he calculated that the diameter of the gold nucleus was about 10^{-14} m. It has since been shown that the very large deflection of the α-particle is due to the electrostatic repulsion between the positive charge of the α-particle and the positive charge of the nucleus of the atom. The closer the path of the α-particle gets to the nucleus, the greater will be this repulsion. An α-particle making a 'head-on' collision with a nucleus is back-scattered through 360°. The α-particle and nucleus both experience an equal but opposite repulsive electrostatic force *F*. This force has a much greater effect on the motion of the α-particle than on the massive nucleus of gold.

PRACTICAL ACTIVITY 15.1

An analogy for Rutherford scattering

Roll a ball-bearing down a slope towards a cymbal. It may be deflected but, even if you roll it directly at the cymbal's centre, it will not come back – it will roll over the centre and carry on to the other side. However, if you roll the ball-bearing towards a 'tin hat' shape (with a much narrower but higher central bulge) any ball-bearings that you roll close to the centre will deflect a lot, and any ball-bearings that you roll directly towards the centre will roll straight back. This is a very simple analogy (or model) of Rutherford's experiment.

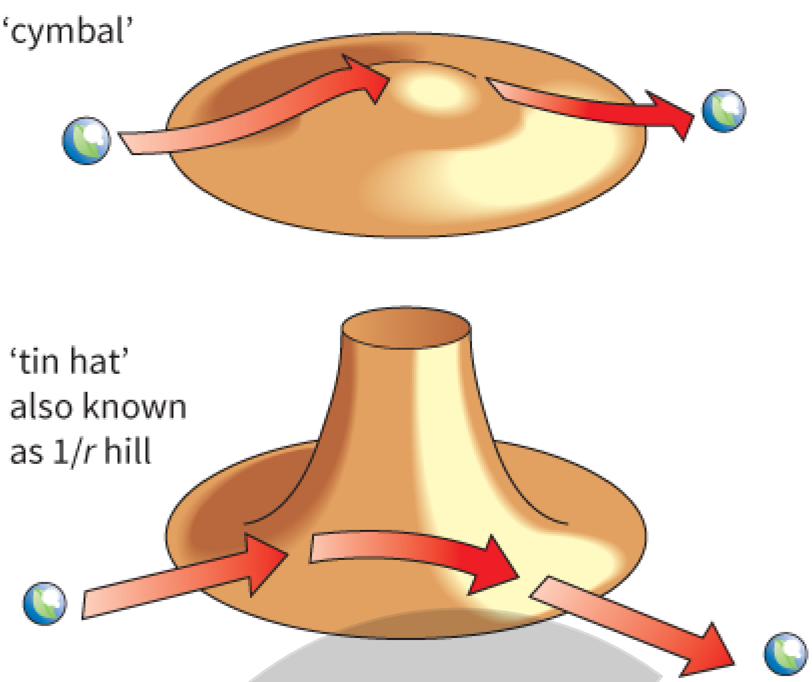


Figure 15.5: An analogy for Rutherford's experiment.

The shape of the cymbal represents the shape of the electric field of an atom in the 'plum pudding' model: low central intensity and spread out. The 'tin hat' represents the shape of the electric field for the nuclear model: high central intensity and concentrated.

From the α-particle scattering experiment, Rutherford deduced the following.

- An α-particle is deviated due to the repulsive force between the α-particle and the positive charge in the atom.
- Most α-particles have little or no deviation–so most of atom is empty space.
- A very few α-particles are deviated more than 90° – so most of the mass of an atom is concentrated in a small space (the nucleus) and most of the atom is empty space.

Questions

1 Rutherford's scattering experiments were done in an evacuated container. Explain why this is necessary.

2 In Rutherford's experiment, α-particles were directed at a thin gold foil. A small fraction of the α-particles were back-scattered through 180°.

Describe and explain how the fraction back-scattered changes if each of the following changes are (separately) made.

a A thicker foil is used.

b Faster α-particles are used.

c A silver foil is used – a silver nucleus has less positive charge than a gold nucleus.

15.3 A simple model of the atom

After Rutherford had presented his findings, the nuclear model of the atom gained rapid acceptance. This was partly because it helped chemists to explain the phenomenon of chemical bonding (the way in which atoms bond together to form molecules). Subsequently, the proton was discovered. It had a positive charge, equal and opposite to that of the electron. However, its mass was too small to account for the entire mass of the atom and it was not until the early 1930s that this puzzle was solved by the discovery of the neutron, an uncharged particle with a similar mass to that of the proton. This suggests a model for the atom like the one shown in Figure 15.6:

- Protons and neutrons make up the nucleus of the atom.
- The electrons move around the nucleus in a cloud, some closer to and some further from the centre of the nucleus.

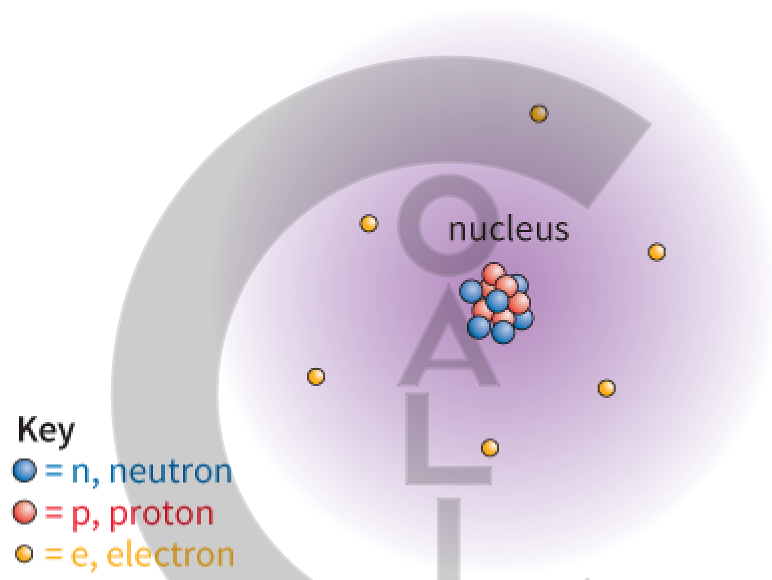


Figure 15.6: A simple model of the atom. If the nucleus were drawn to scale, it would be invisible (and the electrons are even smaller!).

From this model it looks as though all matter, including ourselves, is mostly empty space. For example, if we scaled up the hydrogen atom so that the nucleus was the size of a 1 cm diameter marble, the orbiting electron would be a grain of sand about 800 m away!

The scale of things

It is useful to have an idea of the approximate sizes of typical particles:

- radius of proton ~ radius of neutron ~ 10^{-15} m
- radius of nucleus ~ 10^{-15} m to 10^{-14} m
- radius of atom ~ 10^{-10} m
- size of molecule ~ 10^{-10} m to 10^{-6} m.

(Some molecules, such as large protein molecules, are very large indeed – compared to an atom!)

The radii of nuclear particles are often quoted in femtometres (fm), where 1 fm = 10^{-15} m.

Nuclear density

We can picture a proton as a small, positively charged sphere. Knowing its mass and radius, we can calculate its density:

$$\text{mass of proton } m_p = 1.67 \times 10^{-27} \text{ kg}$$

$$\text{radius of proton } r = 0.80 \text{ fm} = 0.80 \times 10^{-15} \text{ m}$$

(In fact, the radius of the proton is not very accurately known; it is probably between 0.80×10^{-15} m and 0.86×10^{-15} m.)

$$\begin{aligned}
\text{volume of proton} &= \tfrac{4}{3}\pi r^3 \\
&= \tfrac{4}{3}\pi \times (0.80 \times 10^{-15})^3 \\
&= 2.14 \times 10^{-45} \text{ m}^3 \\
\text{density} &= \frac{\text{mass}}{\text{volume}} \\
&= \frac{1.67 \times 10^{-27}}{2.14 \times 10^{-45}} \\
&\approx 7.8 \times 10^{17} \text{ kg m}^{-3}
\end{aligned}$$

So the proton has a density of roughly 10^{18} kg m^{-3}. This is also the density of a neutron, and of an atomic nucleus, because nuclei are made of protons and neutrons held closely together.

Compare the density of nuclear material with that of water whose density is 1000 kg m^{-3} – the nucleus is 10^{15} times as dense. Nuclear matter the size of a tiny grain of sand would have a mass of about a million tonnes! This is a consequence of the fact that the nucleus occupies only a tiny fraction of the volume of an atom. The remainder is occupied by the cloud of orbiting electrons whose mass makes up less than one-thousandth of the atomic mass.

Question

3 Gold has a density of 19 700 kg m^{-3}. A mass of 193 g of gold contains 6.02×10^{23} atoms. Use this information to estimate the volume of a gold atom, and hence its radius. State any assumptions you make.

15.4 Nucleons and electrons

We will start this topic with a summary of the particles mentioned so far (Table 15.1).

Particle	Relative mass (proton = 1)[(a)]	Charge[(b)]
proton (p)	1	$+e$
neutron (n)	1	0
electron (e)	0.0005	$-e$
alpha-particle (α)	4	$+2e$

[(a)] The numbers given for the masses are approximate.

[(b)] $e = 1.60 \times 10^{-19}$ C.

Table 15.1: Summary of the particles that we have met so far in this chapter. The α-particle is in fact a helium nucleus (with two protons and two neutrons).

All nuclei, except the lightest form of hydrogen, contain protons and neutrons, and each nucleus is described by the number of protons and neutrons that it contains.

- Protons and neutrons in a nucleus are collectively called **nucleons**. For example, in a nucleus of gold, there are 79 protons and 118 neutrons, giving a total of 197 nucleons altogether.
- The total number of nucleons in a nucleus is called the **nucleon number** (or mass number) *A*.
- The nucleon number is the sum of the number of neutrons and protons in the nucleus, or A = N +Z (where A = nucleon number, N = **neutron number** and Z = **proton number**).

The unit used to measure masses at this level is the **unified atomic mass unit** (u).

1 u is defined as being one-twelfth of the mass of a carbon-12 atom.

An isolated proton has a mass of 1.007 276 466 77 u and an isolated neutron has a mass 1.008 665 u. You can see that there is a discrepancy between the sum of the masses of the protons and neutrons in a carbon-12 atom and the sum of the masses of six isolated protons and six isolated neutrons

The reasons for these discrepancies are explored in detail in Chapter 29.

A specific combination of protons and neutrons in a nucleus is called a **nuclide**.

The nucleus of any atom can be represented by the symbol for the element (shown here as X) along with the nucleon number *A* and proton number *Z*:

$$\begin{matrix}\text{nucleon number}\rightarrow \\ \text{proton number}\rightarrow\end{matrix} {}^{A}_{Z}\mathrm{X}$$

For example:

Element	Symbol	Nucleon number *A*	Proton number *Z*	Represented as:
oxygen	O	18	8	${}^{16}_{8}\mathrm{O}$
gold	Au	197	79	${}^{197}_{79}\mathrm{Au}$
uranium	U	238	92	${}^{238}_{92}\mathrm{U}$

The proton and nucleon numbers of some common nuclides are shown in Table 15.2.

Element	Nucleon number *A*	Proton number *Z*	Element	Nucleon number *A*	Proton number *Z*
hydrogen	1	1	bromine	79	35
helium	4	2	silver	107	47
lithium	7	3	tin	120	50
beryllium	9	4	iodine	130	53
boron	11	5	caesium	133	55
carbon	12	6	barium	138	56
nitrogen	14	7	tungsten	184	74
oxygen	16	8	platinum	195	78
neon	20	10	gold	197	79
sodium	23	11	mercury	202	80
magnesium	24	12	lead	206	82
aluminium	27	13	bismuth	209	83
chlorine	35	17	radium	226	88
calcium	40	20	uranium	238	92
iron	56	26	plutonium	239	94
nickel	58	28	americium	241	95

Table 15.2: Proton and nucleon numbers of some nuclides.

Questions

4 Table 15.2 shows the proton and nucleon numbers of several nuclei. Determine the number of neutrons in the nuclei of the following elements shown in the table:

- **a** nitrogen
- **b** bromine
- **c** silver
- **d** gold
- **e** mercury.

5 State the charge of each of the following in terms of the elementary charge *e*:

- **a** proton
- **b** neutron
- **c** nucleus
- **d** molecule
- **e** α-particle.

Isotopes

Although atoms of the same element may be identical chemically, their nuclei may be slightly different. The number of protons in the nucleus of an atom determines what element it is: helium always has two protons, carbon six protons, oxygen eight protons, neon 10 protons, radium 88 protons, uranium 92 protons and so on.

However, the number of neutrons in the nuclei for a given element can vary. Take neon as an example. Three different naturally occurring forms of neon are:

$$^{20}_{10}\text{Ne} \quad ^{21}_{10}\text{Ne} \quad ^{22}_{10}\text{Ne}$$

The first has 10 neutrons in the nucleus, the second 11 neutrons and the third 12 neutrons. These three types of neon nuclei are called **isotopes** of neon. Each isotope has the same number of protons (for neon this is 10) but a different number of neutrons. The word 'isotope' comes from the Greek *isotopos* (same place), because all isotopes of the same element have the same place in the Periodic Table of elements.

Isotopes are nuclei of the same element with different numbers of neutrons but the same number of protons.

Any atom is electrically neutral (it has no net positive or negative charge), so the number of electrons surrounding the nucleus must equal the number of protons in the nucleus of the atom. If an atom gains or loses an electron, it is no longer electrically neutral and is called an **ion**.

For an atom, the number of protons (and hence the number of electrons) determines the chemical properties of the atom. The number of protons and the number of neutrons determine the nuclear properties. It is important to realise that, since the number of protons, and therefore the number of electrons, in isotopes of the same element are identical, they will all have the same chemical properties but very different nuclear properties.

Hydrogen has three important isotopes, $^{1}_{1}\text{H}$ (sometimes called protium), $^{2}_{1}\text{H}$ (deuterium) and $^{3}_{1}\text{H}$ (tritium) (Figure 15.7).

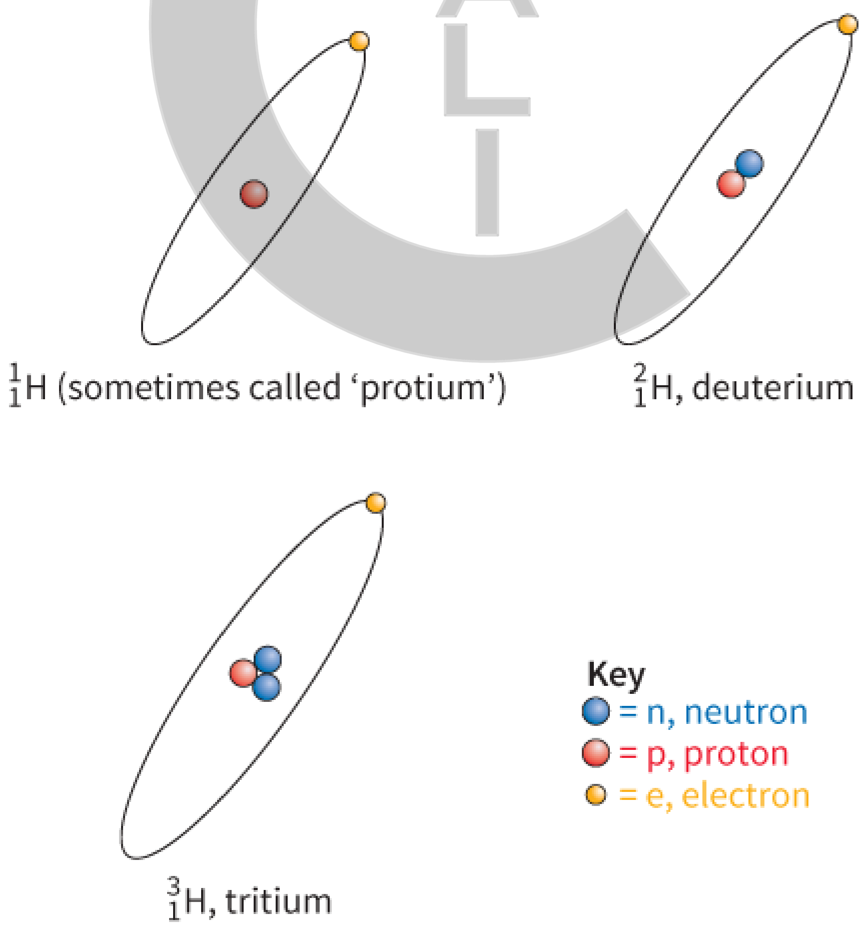


Figure 15.7: The isotopes of hydrogen.

Protium and deuterium occur naturally, but tritium has to be made. Deuterium and tritium form the fuel of many fusion research reactors. Hydrogen is the most abundant element in the Universe (Figure 15.8), because it consists of just one proton and one electron, which is the simplest structure possible for an atom.

Figure 15.8: The Horsehead Nebula in Orion. The large coloured regions are expanses of dust and gas, mostly hydrogen, that are ionised by nearby stars so that they emit light. The dark 'horse head' is where the areas of gas and dust remain in atomic form and block out the light from behind.

The different numbers of neutrons in the isotopes of an element means that the isotopes will have different relative atomic masses. There are differences too in some of their physical properties, such as density and boiling point. For example, heavy water, which is water containing deuterium, has a boiling point of 104 °C under normal atmospheric pressure.

Table 15.3 gives details of some other commonly occurring isotopes.

Element	Nucleon number *A*	Proton number *Z*	Neutron number *N*
hydrogen	1	1	0
	2	1	1
carbon	12	6	6
	14	6	8
oxygen	16	8	8
	18	8	10
neon	20	10	10
	21	10	11
potassium	39	19	20

Element	Nucleon number *A*	Proton number *Z*	Neutron number *N*
	40	19	21
strontium	88	38	50
	90	38	52
caesium	135	55	80
	137	55	82
lead	206	82	124
	208	82	126
radium	226	88	138
	228	88	140
uranium	235	92	143
	238	92	146

Table 15.3: Some commonly occurring isotopes.

Questions

6 Uranium has atomic number 92. Two of its common isotopes have nucleon numbers 235 and 238. Determine the number of neutrons for these isotopes.

7 There are seven naturally occurring isotopes of mercury, with nucleon numbers (and relative abundances) of 196 (0.2%), 198 (10%), 199 (16.8%), 200 (23.1%), 201 (13.2%), 202 (29.8%) and 204 (6.9%).

  a Determine the proton and neutron numbers for each isotope.

  b Determine the average relative atomic mass (equivalent to the 'average nucleon number') of naturally occurring mercury.

8 Eight different atoms are labelled A to H. Group the elements A–H into isotopes and name them using the Periodic Table in Appendix 3.

	A	B	C	D	E	F	G	H
Proton number	20	23	21	22	20	22	22	23
Nucleon number	44	50	46	46	46	48	50	51

15.5 Forces in the nucleus

As you know from earlier in this chapter, there are two kinds of particle in the nucleus of an atom: protons, which carry positive charge +*e*; and neutrons, which are uncharged. It is therefore quite surprising that the nucleus holds together at all. You would expect the electrostatic repulsions from all those positively charged protons to blow it apart. The fact that this does not happen is very good evidence for the existence of an attractive force between the nucleons. This is called the **strong nuclear force**. It only acts over very short distances (10^{-14} m), and it is what holds the nucleus together.

Why are some atoms are unstable?

In small nuclei, the strong nuclear force from all the nucleons reaches most of the others in the nucleus, but as we go on adding protons and neutrons the balance becomes much finer. The longer-range electrostatic force affects the whole nucleus, but the short-range strong nuclear force of any particular nucleon only affects those nucleons around it – the rest of the nucleus is unaffected. In a large nucleus, the nucleons are not held together so tightly and this can make the nucleus unstable. The more protons there are in a nucleus, the greater the electric forces between them and we need a few extra neutrons to help ‘keep the protons apart’. This is why heavy nuclei have more neutrons than protons. The strong interaction can explain α-decay, but not β-decay; we will look at this later in the chapter.

The proton and neutron numbers for some common nuclides are shown in Table 15.3. You can see that for light elements these two numbers are the same, but they become very different for heavy elements. Adding more neutrons helps to keep the nucleus stable, but when the number of protons is greater than 83, adding more neutrons is not enough. Elements with a proton number greater than 83 are all unstable – they undergo radioactive decay.

Most atoms that make up our world have stable nuclei; that is, they do not change as time goes by, which is quite fortunate really! However, some are less stable and give out radiation. Whether or not an atom is unstable depends on the numbers of protons and neutrons in its nucleus. Hydrogen-1 (1p), helium-4 (2p, 2n), carbon-12 (6p, 6n) and oxygen-16 (8p, 8n) are all stable – but add or subtract neutrons and the situation changes.

For example, add a neutron to helium-4 and you get helium-5, a very unstable nucleus – it undergoes radioactive emission. (There is much more about radioactive decay later in this chapter.)

Question

9 State which of the following forces act between protons and neutrons in a nucleus.

- **a** gravitational
- **b** electrostatic
- **c** strong nuclear.

15.6 Discovering radioactivity

The French physicist Henri Becquerel (Figure 15.9) is credited with the discovery of radioactivity in 1896. He had been looking at the properties of uranium compounds when he noticed that they affected photographic film– he realised that they were giving out radiation all the time and he performed several ingenious experiments to shed light on the phenomenon.

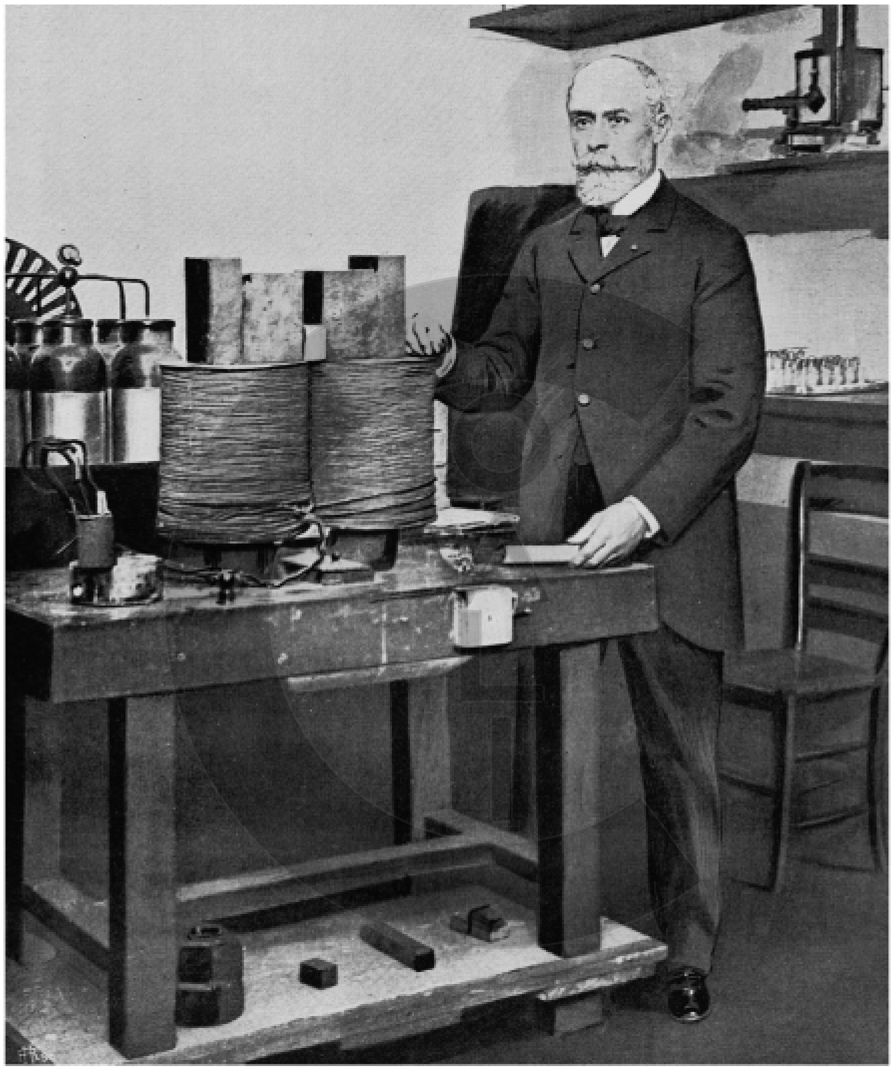


Figure 15.9: Henri Becquerel, the discoverer of radioactivity, in his laboratory. His father and grandfather had been professors of physics in Paris.

15.7 Radiation from radioactive substances

The three types of radiation commonly emitted by radioactive substances – alpha (α), beta (β) and gamma (γ)– come from the unstable nuclei of atoms. Nuclei consist of protons and neutrons, and if the balance between these two types of particles is too far to one side, or the nucleus is just too big to hold together, the nucleus may emit α- or β-radiation as a way of achieving greater stability. Gamma-radiation is usually emitted after α or β decay, to release excess energy from the nuclei.

Table 15.4 shows the basic characteristics of the different types of radiation. The masses are given relative to the mass of a proton; charge is measured in units of *e*, the elementary charge. Figure 15.10 summarises the penetrating powers of the different types of radiation.

Radiation	Symbol	Mass (relative to proton)	Charge	Typical speed
α-particle	$\alpha, {}^4_2\text{He}$	4	$+2e$	'slow' (10^6 m s^{-1})
β^--particle	$\beta, \beta^-, e, {}^{0}_{-1}e$	$\frac{1}{1840}$	$-e$	'fast' (10^8 m s^{-1})
β^+-particle	$\beta, \beta^+, e^+, {}^{0}_{+1}e$	$\frac{1}{1840}$	$+e$	'fast' (10^8 m s^{-1})
γ-ray	γ	0	0	speed of light (3×10^8 m s^{-1})

Table 15.4: The basic characteristics of ionising radiations.

Note the following points:

- α- and β-radiation are particles of matter. A γ-ray is a photon of electromagnetic radiation, similar to an X-ray. (X-rays are produced when electrons are decelerated; γ-rays are produced in nuclear reactions.)
- An α-particle consists of two protons and two neutrons; it is a nucleus of helium-4. A β^--particle is simply an electron and a β^+-particle is a positron.
- The mass of an α-particle is nearly 10 000 times that of an electron and it travels at roughly one-hundredth of the speed of a β-particle.

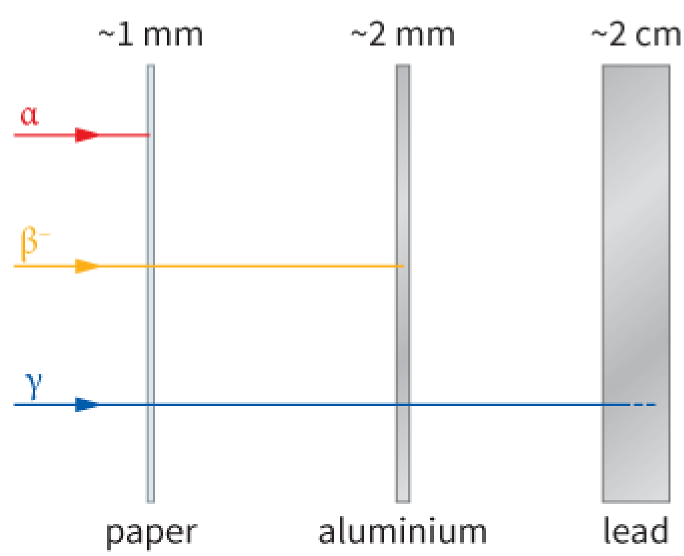


Figure 15.10: A summary of the penetrating powers of α-, β- and γ-radiations. The approximate thickness of the absorbing material is also shown.

Identification and properties of α-radiation

α-particles are relatively slow moving and large particles. They were identified as helium nuclei (He^{2+} ions) by their deflection in electric and magnetic fields (see Chapter 25). The helium nucleus, which consists of two protons and two neutrons, is extremely stable. Scientists believe that, within larger nuclei, α groups are continually forming, breaking apart and reforming. Occasionally, such a group will have enough energy to break away from the strong nuclear forces holding the mother nucleus together and will escape as an α-particle. The α-particles (which are relatively large and carry a charge) interact with atoms in the medium through which they are travelling, causing ionisation within the medium. They lose energy rapidly. This means they are not very penetrative (they are absorbed by a thin sheet of paper) and have a very short range (only a few centimetres in air).

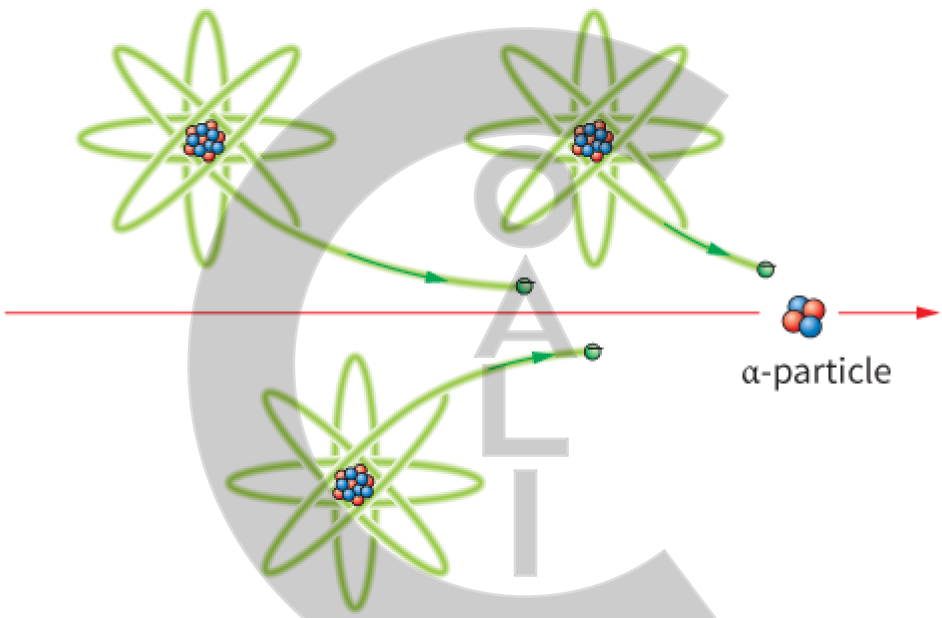


Figure 15.11: As an α-particle passes through a material, it causes ionisation of atoms.

Figure 15.12: Alpha-particle tracks show up in this photograph of a cloud chamber. Notice that the particles all travel approximately the same distance. What does this suggest?

Identification and properties of β-radiation

β-particles were identified as very fast electrons. Like α-particles, β-particles carry a charge. But, because β-particles are much smaller, they cause less ionisation and penetrate further into matter. They are absorbed by approximately one centimetre of aluminium or one millimeter of lead.

β-decay occurs when there is an imbalance of protons and neutrons in the nucleus, usually too many neutrons. A neutron will then decay into a proton (positive) and an electron (negative). The proton remains in the new nucleus and the electron is expelled at a very high velocity. However, some isotopes (such as V-48) have excess protons; because of this, a proton decays into a neutron and emits a positively charged electron or **positron**. This is known as **β^+ (beta plus) decay**. The decay of a neutron into a proton and an electron is known as **β^- (beta minus) decay**.

The positron was the first example of antimatter to be identified. It is now known that all particles have an antiparticle, which has the same mass as the particle but the opposite charge. The general term for antiparticles is **antimatter**.

What happens when matter meets antimatter?

When an antiparticle meets its particle, such as a positron meets an electron, they annihilate each other and two gamma ray photons are produced and the two masses become pure energy!

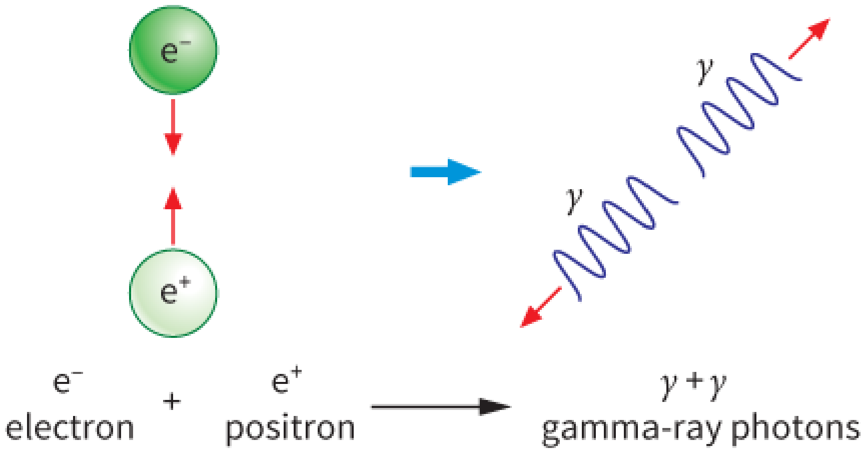


Figure 15.13: Energy is released in the annihilation of matter and antimatter.

Identification and properties of γ-radiation

γ- radiation was identified from its speed in a vacuum, 3×10^8 m s^{-1}, the speed of all electromagnetic radiation. It is very high frequency electromagnetic radiation; as such, it has no rest mass and no charge. Consequently, it does not interact with matter to the same degree as alpha or beta radiation. It produces only a small amount of ionisation and is highly penetrative – it will penetrate through several centimetres of lead.

It is generally emitted following alpha or beta decay. After the initial decay, the nucleus is left in an unstable high energy state – it will drop into a lower energy, more stable state with the emission of a gamma ray.

Question

10 a Explain why you would expect β^--particles to travel further through air than α-particles.

b Explain why you would expect β^--particles to travel further through air than through metal.

15.8 Energies in α and β decay

Look back at Figure 15.12. You were asked what conclusion could be drawn from the observation that all the α-particle tracks were the same length. The answer is quite simple: it suggests that they all have the same initial kinetic energy. This should not surprise you, as they are all the result of the similar reactions in identical nuclei. However, when we look at the energies of β-particles (both β^- and β^+) the results are quite different, as shown by the graph in Figure 15.14.

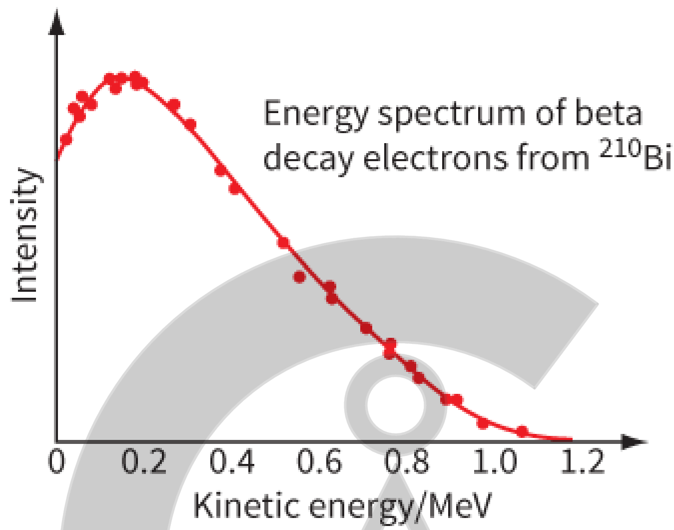


Figure 15.14: The energy spectrum for β^--decay of bismuth-210.

You will notice that the energy of the β-particles is measured in MeV (mega electronvolts). Alpha and beta particles move quickly; gamma photons travel at the speed of light. These types of radiation all have energy, but the energy of a single particle or photon is very small and far less than a joule. So we use another, much smaller unit of energy, the electronvolt, when considering the energy of individual particles or photons.

When an electron (with a charge of magnitude 1.60×10^{-19} C) travels through a potential difference, energy is transferred. The energy change W is given by:

$$W = QV = 1.60 \times 10^{-19} \times 1 = 1.60 \times 10^{-19}\ \text{J}$$

One electronvolt (1 eV) is the energy transferred when an electron travels through a potential difference of one volt.

Therefore:

$$1\ \text{eV} = 1.60 \times 10^{-19}\ \text{J}$$

There is more about the electronvolt and its use in energy calculations in Chapter 28.

The graph shows that the β-particles have a wide range of energies. One of the great physicists of the early 20th century, Wolfgang Pauli, suggested that another particle carries off some of the kinetic energy. This particle was not easy to detect–Pauli hypothesised its existence in 1930, but it was not detected until 1956. The particle has no charge and virtually no rest mass (much less than an electron) and barely interacts with matter at all. We now know that there is a steady stream of them given off by the Sun, some of which travel straight through the Earth without any interaction with it at all (which is why it's difficult to detect them!) The particle was named the antineutrino, and is now known as the **electron antineutrino**. The particle given off when a positron is emitted is called the **electron neutrino**. The symbol used for the electron neutrino is the Greek letter (nu), and the electron antineutrino is (nu bar).

COALI

15.9 Equations of radioactive decay

In radioactive decay, the nucleus changes. It is important to realise that both the nucleon number and the proton number are conserved in the reaction.

We have already established that an α-particle is a helium nucleus and can be represented as:

$${}^{4}_{2}\alpha$$

The isotope radon-222 decays by α-emission, we can describe this by the equation:

$${}^{226}_{86}\mathrm{Rn} \rightarrow {}^{218}_{84}\mathrm{Po} + {}^{4}_{2}\alpha$$

A quick glance tells us there are 222 nucleons before the decay and 218 + 4 = 222 after the decay. Similarly, there are 86 protons before and 84 + 2 = 86 after the decay.

The isotope Mn-56 decays by β^- emission:

$${}^{56}_{25}\mathrm{Mn} \rightarrow {}^{56}_{26}\mathrm{Fe} + {}^{0}_{-1}\beta + {}^{0}_{0}\nu$$

Again, it is easy to see the nucleons are conserved; however, we need to recognise that the electron is regarded as −1 proton.

Note also that an antineutrino is also emitted; interestingly, this suggests the number of particles/antiparticles are the same before and after the decay.

For the β^+ decay we look at the isotope V-48:

$${}^{48}_{23}\mathrm{V} \rightarrow {}^{48}_{22}\mathrm{Ti} + {}^{0}_{1}\beta + {}^{0}_{0}\nu$$

Once again there is a balance of proton numbers, nucleon numbers and particles/antiparticles (remember that the positron is an antiparticle).

There is another quantity that is conserved. You might expect mass to be conserved, but this is not so. For example, in the α decay equation given previously, the combined mass of the polonium nucleus and the alpha particle is slightly less than that of the original radon nucleus. The 'lost' mass has become energy – this is where the fast-moving alpha particle gets its kinetic energy. The relationship between mass m and energy E is given by Einstein's equation $E = mc^2$, where c is the speed of light in free space. So, instead of saying that mass is conserved in nuclear processes, we have to say that mass–energy is conserved. There is much more about this in Chapter 29.

Questions

In these questions, use the Periodic Table in Appendix 3 to determine the identity, or the proton number, of the relevant elements.

11 The isotope thorium-227 decays by α-emission.

Write down an equation to describe this decay and identify the element that is produced.

12 Copper-64 can decay by either β^+ or β^- emission.

Give equations for both processes and identify the resulting elements.

13 Uranium 238 decays through a series of α and β^- decays to eventually form the stable isotope lead-206 in what is known as a decay chain.

Determine the number of each type of decay in the decay chain.

15.10 Fundamental particles

Chemistry is complicated because there are billions of different molecules that can exist. The discovery of the Periodic Table simplified things because it suggested that there were roughly 92 different elements whose atoms could be arranged to make the billions of molecules. The idea that atoms are made up of just three types of particle (protons, neutrons and electrons) seemed to simplify things still more, and scientists were happy because it provided a simple explanation of a complex world.

Protons, neutrons and electrons were thought of as fundamental particles, which could not be subdivided further. However, in the middle decades of the 20th century, physicists discovered many other particles that did not fit this pattern. They gave them names such as pions, kaons, muons and so on, using up most of the letters of the Greek alphabet.

These new particles were found in two ways:

- by looking at cosmic rays, which are particles that arrive at the Earth from outer space
- by looking at the particles produced by high-energy collisions in particle accelerators (Figure 15.15).

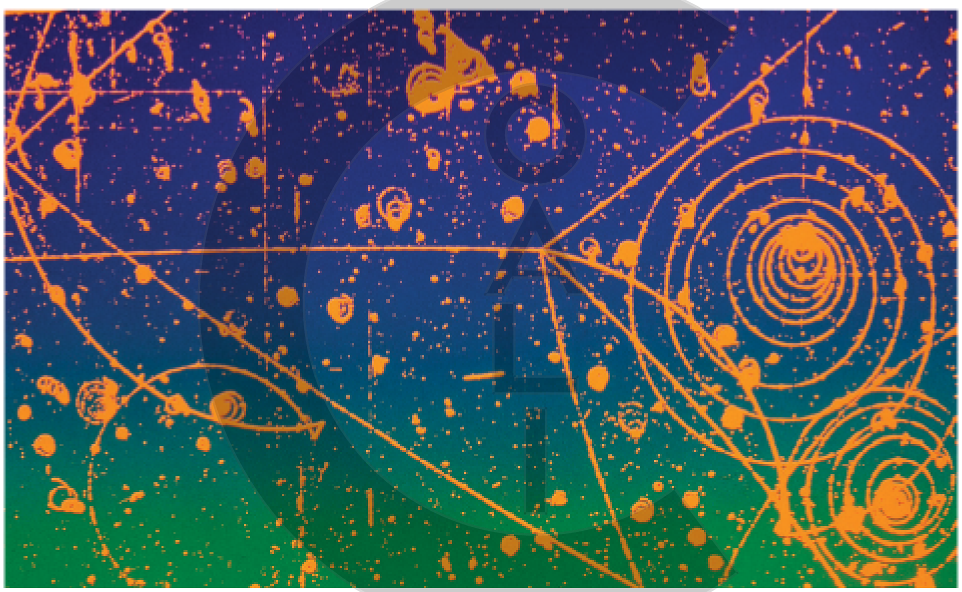


Figure 15.15: Particle tracks in a bubble chamber detector. A particle has entered from the left and then struck another particle just to the right of the centre. Four new particles fly out from the point of impact.

The discovery of dozens of new particles with masses different from those of protons, neutrons and electrons suggested that these were not fundamental particles. Various attempts were made to tidy up this very confusing picture.

15.11 Families of particles

Today, sub-atomic particles are divided into two families:

- **Leptons** such as electrons and neutrinos. These are particles that are unaffected by the strong nuclear force.
- **Hadrons** such as protons and neutrons. These are all particles that are affected by the strong nuclear force.

The word 'lepton' comes from a Greek word that means 'light' (in mass) while 'hadron' means 'bulky'. It is certainly true that protons and neutrons are bulky compared to electrons.

Leptons

Leptons are (currently) considered to be fundamental particles, although, in principle, we can never know for certain whether a particle, such as the electron, is truly fundamental; the possibility will always remain that a physicist will discover some deeper underlying structure.

Hadrons

Figure 15.16 shows the Large Hadron Collider at the CERN laboratory in Geneva. Physicists are experimenting with hadrons in the hope of finding answers to some fundamental questions about this family of particles. In 2013, they announced the discovery of the Higgs boson, a particle that was predicted 50 years earlier and which is required to explain why matter has mass.

Figure 15.16: Particle accelerators have become bigger and bigger, accelerating particles to higher and higher energies as scientists have sought to look further and further into the fundamental nature of matter. This is one of the particle detectors of the Large Hadron Collider (LHC), as it was about to be installed. The entire collider is 27 km in circumference.

Type of quark	up	down	charm	strange	top	bottom

Type of quark	up	down	charm	strange	top	bottom
Symbol	u	d	c	s	t	b
Charge	$+\frac{2}{3}e$	$-\frac{1}{3}e$	$+\frac{2}{3}e$	$-\frac{1}{3}e$	$+\frac{2}{3}e$	$-\frac{1}{3}e$
Type of antiquark	antiup	antidown	anticharm	antistrange	antitop	antibottom
Symbol	$\bar{u}$	$\bar{d}$	$\bar{c}$	$\bar{s}$	$\bar{t}$	$\bar{u}$
Charge	$-\frac{2}{3}e$	$+\frac{1}{3}e$	$-\frac{2}{3}e$	$+\frac{1}{3}e$	$-\frac{2}{3}e$	$+\frac{1}{3}e$

Table 15.5: The charges on the different types of quark and antiquark.

Quarks

To sort out the complicated picture of the hadron family of particles, Murray Gell-Mann in 1964 proposed a new model. He suggested that they were made up of just a few different particles, which he called **quarks**.

There are many surprising things about quarks. First, quarks have charges of less than the fundamental charge, *e*. However, quarks are never found outside a hadron. The quarks combine so that the resulting hadron will have a charge of *e* or a multiple of *e.*

There are six types (or 'flavours') of quark, each with an associated antiquark. Table 15.5 lists these quarks together with their charges.

In addition to the property of charge, quarks have other properties such as strangeness, charm, upness and downness. We do not need to concern ourselves about these properties; however, recognising that they exist should help you to understand how a large number of different hadrons can be made up from these half-dozen flavours of quark.

There are two ways in which quarks can combine to produce hadrons:

- three quarks make up a class of hadrons called **baryons**
- a quark and an antiquark make up a class of hadron called **mesons**.

Baryons

Examples of baryon are the proton and the neutron.

- A proton is made up of two up quarks and a down quark; proton = (uud).
- A neutron is made up of one up quark and two down quarks; neutron = (udd).

Mesons

There are many examples of mesons; here are two described:

- A π^+ meson is made up of an up quark and a down antiquark; π^+ meson = (u$\bar{d}$).
- A ϕ meson is made up of a strange quark and an antistrange quark; ϕ meson = (s$\bar{s}$)

Questions

14 a Show that the charges on the quarks making up a proton give it a charge of +1*e*.

b Show that the charges on the quarks making up a neutron give it a charge of 0.

15 A ρ-meson is made up of an up quark and an antidown quark. Calculate its charge.

16 Suggest which quarks or antiquarks make up a π^- meson.

17 Show that the ϕ-meson is neutral.

15.12 Another look at β decay

This is interesting as a hadron decays into another hadron emitting a lepton and an antilepton.

So, what happens in β decay?

The neutron is made up of an up quark and two down quarks (udd), the proton is made up of two up quarks and a down quark (uud)

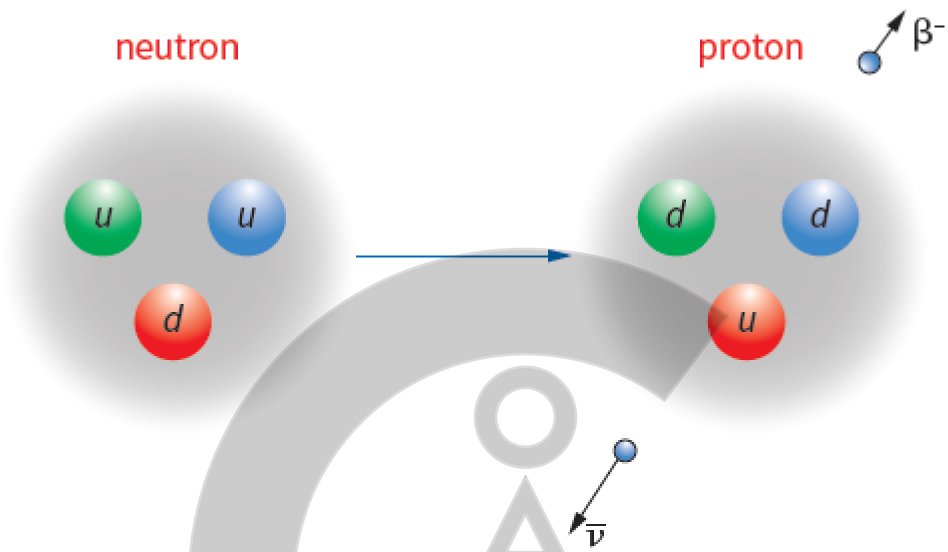


Figure 15.17: A visual representation of the change within the neutron as it decays into a proton.

The equation representing this at a quark level is:

$$d \rightarrow u + {}^{0}_{-1}e + v$$

KEY EQUATION

$$d \rightarrow u + {}^{0}_{-1}e + \overline{v}$$

The decay of a down quark into an up quark in β^- decay.

Question

18 **a** Draw a diagram similar to Figure 15.17 to show β^+ decay.

b Write an equation to describe the changes in β^+ decay.

15.13 Another nuclear force

We have met and described the strong nuclear force in some detail, as a short range force that holds the nucleons in a nucleus together. It is a force that is felt only by hadrons, not leptons.

There is a second nuclear force, known as the weak nuclear force or weak interaction. It is felt by both hadrons and leptons and, more importantly, it is the interaction that causes β-decay, in which a hadron changes to a different hadron with the emission of a lepton and an antilepton.

Questions

19 The equation $^1_1p \rightarrow ^1_0n + ^{\ 0}_{+1}\beta + \nu$ represents β^+ decay.

Use the equation to explain why the neutrino ν can have no charge and very little mass.

20 What are the differences between a proton, a positron and a photon? You can describe how their masses differ, how their charges differ or whether they are particles or antiparticles.

21 State **two** differences between hadrons and leptons.

REFLECTION

We have mentioned that quarks have different properties known as strangeness, charm, upness and downness. What is charm? What is strangeness? And what is electric charge?

Discuss your ideas in a group.

What were some of the most interesting discoveries you made while working through this chapter?

SUMMARY

The α-particle scattering experiment provides evidence for the existence of a small, massive and positively charged nucleus at the centre of the atom.
Most of the mass of an atom is concentrated in its nucleus.
The nucleus consists of protons and neutrons, and is surrounded by a cloud of electrons.
The number of protons and neutrons in the nucleus of an atom is called its nucleon number *A*.
The number of protons in the nucleus of an atom is called its proton number (or atomic number) *Z*.
Isotopes are nuclei of the same element with a different number of neutrons but the same number of protons.
Different isotopes (or nuclides, if referring to the nucleus only) can be represented by the notation $^{A}_{Z}\text{X}$, where X is the chemical symbol for the element.
There are three types of ionising radiation produced by radioactive substances: α-particles, β-particles and γ-rays.
In radioactive decay, the following quantities are conserved: proton number, nucleon number and mass–energy.
The most strongly ionising, and hence the least penetrating, is α-radiation. The least strongly ionising is γ-radiation.
Because of their different charges, masses and speeds, the different types of radiation can be identified by the effect of an electric or magnetic field.
Antimatter is material made up of antiparticles of the corresponding particles of ordinary matter. All particles have an antiparticle, which has the same mass as the particle but the opposite charge.
Quarks are particles that make up hadrons. There are six flavours of quark: up, down, strange, charm, top and bottom. Quarks have charges of $+\frac{2}{3}e$ or $-\frac{1}{3}e$.
The strong nuclear force is the force that acts between quarks and holds the nucleus together.
Leptons (such as the electron) are particles that are unaffected by the strong nuclear force.
Hadrons (such as the neutron) are particles that consist of quarks and hence are affected by the strong nuclear force.

EXAM-STYLE QUESTIONS

1 Which of the interactions is **not** possible? **[1]**

A $^{4}_{2}\text{He} + {}^{1}_{0}\text{n} \rightarrow {}^{5}_{2}\text{He}$

B $^{209}_{84}\text{Po} \rightarrow {}^{205}_{82}\text{Pb} + {}^{4}_{2}\alpha$

C $^{14}_{8}\text{O} \rightarrow {}^{14}_{7}\text{N} + {}^{0}_{1}\beta + {}^{0}_{0}\nu$

D $^{31}_{14}\text{Si} \rightarrow {}^{31}_{15}\text{N} + {}^{0}_{-1}\beta + {}^{0}_{0}\nu$

2 Hadrons are made up from quarks.

Which combination of quarks could make up a meson? **[1]**

A $dd\bar{s}$

B ssc

C sb

D sc

3 Explain why the most strongly ionising radiation (α-particles) is the least penetrating, while the least ionising (γ-rays) is the most penetrating. **[1]**

4 Before Rutherford's model, scientists believed that the atom was made up of negatively charged electrons embedded in a 'plum pudding' of positive charge that was spread throughout the atom. Explain how the α-particle scattering experiment proved that this old model of the atom was incorrect. **[3]**

5 A nucleus of strontium has a nucleon number of 90 and a proton number of 38. Describe the structure of this strontium nucleus. **[1]**

6 State the changes that take place in a nucleus when it emits an α-particle and then two β^--particles. **[5]**

7 The nuclide of iodine with a nucleon number of 131 and a proton number 53 emits a β^--particle. Write a nuclear equation for this decay. **[3]**

8 An isotope of carbon $^{14}_{6}\text{C}$ emits a β^--particle and changes into an isotope of nitrogen (N).

a What are β^--particles? **[1]**

b Write a nuclear decay equation for the decay. **[2]**

c Draw a graph with the *y*-axis representing nucleon numbers between 10 and 16 and the *x*-axis representing proton numbers between 4 and 10. On your graph, mark:

i the isotope $^{14}_{6}\text{C}$ **[2]**

ii the daughter nucleus produced in the decay. **[1]**

[Total: 6]

9 The uranium isotopes U-236 and U-237 both emit radioactive particles. A nucleus of uranium-237 may be written as $^{237}_{92}\text{U}$ and emits a β^--particle. A nucleus of uranium-236 emits an α-particle. The number of protons in a nucleus of uranium is 92.

a Describe the differences between an α-particle and a β^--particle. **[4]**

b Explain how uranium can exist in a number of different isotopes. **[2]**

c Write down the nuclear equation for the decay of U-236. **[2]**

[Total: 8]

10 Approximate values for the radius of a gold atom and the radius of a gold nucleus are 10^{-10} m and 10^{-15} m, respectively.

a Estimate the ratio of the volume of a gold atom to the volume of a gold nucleus. **[2]**

b The density of gold is 19 000 kg m^{-3}. Estimate the density of a gold nucleus, stating any assumptions that you make in your answer. **[3]**

[Total: 5]

11 The nuclide of lead $^{210}_{82}\text{Pb}$ decays in three separate stages by α and β^- emission to another lead nuclide, $^{206}_{82}\text{Pb}$.

a Describe the structure of a nucleus of $^{206}_{82}\text{Pb}$ **[2]**

b α- and β^--particles are known as **ionising radiations**. State and explain why such radiations can be described as **ionising**. **[2]**

c The two lead nuclides are shown in the graph, which plots nucleon number A against proton number Z.

Copy the graph and, on your copy, draw **three** arrows to represent one possible route for the three decays between the two isotopes of lead. Label each arrow to show whether an α-particle or a β^--particle is emitted. **[3]**

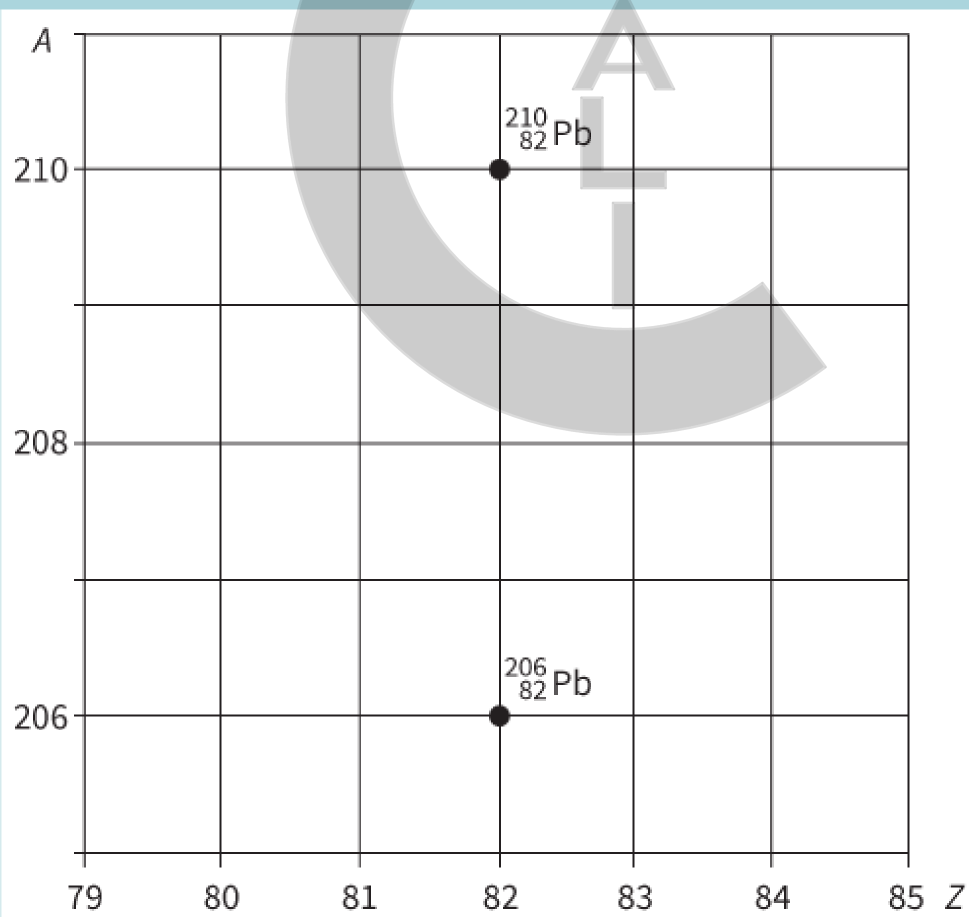


Figure 15.18

[Total: 7]

12 Geiger and Marsden carried out an experiment to investigate the structure of the atom. In this experiment, α-particles were scattered by a thin film of gold.

a When Rutherford analysed their results, what conclusions did he draw about the distribution of mass and charge in the atom? **[2]**

b Describe and explain the experimental observations that led to these conclusions. **[3]**

[Total: 5]

13 Beta decay occurs as either β^+ decay or β^- decay. An isotope of calcium Ca decays by β^+ emission into the isotope ${}^{46}_{21}\mathrm{Sc}$ and an isotope of magnesium ${}^{23}_{12}\mathrm{Mg}$ decays by β emission into the isotope ${}^{23}_{11}\mathrm{Na}$

a Copy and complete the following decay equations for the calcium and magnesium isotopes.

i decay of calcium: ${}^{46}_{20}\mathrm{Ca} \rightarrow \mathrm{Sc} + \beta^+ + \ldots$ **[1]**

ii decay of magnesium: ${}^{23}_{12}\mathrm{Mg} \rightarrow \mathrm{Na} + \beta^+ + \ldots$ **[1]**

b State what happens in each type of β decay in terms of the quark model of nucleons.

i β^- decay **[1]**

ii β^+ decay. **[1]**

c Name the force responsible for β decay. **[1]**

[Total: 5]

14 a A quark is a fundamental particle but a neutron is not. Explain what this statement means. **[1]**

b A proton and a neutron each contain three quarks, either up or down quarks.

i Copy and complete the table to show the charge on a proton and a neutron and the quarks that they contain. **[2]**

	Charge	Quarks
proton		
neutron		

Table 15.6

ii Using information from your table, suggest why some quarks must have a positive charge and some quarks a negative charge. **[2]**

c State what interaction is responsible for holding the nucleus together. **[1]**

d When a neutron decays it produces an electron and two other particles. Copy and complete the decay equation for a neutron. ${}^{0}_{1}\mathrm{n} \rightarrow \ldots\ldots\ldots\ldots\ldots\ldots$ **[2]**

e The electron and the neutron belong to different groups of particles.

Copy and complete the table to show the group of particles to which the electron and neutron belong and state the name of another member of each group. **[2]**

	Group to which it belongs	Another particle in the same group
electron		
neutron		

Table 15.7

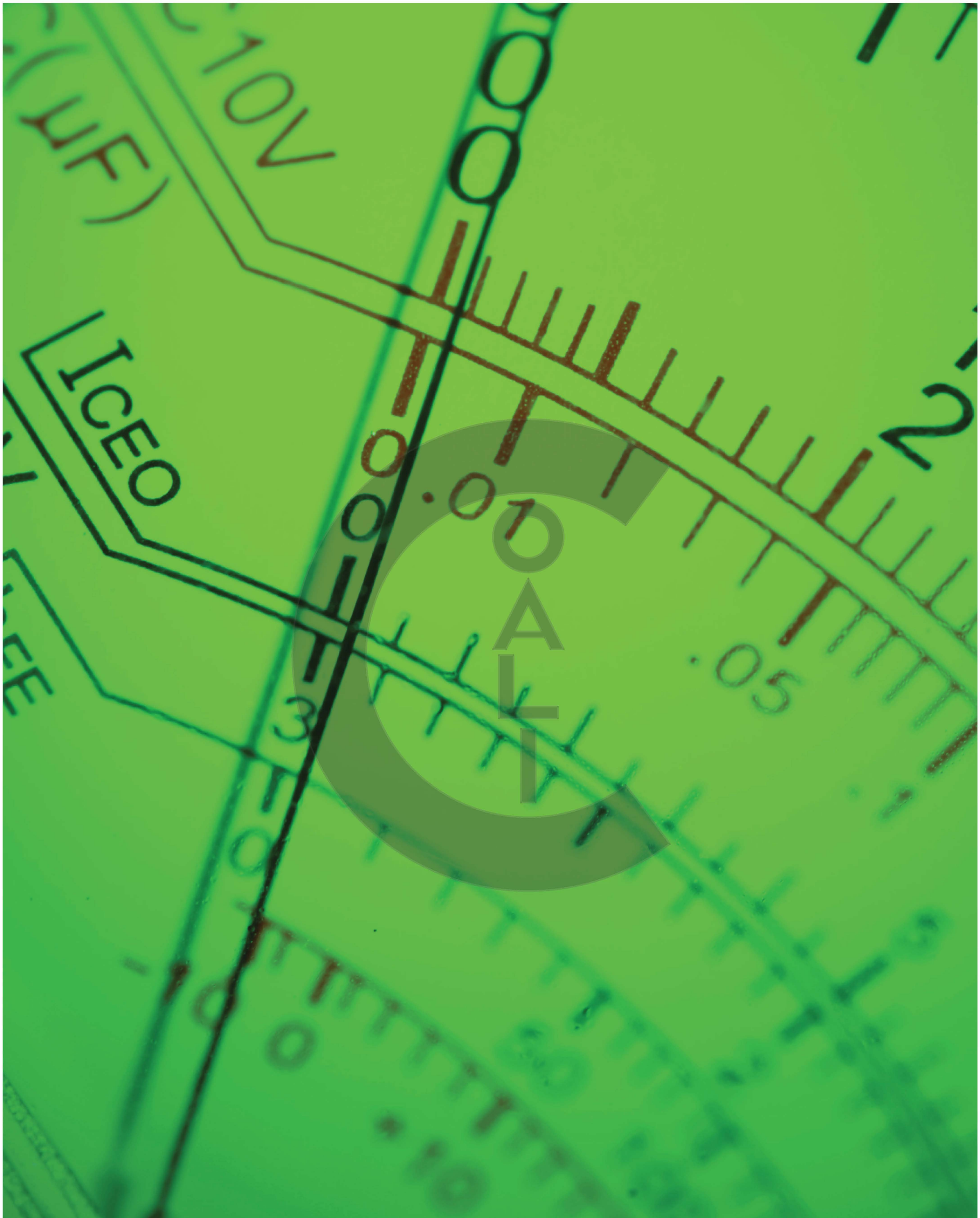
[Total: 10]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
describe the structure of the atom and understand that the nucleus is made up of protons and neutrons	15.3			
recognise that the nucleon number of a nuclide is the number of neutrons plus the number of protons	15.4			
recognise that the number of protons is known as the proton number	15.4			
use the nuclide notation $^{A}_{Z}X$ in nuclide equations	15.4			
understand and use the term isotope	15.4			
understand that there are three main types of radiation and that their penetration through matter is inversely related to their ionising ability	15.7			
recognise that there are two types of β-radiation: β^- and β^+	15.7			
understand that the positron is the antiparticle of the electron	15.7			
understand that α-particles emitted from a single isotope all have the same initial kinetic energy	15.8			
understand that β-particles emitted from a single isotope have a range of kinetic energies	15.8			
recognise that the energy spectrum of β emission led to the hypothesis that a neutrino (or antineutrino) is emitted as well as the β-particle in β decay	15.8			
understand that leptons are fundamental particles and are not affected by the strong nuclear forces	15.11			
understand that baryons (such as protons or neutrons) are made up of three quarks	15.12			

I can	See topic…	Needs more work	Almost there	Ready to move on
understand that mesons (such as pions or muons) are made up of a quark and an antiquark	15.12			
understand that the strong force holds the nucleus together and affects hadrons but not leptons.	15.13			


ICEO
.01
.05

〉Chapter P1

Practical skills at AS Level

LEARNING INTENTIONS

In this chapter you will learn how to:

- recognise random, systematic and zero errors
- calculate uncertainties in measurements made with a range of instruments
- distinguish between precision and accuracy
- estimate absolute uncertainties and combine uncertainties when quantities are added, subtracted, multiplied and divided
- set up apparatus, follow instructions and make a variety of measurements
- present data in an adequate table, produce best fit straight-line graphs and obtain the intercept and gradient
- use readings to draw conclusions from an experiment and to test a relationship
- identify limitations in an experiment and identify the main sources of uncertainty
- suggest changes to an experiment to improve accuracy and extend an investigation.

BEFORE YOU START

- What are physical properties of materials?
- What quantities do all these instruments measure: protractor, 30 cm ruler, metre rule, micrometer screw gauge, calipers, newton-meter, balance, measuring cylinder, thermometer, stopwatch, ammeter and voltmeter?
- Can you suggest, for each instrument in the list, what is its range and its smallest scale division, and suggest a simple experimental problem in using it?

P1.1 Practical work in physics

Throughout your A Level physics course, you will develop your skills in practical work, and they will be assessed at both AS & A Level. This chapter outlines the skills you will develop in the first year of the course; it includes some activities to test your understanding as you go along.

The sciences differ from most other subjects in that they involve not only theory but also practical work. The very essence of science is that theory can be tested by practical experiment. So, the ability to carry out practical exercises in a logical and scientific manner is essential.

P1.2 Using apparatus and following instructions

You need to familiarise yourself with the use of simple measuring instruments such as metre rules, balances, protractors, stopwatches, ammeters and voltmeters, and even more complicated ones such as a micrometer screw gauge and calipers.

When using measuring instruments like these you need to ensure that you are fully aware of what each division on a scale represents. If you look at Figure P1.1 you will see that on the first ruler each division is 1 mm, and on the second each division is 2 mm.

If you use instruments incorrectly, you may introduce errors into your readings. For example, when taking a reading your line of sight should always be perpendicular to the scale that you are using. Otherwise, you will introduce a parallax error; this is shown in Figure P1.2. Looking from point A the length of the rod appears to be 21 mm, from point C it appears to be 25 mm and from point B, the correct position, the length is 23 mm.

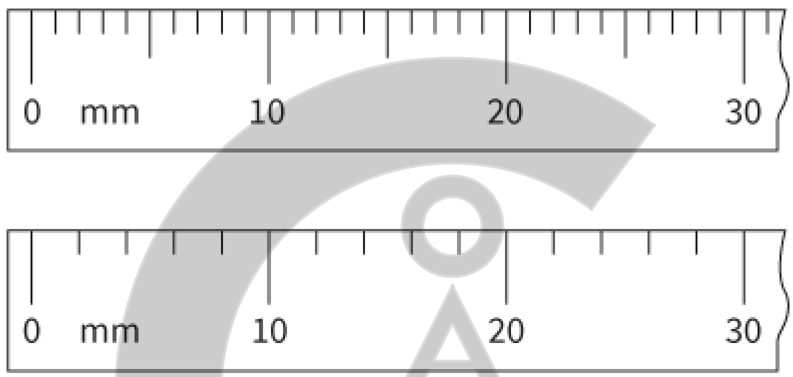


Figure P1.1: When reading from a scale, make sure that you know what each division on the scale represents.

A rule, for example, a metre rule, or a ruler, for example, an ordinary school ruler of length 30 cm, are simple measuring instruments with a smallest division of 1 mm. Other instruments have a greater precision because their smallest scale division is less than 1 mm. Here, we will look at two of them.

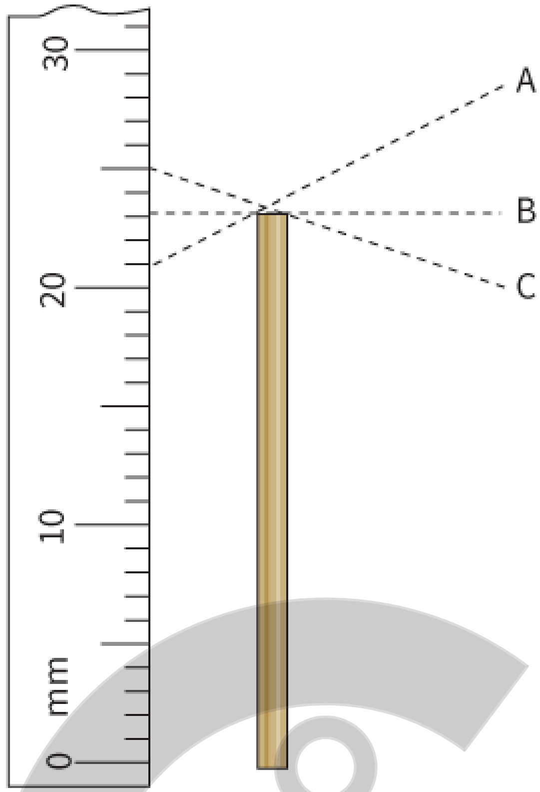


Figure P1.2: Parallax error.

Calipers

Calipers are designed to grip an object with two jaws and, in the example shown in Figure P1.3, to measure the diameter of the object. They can also be used to measure the internal diameter of a tube, for example, if the two prongs are placed inside the tube and the moving part of the calipers is adjusted until the prongs just grip the inside of the tube.

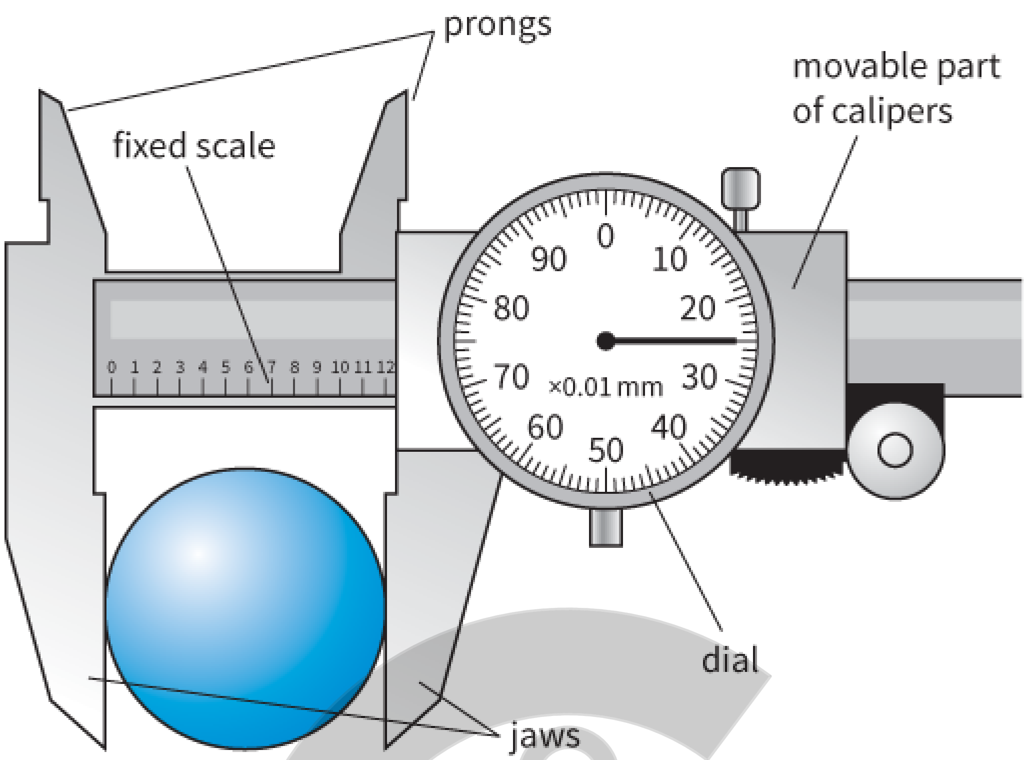


Figure P1.3: Using dial calipers.

The calipers shown in Figure P1.3 are dial calipers, although other versions such as vernier calipers are still sometimes used. As the sliding scale moves along, one rotation of the dial moves the jaws 1 mm further apart. Since the dial shown has 100 divisions, each of these divisions is $\frac{1}{100} = 0.01$ mm. The object shown has a diameter of 12 mm on the fixed scale and 25 divisions or 0.25 mm on the dial, so the diameter of the object is 12.25 mm.

Micrometer screw gauge

A micrometer screw gauge, or more simply a micrometer, is shown in Figure P1.4. This also has two scales. The main scale is on the shaft and the fractional scale is on the rotating barrel. One rotation of the barrel moves the end of the barrel 0.50 mm along the shaft. The barrel has 50 divisions so each division represents $\frac{0.50}{50} = 0.01$ mm

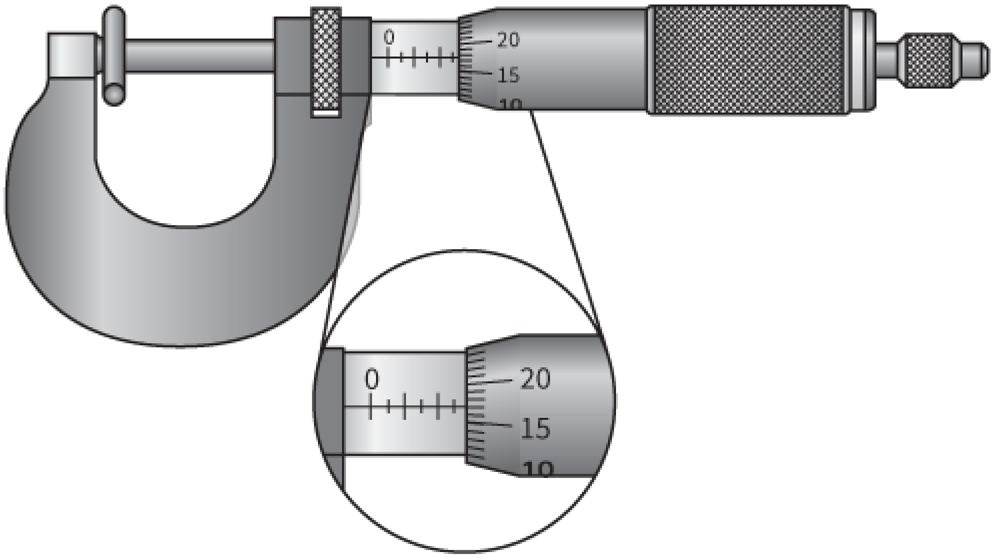


Figure P1.4: Using a micrometer screw gauge.

To use the micrometer, turn the barrel until the jaws just tighten on the object. Some micrometers have a ratchet or slip mechanism to prevent the user from tightening too hard and damaging the micrometer or object. Read the main scale to the nearest 0.5 mm, then read the number of divisions on the sleeve, which will be in 0.01 mm, and finally add the two readings. You should realise that the smallest division on the micrometer is 0.01 mm.

Before you start to use a micrometer or dial calipers, it is usual to check if there is a zero error. This is done by bringing the jaws together without any object between them. Obviously, the reading should be zero, but if the instrument is worn or has been used badly the reading may not be zero. When you have taken this zero error reading, it should be added to or subtracted from every other reading that you take with the instrument. If the jaws do not quite close to the zero mark, there is a positive zero error, and this zero error reading should be subtracted. The zero error is an example of a systematic error, which is dealt with later in this chapter.

It is also important that you become familiar with setting up apparatus. When instructions are given, the only way to become confident is through practice. You may face a variety of tasks, from setting up a pendulum system to measuring the angle at which a tilted bottle falls.

You should also learn to set up simple circuits from circuit diagrams. The most common error in building circuits comes where components need to be connected in parallel. A good piece of advice here is to build the main circuit first, and then add the components that need to be connected in parallel.

P1.3 Gathering evidence

When gathering evidence, you should take into account the range of results that you are going to obtain. If you are investigating the extension of a spring with load, for loads of between 0 N and 20 N, you should take a fair spread of readings throughout that range. For instance, six readings between 12 N and 20 N would not be sensible because you are not investigating what happens with smaller loads. Equally, taking three readings below 5 N and three more between 15 N and 20 N does not test what happens with intermediate loads.

A sensible set of readings might be at 0 N, 4 N, 8 N, 12 N, 16 N and 20 N. This covers the whole range in equal steps.

Question

1 You are investigating how the current through a resistor depends on its resistance when connected in a circuit. You are given resistors of the following values:

50Ω, 100Ω, 150Ω, 200Ω, 250Ω, 300Ω, 350Ω, 400Ω, 450Ω, 500Ω

You are asked to take measurements with just six of these resistors. Which six resistors would you choose? Explain your choice.

P1.4 Precision, accuracy, errors and uncertainties

Whenever you make a measurement, you are trying to find the true value of a quantity. This is the value you would find if your measurement was perfect. However, no measurement can ever be perfect; there will always be some **uncertainty**. Your equipment may be imperfect or your technique may be capable of improvement. So, whenever you carry out practical work, you should think about two things:

- how the equipment or your technique could be improved to give better results, with less uncertainty
- how to present the uncertainty in your findings.

As you will see later in this chapter, both of these need to be reflected in the way you present your findings.

We will first consider the **precision** of a measurement. The level of precision is high if you make several measurements of a quantity and they are all very similar. A precise measurement, when repeated, will be the same, or nearly so. However, if your measurements are spread widely around the average, they are less precise. This can arise because of practical difficulties in making the measurements.

Precision is reflected in how the results are recorded. If a distance is quoted as ‘15 m’ then it implies that it was only measured to the nearest metre, whereas if it is quoted as ‘15.0 m’ then it suggests that it was measured to the nearest 0.1 m.

Take care not to confuse precision with **accuracy**. A measurement is described as ‘accurate’ if the value obtained is close to the true value. Even if a measurement is precise, and always produces the same result, it may not be accurate because every reading may have the same error. For example, you can make very precise measurements of the diameter of a wire using a micrometer screw gauge to the nearest 0.01 mm, but every reading may be inaccurate if the gauge has a zero error.

Figure P1.5 shows two attempts at making holes in the centre of a target. Imagine that the positions of the holes represent readings, with the true value at the centre. On the left, the readings are close together so we can say that they are precise. However, they are not accurate as the average is far from the centre. In the second, the measurement can be said to be accurate as the average position of the holes is close to the centre, but the readings are not precise as the holes are spread out.

Whenever you make a measurement, you should be aware of the uncertainty in the measurement. It will often, but not always, be determined by the smallest division on the measuring instrument. On a metre rule, which is graduated in millimetres, we should be able to read to the nearest half millimetre, but beware! If we are measuring the length of a rod there are two readings to be taken, one at each end of the rod. Each of these readings has an uncertainty of 0.5 mm, giving a total uncertainty of 1 mm.

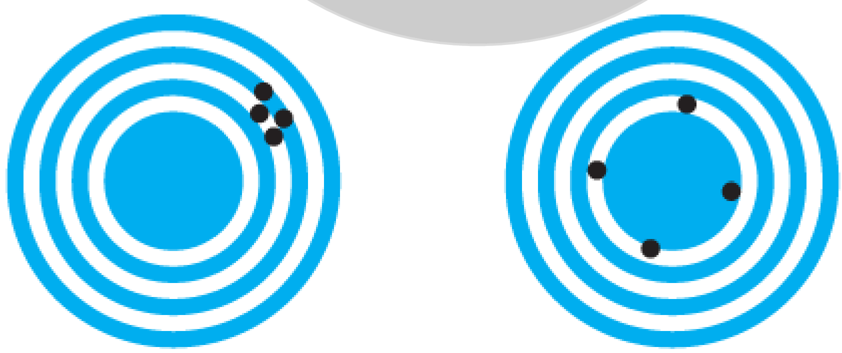


Figure P1.5: The left-hand diagram represents readings that are precise but not accurate; the right-hand diagram represents readings that are accurate but without precision.

The uncertainty will depend not only on the precision of the calibrations on the instrument you are using, but also on your ability to observe and on **errors** introduced by less than perfect equipment or poor technique in taking the observations. Here are some examples of where uncertainties might arise:

Systematic error – A spring on a force meter might, over time, become weaker so that the force meter reads consistently high. Similarly, the magnet in an ammeter might, over the years, become weaker and the needle may not move quite as far round the scale as might be expected. Parallax errors, described earlier, may be another example of a systematic error if one always looks from the same angle, and not directly from above, when taking a measurement. In principle, systematic errors can be corrected for by recalibrating the instrument or by correcting the technique being used.

Zero error – The zero on a ruler might not be at the very beginning of the ruler. This will introduce a fixed error into any reading unless it is allowed for. This is a type of systematic error.

Random errors – When a judgement has to be made by the observer, a measurement will sometimes be above and sometimes below the true value. Random errors can be reduced by making multiple measurements and averaging the results.

Good equipment and good technique will reduce the uncertainties introduced, but difficulties and judgements in making observations will limit the precision of your measurements. Here are two examples of how difficulties in observation will determine the uncertainty in your measurement.

Example 1: Using a stopwatch

Tambo has a digital stopwatch that measures to the nearest one-hundredth of a second. He is timing his sister Nana in a 100 metre race (Figure P1.6). He shows her the stopwatch, which reads 11.87 s. She records in her notebook the time 11.9 s. She explains to Tambo that he cannot possibly measure to the nearest one-hundredth of a second as he has to judge both when the starting pistol was fired and the exact moment at which she crossed the finishing line. To do this to any closer than the nearest one-tenth of a second is impossible. In addition, sometimes he will press the button too early and sometimes too late.

Figure P1.6: Uncertainty in timing using a stopwatch.

Example 2: Measuring displacement of a pendulum

Fatima is asked to measure the maximum displacement of a pendulum bob as it oscillates, as shown in Figure P1.7. She uses a ruler calibrated in millimetres. She argues that she can measure the displacement to the nearest millimetre. Joanne, however, correctly argues that she can only measure it to the nearest two millimetres, as not only is there the uncertainty at either end (0.5 mm) but she also has to judge precisely the point at which the bob is at its greatest displacement, which adds an extra millimetre to the uncertainty.

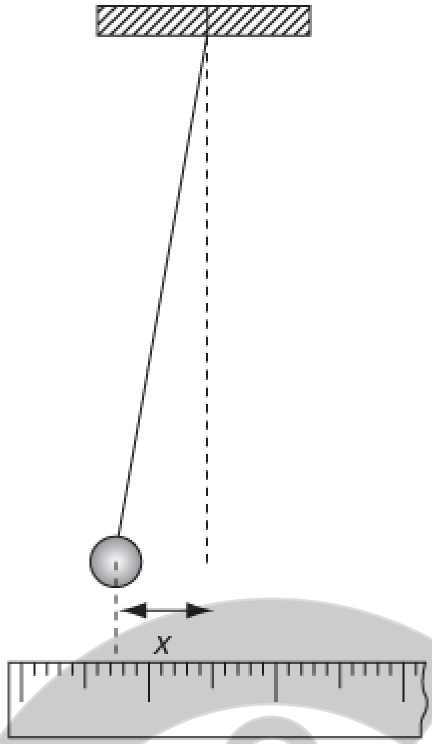


Figure P1.7: Displacement of a pendulum bob.

Questions

2 Look at Figure P1.5. Draw similar diagrams to represent:

a a target where the holes are both precise and accurate

b a target where the holes are neither precise nor accurate.

3 The position of the holes in Figure P1.5 represents attempts at measuring the position of the centre of the circle. Which one shows more random error and which shows more systematic error?

P1.5 Finding the value of an uncertainty

We have used the terms uncertainty and error; they are not quite the same thing. In general, an 'error' is just a problem that causes the reading to be different from the true value (although a zero error can have an actual value). The uncertainty, however, is an actual range of values around a measurement, within which you expect the true value to lie. The uncertainty is an actual number with a unit.

For example, if you happen to know that the true value of a length is 21.0 cm and an 'error' or problem causes the actual reading to be 21.5 cm, then, since the true value is 0.5 cm away from the measurement, the uncertainty is ±0.5 cm.

But how do you estimate the uncertainty in your reading without knowing the true value? Obviously, if a reading is 21.5 cm and you know the true value is 21.0 cm, then the uncertainty in the reading is 0.5 cm. However, you may still have to estimate the uncertainty in your reading without knowing the true value. So how is this done?

KEY IDEA

You can find the uncertainty from whichever is the largest out of:

- the smallest division on the instrument used, or
- half the range of a number of readings of the measurement.

First, it should be understood that the uncertainty is only an estimate of the difference between the actual reading and the true value. We should not feel too worried if the difference between a single measurement and the true value is as much as twice the uncertainty. Because it is an estimate, the uncertainty is likely to be given to only one significant figure. For example, we write the uncertainty as 0.5 cm and not 0.50 cm.

The uncertainty can be estimated in two ways.

Using the division on the scale – Look at the smallest division on the scale used for the reading. You then have to decide whether you can read the scale to better than this smallest division. For example, what is the uncertainty in the level of point B in Figure P1.2? The smallest division on the scale is 1 mm but is it possible to measure to better than 1 mm? This will depend on the instrument being used and whether the scale itself is accurate. In Figure P1.2, the width of the line itself is quite small but there may be some parallax error that would lead you to think that 0.5 mm or 1 mm is a reasonable uncertainty. In general, the position of a mark on a ruler can generally be measured to an uncertainty of ±0.5 mm. In Figure P1.8, the smallest division on the scale is 20 g. Can you read more accurately than this? In this case, it is doubtful that every marking on the scale is accurate and so 20 g would be reasonable as the uncertainty.

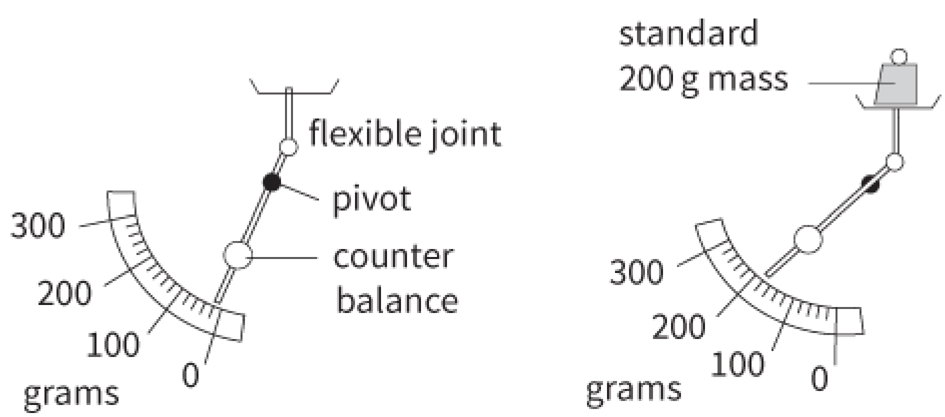


Figure P1.8: The scales on a lever-arm balance.

You need to think carefully about the smallest division you can read on any scale. As another example, look at a protractor. The smallest division is probably 1° but it is unlikely you can use a protractor to measure an angle to better than ±0.5° with your eye.

Repeating the readings – Repeat the reading several times. The uncertainty can then be taken as half of the range of the values obtained; in other words, the smallest reading is subtracted from the largest and the result is halved. This method deals with random errors made in the readings but does not account for systematic errors. This method should always be tried, wherever possible, because it may reveal random errors and gives an easy way to estimate the uncertainty. However, if the repeated readings are all the same, do not think that the uncertainty is zero. The uncertainty can never be less than the value you obtained by looking at the smallest scale division.

Which method should you actually use to estimate the uncertainty? If possible, readings should be repeated and the second method used. But if all the readings are the same, you have to try both methods!

The uncertainty in using a stopwatch is something of a special case as you may not be able to repeat the measurement. Usually, the smallest division on a stopwatch is 0.01 s, so can you measure a time interval with this uncertainty? You may know that your own reaction time is larger than this and is likely to be at least 0.1 s. The stopwatch is recording the time when you press the switch but this is not pressed at exactly the correct moment. If you do not repeat the reading then the uncertainty is likely to be at least 0.1 s, as shown in Figure P1.7. If several people take the reading at the same time, you are likely to see that 0.01 s is far too small to be the uncertainty.

Even using a digital meter is not without difficulties. For example, if a digital ammeter reads 0.35 A, then, without any more information, the uncertainty is ±0.01 A, the smallest digit on the meter. But if you look at the handbook for the ammeter, you may well find that the uncertainty is ±0.02 or 0.03 A (although you cannot be expected to know this).

WORKED EXAMPLE

1 A length is measured five times with a ruler whose smallest division is 0.1 cm and the readings obtained, in cm, are: 22.9, 22.7, 22.9, 23.0, 23.1. What is the reading obtained and the uncertainty?

Step 1 Find the average by adding the values and dividing by the number of values:

$$\frac{22.9+22.7+22.9+23.0+23.1}{5} = 22.92 \text{ cm}$$

This is written to four significant figures. At this stage, you are not sure how many figures to write in the answer.

Step 2 The maximum value is 23.1 and the minimum value is 22.7. Use these values to find half the range.

$$\text{half the range} = \frac{23.1-22.7}{2} = 0.2 \text{ cm}$$

Step 3 Check that the uncertainty calculated in Step 2 is larger than the smallest division you can read on the scale.

Step 4 Write down the average value, the uncertainty to a reasonable number of significant figures and the unit. Obviously, the last digit in 22.92 is meaningless as it is much smaller than the uncertainty; it should not be written down.

The final value is (22.9 ± 0.2) cm.

You do not usually write down the final value of the answer to a greater number of decimal places than the uncertainty. Uncertainties are usually quoted to one or perhaps two significant figures.

Questions

4 Figure P1.8 shows a lever-arm balance, initially with no mass in the pan and then with a standard 200 g mass in the pan.

Explain what types of error might arise in using this equipment.

5 Estimate the uncertainty when a student measures the length of a room using a steel tape measure calibrated in millimetres.

6 Estimate the uncertainty when a girl measures the temperature of a bath of water using the thermometer in Figure P1.9.

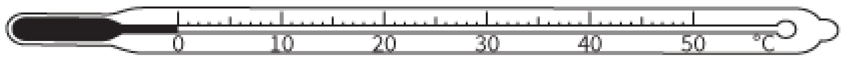


Figure P1.9: For Question 6.

7 A student is asked to measure the wavelength of waves on a ripple tank using a metre rule that is graduated in millimetres. Estimate the uncertainty in his measurement.

8 Estimate the uncertainty when a student attempts to measure the time for a single swing of a pendulum.

9 What is the average value and uncertainty in the following sets of readings? All are quoted to be consistent with the smallest scale division used.

- **a** 20.6, 20.8
- **b** 20, 30, 36
- **c** 0.6, 1.0, 0.8, 1.2
- **d** 20.5, 20.5.

P1.6 Percentage uncertainty

The uncertainties we have found so far are sometimes called absolute uncertainties, but percentage uncertainties are also very useful.

The percentage uncertainty expresses the absolute uncertainty as a fraction of the measured value and is found by dividing the uncertainty by the measured value and multiplying by 100%.

$$\text{percentage uncertainty} = \frac{\text{uncertainty}}{\text{measured value}} \times 100\%$$

For example, suppose a student times a single swing of a pendulum. The measured time is 1.4 s and the estimated uncertainty is 0.2 s. Then we have:

$$\begin{aligned}\text{percentage uncertainty} &= \frac{\text{uncertainty}}{\text{measured value}} \times 100\% \\ &= \frac{0.2}{1.4} \times 100\% \\ &= 14\%\end{aligned}$$

This gives a percentage uncertainty of 14%. We can show our measurement in two ways:

- with absolute uncertainty: time for a single swing = 1.4 s ± 0.2 s
- with percentage uncertainty: time for a single swing = 1.4 s ± 14%

(Note that the absolute uncertainty has a unit whereas the percentage uncertainty is a fraction, shown with a % sign.)

A percentage uncertainty of 14% is very high. This could be reduced by measuring the time for 20 swings. In doing so, the absolute uncertainty remains 0.2 s (it is the uncertainty in starting and stopping the stopwatch that is the important thing here, not the accuracy of the stopwatch itself), but the total time recorded might now be 28.4 s.

$$\begin{aligned}\text{percentage uncertainty} &= \frac{0.2}{28.4} \times 100\% \\ &= 0.7\%\end{aligned}$$

So measuring 20 oscillations rather than just one reduces the percentage uncertainty to less than 1%. The time for one swing is now calculated by dividing the total time by 20, giving 1.42 s. Note that, with a smaller uncertainty, we can give the result to two decimal places. The percentage uncertainty remains at 0.7%:

time for a single swing = 1.42 s ± 0.7%

Questions

10 The depth of water in a bottle is measured as 24.3 cm, with an uncertainty of 0.2 cm. (This could be written as (24.3 ± 0.2) cm.) Calculate the percentage uncertainty in this measurement.

11 The angular amplitude of a pendulum is measured as (35 ± 2)°.

- **a** Calculate the percentage uncertainty in the measurement of this angle.
- **b** The protractor used in this measurement was calibrated in degrees. Suggest why the user only feels confident to give the reading to within 2°.

12 A student measures the potential difference across a battery as 12.4 V and states that his measurement has a percentage uncertainty of 2%. Calculate the absolute uncertainty in his measurement.

P1.7 Recording results

It is important that you develop the skill of recording results in a clear and concise manner.

Generally, numerical results will be recorded in a table. The table should be neatly drawn using a ruler and each heading in the table should include both the quantity being measured and the unit it is measured in.

KEY IDEA

Each column of a table must be labelled with a quantity / unit, and, if a reading be given to the precision of the instrument, usually to the same number of decimal places. Calculated quantities may have one more significant figure than the readings used.

Table P1.1 shows how a table may be laid out. The measured quantities are the length of the wire and the current though it; both have their units included. Similarly, the calculated quantity, $\frac{1}{\text{current}}$, is included and this too has a unit, A^{-1}.

When recording your results, you need to think once more about the precision to which the quantities are measured. In the example in Table P1.1, the length of the wire might be measured to the nearest millimetre and the current might be measured to the nearest milliampere.

Note how '.0' is included in the second result for the length of the wire, to show that the measurement is to the nearest millimetre, not the nearest centimetre. Similarly the zero after the 0.35 shows that it is measured to the nearest milliampere or $\frac{1}{1000}$ of an ampere.

The third column is calculated and should show the same number of significant figures, or one more than the quantity (or quantities) it is calculated from. In this example, the current is measured to three significant figures so the inverse of the current is calculated to three significant figures.

Length of wire / cm	Current / A	$\frac{1}{\text{current}}$ / A^{-1}
10.3	0.682	1.47
19.0	0.350	2.86

Table P1.1: A typical results table.

Question

13 A ball is allowed to roll down a ramp from different starting points. Figure P1.10 shows the apparatus used. The ramp is placed at a fixed height above the floor. You are asked to measure the vertical height *h* of the starting point above the bottom of the ramp and the horizontal distance *d* the ball travels after it leaves the ramp.

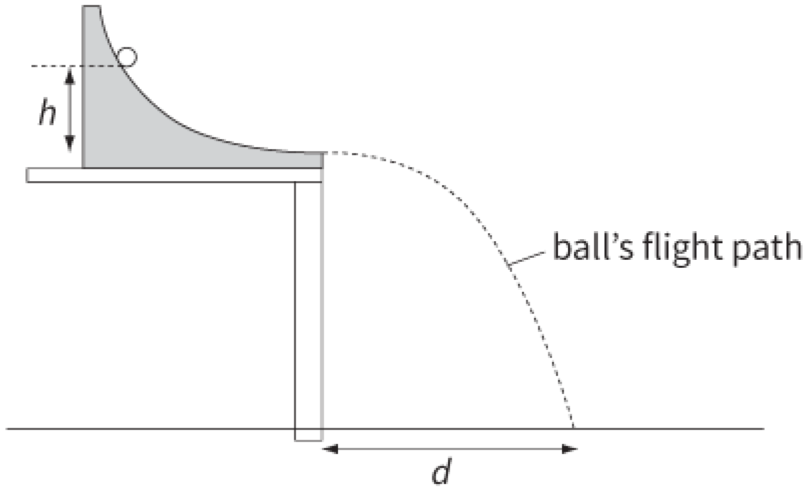


Figure P1.10: For Question 13.

You are also asked to find the square of the horizontal distance the ball travels after it leaves the ramp. Table P1.2 shows the raw results for the experiment. Copy and complete the table.

***h* / cm**	***d* / cm**	**d^2 /**
1.0	18.0	
2.5	28.4	
4.0	35.8	
5.5	41.6	
7.0	47.3	
9.0	53.6	

Table P1.2: For Question 13.

P1.8 Analysing results

When you have obtained your results, the next thing to do is to analyse them. Very often this will be done by plotting a graph.

You may be asked to plot a graph in a particular way, however, the general rule is that the variable you control or alter (the **independent variable**) is plotted on the *x*-axis and the variable that changes as a result (the **dependent variable**) is plotted on the *y*-axis.

In the example in Table P1.1, the length of the wire would be plotted on the *x*-axis and the current (or $\frac{1}{\text{current}}$) would be plotted on the *y*-axis.

You should label your axes with both the quantities you are using and their units. You should then choose your scales to use as much of the graph paper as possible. However, you also need to keep the scales simple. Never choose scales that are multiples of 3, 7, 11 or 13. Try and stick to scales that are simple multiples of 1, 2 or 5.

Plot your points carefully using small crosses; dots tend to disappear into the page and larger dots become blobs, the centre of which is difficult to ascertain.

Many, but not all, graphs you meet will be straight lines. The points may not all lie exactly on the straight line and it is your job to choose the **best fit line**. Choosing this line is a skill that you will develop through the experience of doing practical work.

Generally, there should be equal points either side of the line (but not three on one side at one end and three on the other at the other end). Sometimes, all the points, bar one, lie on the line. The point not on the line is often referred to as an anomalous point, and it should be checked, if possible. If it still appears to be off the line it might be best to ignore it and use the remaining points to give the best line. It is best to mark it clearly as 'anomalous'.

In Figure P1.11, the line chosen on the first graph is too shallow. By swinging it round so that it is steeper, it goes closer to more points and they are more evenly distributed above and below the line.

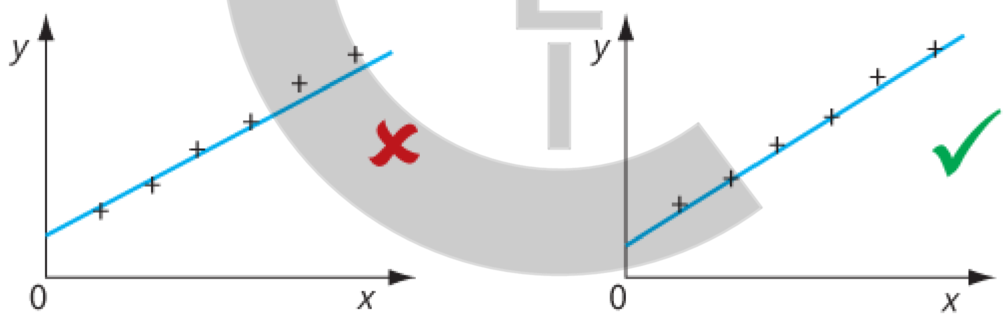


Figure P1.11

Deductions from graphs

There are two major points of information that can be obtained from straight-line graphs: the gradient and the intercept with the *y*-axis. When measuring the gradient, a triangle should drawn, as in Figure P1.12, using at least half of the line that has been drawn.

$$\text{gradient} = \frac{\text{change in y}}{\text{change in x}} = \frac{\Delta y}{\Delta x}$$

In the mathematical equation $y = mx + c$, m is equal to the gradient of the graph and c is the intercept with the *y*-axis. If c is equal to zero, the graph passes through the origin, the equation becomes $y = mx$ and we can say that *y* is proportional to *x*.

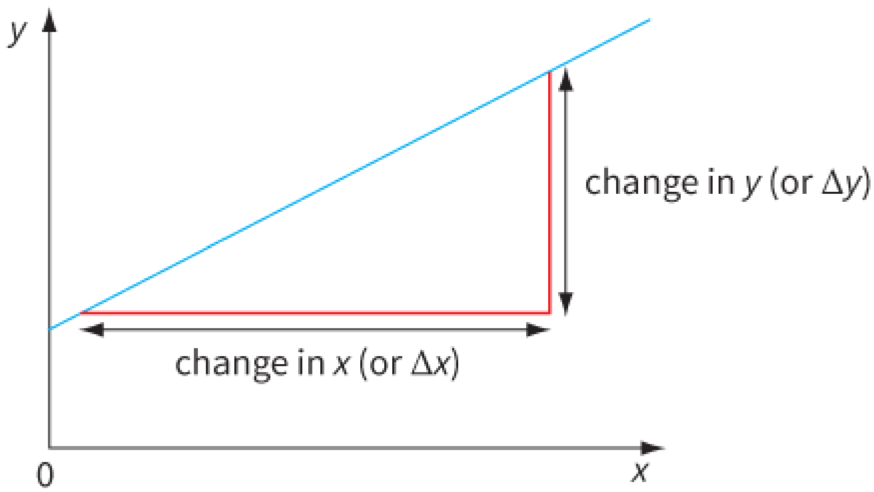


Figure P1.12

Question

14 **a** Use your results from Question 13 to plot a graph of the square of the horizontal distance d^2 (on the *y*-axis) against the height *h* (on the *x*-axis). Draw the best fit line.

b Determine the gradient of the line on your graph and the intercept with the *y*-axis. Remember, both the gradient and the intercept have units; these should be included in your answer.

Curves and tangents

You also need to develop the skill of drawing smooth curves through a set of points, and drawing tangents to those points. When drawing curves, you need to draw a single smooth curve, without any jerks or feathering. As with a straight line, not every point will lie precisely on the curve, and there should be a balance of points on either side.

In the first graph of Figure P1.13, the student has joined each of the points using a series of straight lines. This should never be done. The second graph is much better, although there is some feathering at the left-hand side, as two lines can be seen. The third graph shows a well-drawn curve.

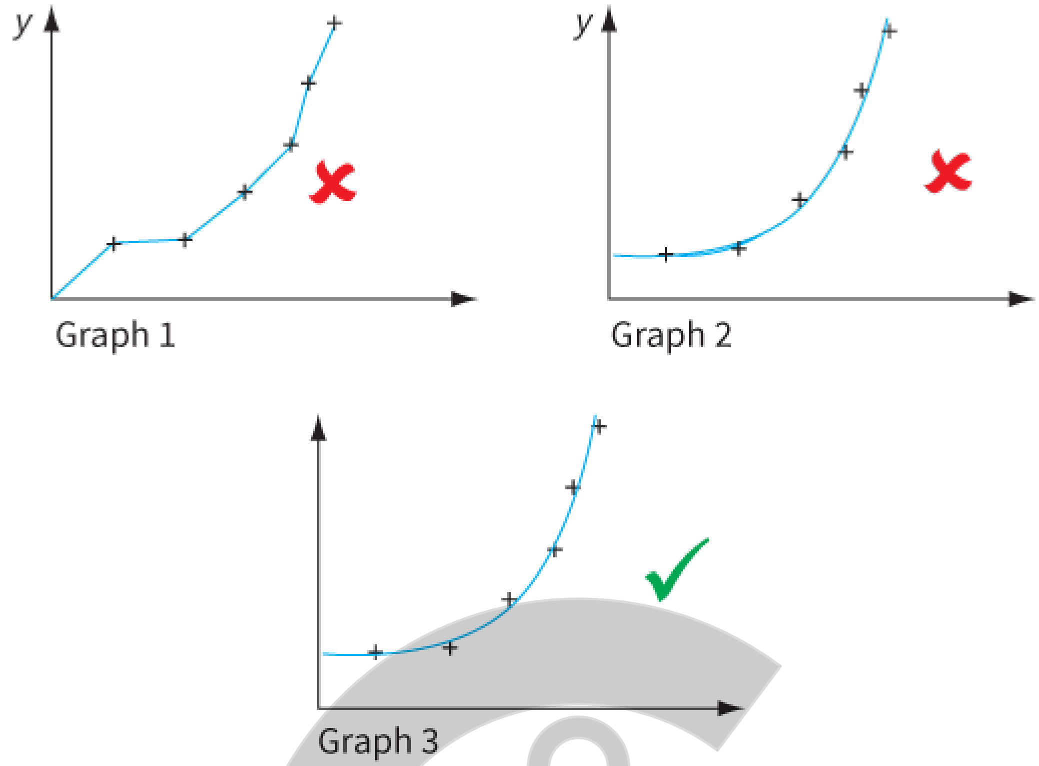


Figure P1.13: For Question 14.

P1.9 Testing a relationship

The readings from an experiment are often used to test a relationship between two quantities, typically whether two quantities are proportional or inversely proportional.

You should know that if two quantities *y* and *x* are directly proportional:

- the formula that relates them is $y = kx$, where *k* is a constant
- if a graph is plotted of *y* against *x* then the graph is a straight line through the origin and the gradient is the value of *k*.

If the two quantities are inversely proportional then $y = \frac{k}{x}$ and a graph of *y* against $\frac{1}{x}$ gives a straight line through the origin.

These statements can be used as a basis for a test. If a graph of *y* against *x* is a straight line through the origin, then *y* and *x* are directly proportional. If you know the values of *y* and *x* for two points, you can then calculate two values of *k* with the formula $y = \frac{k}{x}$ and see whether these two values of *k* are actually the same. But what if the points are not exactly on a straight line or the two values of *k* are not exactly the same – is the relationship actually false or is it just that errors caused large uncertainties in the readings?

Later in this chapter, we will look at how to combine the uncertainties in the values for *y* and *x* to find an uncertainty for *k*. However, you can use a simple check to see whether the difference in the two values of *k* may be due to the uncertainties in the readings. For example, if you found that the two values of *k* differ by 2% but the uncertainties in the readings of *y* and *x* are 5%, then you cannot say that the relationship is proved false. Indeed, you are able to say that the readings are consistent with the relationship.

You should first write down a criterion for checking whether the values of *k* are the same. This criterion is just a simple rule you can invent for yourself and use to compare the two values of *k* with the uncertainties in the readings. If the criterion is obeyed you can then write down that the readings are consistent with the relationship.

KEY IDEA

Write down a criterion.

Calculate the percentage difference between two values of the constant.

Compare the percentage difference with the percentage uncertainty in one of the variables.

Write a conclusion as to whether the criterion is obeyed or not.

Criterion 1

A simple approach is to assume that the percentage uncertainty in the value of *k* is about equal to the percentage uncertainty in either *x* or *y*; choose the larger percentage uncertainty of *x* or *y*.

You first look at the percentage uncertainty in both *x* and *y* and decide which is bigger. Let us assume that the larger percentage uncertainty is in *x*. Your stated criterion is then that 'if the difference in the percentage uncertainty in the two values of *k* is less than the percentage uncertainty in *x*, then the readings are consistent with the relationship'.

If the percentage difference in *k* values is less than the percentage uncertainty in *x* (or *y*), the readings are consistent with the relationship.

KEY IDEA

If the percentage difference in *k* values is less than the percentage uncertainty in *x* (or *y*), the readings are consistent with the relationship.

Criterion 2

Another criterion is to state that the *k* values should be the same within 10% or 20%, depending on the experiment and the uncertainty that you think sensible. It is helpful if the figure of 10% or 20% is related to some uncertainty in the actual experiment.

Whatever criterion you use, it should be stated clearly and a clear conclusion given. The procedure to check whether two values of *k* are reasonably constant is as follows:

- Calculate two values of the constant *k*. The number of significant figures chosen when writing down these values should be equal to the least number of significant figures in the data used. If you are asked to justify the number of significant figures you give for your value of *k*, state the number of significant figures that *x* and *y* were measured to and that you will choose the smallest. Do not quote your values of *k* to one significant figure to make them look equal when *x* and *y* were measured to two significant figures.
- Calculate the percentage difference in the two calculated values of *k*. It is worthwhile using one more significant figure in each actual value of *k* than is completely justified in this calculation.
- Compare the percentage difference in the two values of *k* with your clearly stated criterion. You could compare your percentage difference in *k* values with the larger of the percentage differences in *x* and *y*.

WORKED EXAMPLES

1 A student investigates the depth *D* of a crater made when ball-bearings of different diameters *d* are dropped into sand. He drops two ball bearings from the same height and measures the depth of the craters using a 30 cm ruler. The results are shown in Table P1.3.

Diameter of ball bearing *d* / mm	Depth of the crater *D* / mm	*D*/d
5.42 ± 0.01	36 ± 2	6.64
3.39 ± 0.01	21 ± 2	6.19

Table P1.3: For Worked example 2.

It is suggested that the depth *D* of the crater is directly proportional to the diameter *d* of the ball-bearing, that is:

$$D = kd \text{ or } \frac{D}{d} = k$$

Do the readings support this hypothesis?

Step 1 Calculate the values of $k = \frac{D}{d}$ These values are shown in the third column in Table P1.3, although they should only be given to two significant figures as values of *D* are given to two significant figures and values of *d* to three significant figures. The more precise values for *k* are to be used in the next step.

Step 2 Calculate the percentage difference in the *k* values. The percentage difference is:

$$\frac{0.45}{6.19} \times 100\% = 7.2\%$$

So the *k* values differ by 7% of the smaller value.

Step 3 State a criterion and check it.

'My criterion is that, if the hypothesis is true, then the percentage difference in the *k* values will be less than the percentage uncertainty in *D*. I chose *D* as it obviously has the higher percentage uncertainty.'

The uncertainty in the smaller measurement of *D* can be calculated as:

$$\text{uncertainty in } D = \frac{2}{21} \times 100\% = 9.5\%$$

The percentage difference in the *k* values is less than the uncertainty in the experimental results; therefore, the experiment is consistent with the hypothesis.

Of course, we cannot say for sure that the hypothesis is correct. To do that, we would need to greatly reduce the percentage uncertainties.

3 A student obtains data shown in Table P1.4.

x / cm	d / cm
2.0	3.0
3.5	8.0

Table P1.4: For Worked example 3.

The first reading of *x* was found to have an uncertainty of ±0.1. Do the results show that *d* is proportional to *x*?

Step 1 Calculate the ratio of $\frac{d}{x}$ in both cases:

$$\left(\frac{d}{x}\right)_1 = 1.50 \quad \left(\frac{d}{x}\right)_2 = 2.29$$

Step 2 Calculate how close to each other the two ratios are: 2.29 − 1.50 = 0.79

So the two values of $\left(\frac{d}{x}\right)$ are $\frac{0.79}{1.5} = 53\%$ different.

Step 3 Compare the values and write a conclusion.

The uncertainty in the first value of *x* is 5% and, since the percentage difference between the ratios of 53% is much greater, the evidence does not support the suggested relationship.

Questions

15 A student obtains the following data for two variables *T* and *m* (Table P1.5).

T / s	*m* / kg
4.6	0.90
6.3	1.20

Table P1.5: Data for Question 15.

The first value of *T* has an uncertainty of ±0.2 s. Do the results show that *T* is proportional to *m*?

16 A student obtains the following values of two variables *r* and *t* (Table P1.6).

r / cm	*t* / s
6.2	4.6
12.0	6.0

Table P1.6: Data for Question 16.

The first value of *r* has an uncertainty of ±0.2 cm, which is much greater than the percentage uncertainty in *t*. Do the results show that t^2 is proportional to *r*?

P1.10 Combining uncertainties

When quantities are combined, for example, multiplied or divided, what is the uncertainty in the final result?

KEY IDEA

If quantities are added or subtracted, add absolute uncertainties.

If quantities are multiplied or divided, add percentage uncertainties.

Suppose that quantity $A = 1.0 \pm 0.1$ and that $B = 2.0 \pm 0.2$, so that the value of $A + B$ is 3.0. The maximum likely value of $A + B$, taking into account the uncertainties, is 3.3 and the minimum likely value is 2.7. You can see that the combined uncertainty is ±0.3, so $A + B = 3.0 \pm 0.3$. Similarly, $B - A = 1.0 \pm 0.3$.

When quantities are added or subtracted, their absolute uncertainties are added. A simple example is measuring the length of a stick using a millimetre scale. There is likely to be an uncertainty of 0.5 mm at both ends, giving a total uncertainty of 1.0 mm.

When quantities are multiplied or divided, combining uncertainties is a little more complex. To find the combined uncertainty in this case, we add the percentage uncertainties of the two quantities to find the total percentage uncertainty.

Remember, you always add uncertainties; never subtract.

Where quantities are:

- added or subtracted, then add **absolute** uncertainties
- multiplied or divided, then add **percentage** or fractional uncertainties.

WORKED EXAMPLES

3 The potential difference across a resistor is measured as (6.0 ± 0.2) V, while the current is measured as (2.4 ± 0.1) A.

Calculate the resistance of the resistor and the absolute uncertainty in its measurement.

Step 1 Find the percentage uncertainty in each of the quantities:

$$\text{percentage uncertainty in p.d.} = \frac{0.2}{6.0} \times 100\% = 3.3\%$$

$$\text{percentage uncertainty in current} = \frac{0.1}{2.4} \times 100\% = 4.2\%$$

Step 2 Add the percentage uncertainties. Sum of uncertainties:

(3.3 + 4.2)% = 7.5%

Step 3 Calculate the resistance value and find the absolute uncertainty:

$$R = \frac{V}{I} = \frac{6.0}{2.4} = 2.5\Omega$$

7.5% of 2.5 = 0.1875 ≈ 0.2 Ω

The resistance of the resistor is 2.5 ± 0.2 Ω.

When you calculate the uncertainty in the square of a quantity, since this is an example of multiplication, you should double the percentage uncertainty. For example, if $A = (2.0 \pm 0.2)$ cm, then A has a percentage uncertainty of 10% so $A^2 = 4.0\ \text{cm}^2 \pm 20\%$; or giving the absolute uncertainty, $A^2 = (4.0 \pm 0.8)\ \text{cm}^2$.

Questions

17 You measure the following quantities:

$A = (1.0 \pm 0.4)$ m

$B = (2.0 \pm 0.2)$ m

$C = (2.0 \pm 0.5)$ m s^{-1}

$D = (0.20 \pm 0.01)$ s

Calculate the result and its uncertainty for each of the following expressions. You may express your uncertainty either as an absolute value or as a percentage.

a $A + B$

b $B - A$

c $C \times D$

d $\frac{B}{D}$

e A^2

f $2 \times A$

g the square root of ($A \times B$). (Recall that the square root of x can be written as $x^{1/2}$.)

18 A rifle bullet is photographed in flight using two flashes of light separated by a time interval of (1.00 ± 0.02) ms. The first image of the bullet on the photograph appears to be at a position of (22.5 ± 0.5) cm on a scale underneath the flight path. The position of the second image is (37.5 ± 0.7) cm on the same scale. Find the speed of the bullet and its absolute uncertainty.

P1.11 Identifying limitations in procedures and suggesting improvements

No experiment is perfect and the ability to see weaknesses in the experimental setup and the techniques used is an important skill. You should also take the opportunity to think of ways to improve the experimental technique, thereby reducing the overall percentage uncertainty.

In this topic, we will look at five experiments and discuss **problems** that might arise and the **improvements** that might be made to overcome them. It will help if you try out some of the experiments yourself so that you get a feel for the methods described. The table for each experiment is a summary of ideas that you might use in your answer.

Experiment 1: Ball-bearings and craters

In Worked example 2, the student dropped a ball-bearing of diameter *d* into sand and measured the depth *D* of the crater produced. He dropped two ball-bearings of different diameters from the same height and measured the depth of the crater using a 30 cm ruler. Table P1.7 suggests some of the problems with the simple method used, together with some improvements.

Suggestion	Problem	Improvement
1	'Two results are not enough to draw a valid conclusion.'	'Take more results and plot a graph of *D* against *d*.'
2	'The ruler is too wide to measure the depth of the crater.'	'Use a knitting needle and mark the sand level on the needle and then measure with a ruler.'
3	'There may be a parallax error when measuring the top level of the crater.'	'Keep the eye parallel to the horizontal level of the sand, or use a stiff card.'
4	'It is difficult to release the ball-bearing without giving it a sideways velocity, leading to a distorted crater.'	'Use an electromagnet to release the ball.'
5	'The crater lip is of varying height.'	'Always measure to the highest point.'

Table P1.7: Suggestions for improving Experiment 1.

It is worth making some points regarding these suggestions.

1 This is a simple idea, but it is important to explain how the extra results are to be used. In this case, a graph is suggested – alternatively the ratio $\frac{D}{d}$ could be calculated for each set of readings.

2 The problem is clearly explained. It is not enough to just say that the depth is difficult to measure.

3 It is not enough to just say 'parallax errors'. We need to be specific as to where they might occur. Likewise, make sure you make it clear where you look from when you suggest a cure.

4 There is no evidence that this will affect the crater depth, but it is a point worthy of consideration.

5 An interesting point: does the crater depth include the lip or is it just to the horizontal sand surface? Consistency in measurement is what is needed here.

Experiment 2: Timing with a stopwatch

Many years ago, Galileo suggested that heavy and light objects take the same time to fall to the ground from the same height, as illustrated in Figure P1.14. Imagine that you want to test this hypothesis.

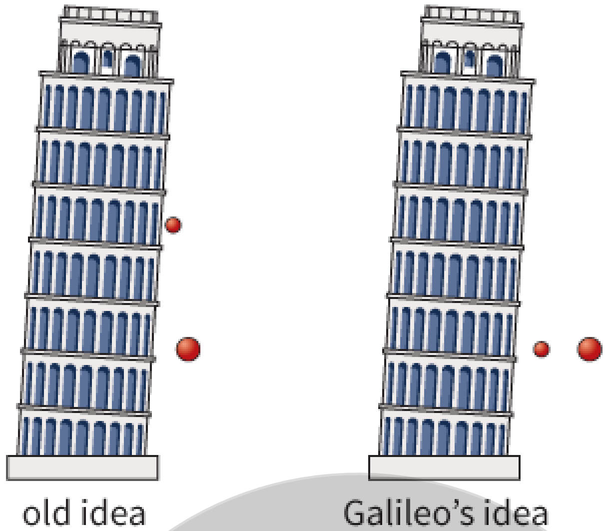


Figure P1.14: It was believed that Galileo dropped two different masses from the top of the Leaning Tower of Pisa to prove his idea. But people now think it probably didn't happen. He just did a 'thought experiment'.

This is an experiment you can do yourself with two objects and a stopwatch, or even a digital wrist watch or a cell phone with a timing app. Drop two different objects, for example two stones, and measure the time they take to fall the same distance to the ground.

Of course, the times you obtain are likely to be different. Does this prove Galileo wrong? You can test the relationship and establish whether your readings are consistent with his hypothesis. However, if you improve the experiment and reduce the uncertainties, the conclusion will be much more useful.

When you consider improving an experiment, first consider any practical difficulties and possible sources of inaccuracy. Write them down in detail. Do not just write, for example, 'reaction time' or 'parallax error'. It is always a good idea to start with the idea that more readings need to be taken, possibly over a greater range (for example, in this case, if the masses of the stones were almost equal). Table P1.8 gives other possibilities.

Problem	**Improvement**
'Taking readings for just two masses was not enough.'	'I should use a great range of different masses and plot a graph of the average time to fall to the ground against the mass of the object.'
'It was difficult to start the stopwatch at the same instant that I dropped the stone and to stop it exactly as it hit the ground. I may have been late because of my reaction time.'	'Film the fall of each stone with a video camera which has a timer in the background. When the video is played back, frame by frame, I will see the time when the ball hits the ground on the timer. '(Alternatively, you can use light gates connected to a timer to measure the time electronically. You should draw a diagram, explaining that the timer starts when the first light gate is broken and stops when the second is broken.)
'My hand was not steady and so I may not have dropped the stones from exactly the same height each time.'	'Use iron objects which hang from an electromagnet. When the current in the electromagnet is switched off, the object falls.' (A diagram would help – see Figure P.15.)

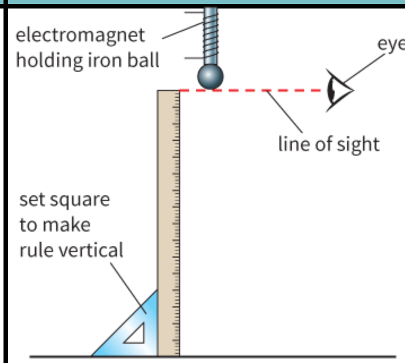
Problem	Improvement	
'The heavier stone was larger in size and it was important that the bottom of each stone started at the same height. There may have been parallax error.'	'Clamp a metre rule vertically and start the bottom of each stone at exactly the top of the ruler each time. To avoid parallax error, I will make sure my line of sight is horizontal, at right angles to the rule.' (A diagram will show this clearly – see Figure P1.15.)	 **Figure P1.15:** Using an electromagnet to release iron objects. The line of sight is clearly shown.
'The times that I measured were very short – not much greater than my reaction time – so reaction time had a great effect.'	'Increase the distance of fall so that the times are larger. This will make the uncertainty in each time measurement smaller in proportion to the time being measured.'	

Table P1.8: Suggestions for improving Experiment 2.

Question

19 Use a stopwatch and a metre rule to measure the average speed as an object falls from a table to the ground. What are the difficulties and how might they be reduced? Some of the suggestions will be the same as those in Experiment 2, but you should also consider difficulties in measuring the distance to the ground and how they can be avoided. Remember, rules have battered ends and the ends may not be at 0 and 100 cm.

Experiment 3: Timing oscillations

In physics, the study of oscillations is of great importance. Indeed, the observation of a pendulum led Galileo to study time intervals and allowed pendulum clocks to be developed.

One skill you will need to develop is finding the time for an oscillation. Figure P1.16 shows a simple pendulum and one complete oscillation. The pendulum is just a small weight, the bob, which hangs on a string.

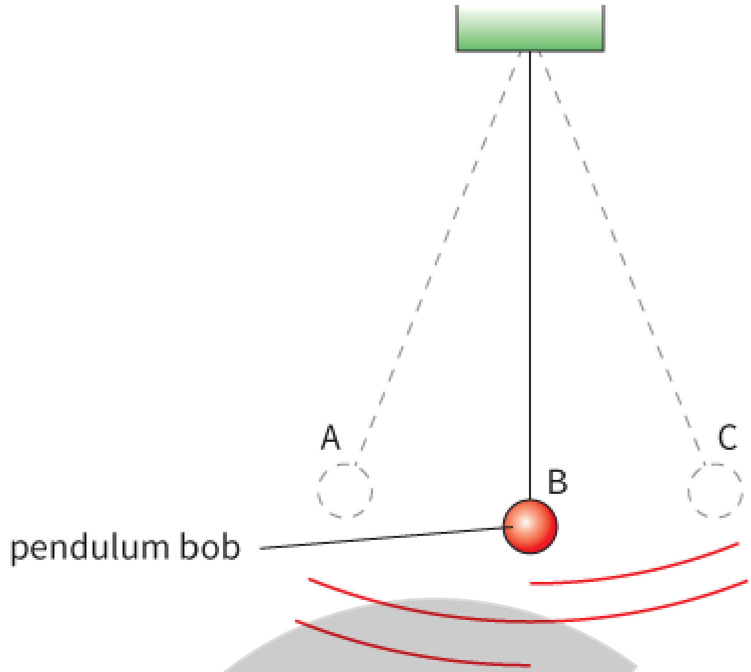


Figure P1.16: One complete oscillation is either from A to C and then back to A, or from B to C then back to B, then to A and back to B, as shown.

Figure P1.16 shows that one complete oscillation can be measured in two ways. Which way is better? In fact, the second way is better. This is because it is difficult to judge exactly when the pendulum bob is at the end of its swing. It is easier to start timing when the bob is moving quickly past a point; this happens in the middle of the swing. To time from the middle of the swing, you should use a **fiducial** mark. This can be a line on the bench underneath the bob at the centre of the swing, or it can be another object in the laboratory that appears to be in line with the bob when it hangs stationary, as seen from where you are standing. As long as you do not move your position, every time the bob passes this point it passes the centre.

Another way to reduce the uncertainty in the time for one oscillation is to time more than one swing, as explained in the topic on percentage uncertainty.

A simple practical task is to test the hypothesis that the time for one oscillation T is related to the length l of a simple pendulum by the formula $T^2 = kl$, where k is a constant.

What difficulties would you face and what are possible improvements? Table P1.9 gives some possibilities.

Problem	**Improvement**
'Taking readings for just two lengths was not enough.'	'Use more than two lengths and plot a graph of the average time squared against the length of the string.'
'It was difficult to judge the end of the swing.'	'Use a fiducial mark at the centre of the oscillation as the position to start and stop the stopwatch.' 'Use an electronic timer placed at the centre of the oscillation to measure the time.' 'Make a video of the oscillation with a timer in the background and play it back frame by frame.'
'The oscillations died away too quickly.'	'Use a heavier mass which swings longer.'

Problem	Improvement
'The times were too small to measure accurately, as my reaction time was a significant fraction of the total time.'	'Use longer strings.' 'Time 20 rather than 10 oscillations.'
'It was difficult to measure the length to the centre of gravity of the weight accurately.'	'Use a longer string so any errors are less important.' 'Measure the length to the top of the weight and use a micrometer to measure the diameter of the bob and add on half the diameter to the length of the string.'

Table P1.9: Suggestions for improving Experiment 3.

Question

20 Hang a mass from a spring or from a rubber band. Use a stopwatch to time the mass as it oscillates up and down. Measure the time for just one oscillation, the time for 10 oscillations and the time for 20 oscillations. Repeat each reading several times. Use your readings to find the time for one complete oscillation and the uncertainty in each time. Draw up a table to show the problems of such measurements and how to reduce them.

Experiment 4: Using force meters

You need to be able to read instruments, estimating the uncertainty, looking for sources of error and trying to improve their use. One such instrument is a force meter or newton-meter, shown in Figure P1.17.

In this experiment, the block is pulled using the force meter to find the force F needed to make a block just start to move. An extra mass is added on top of the block to see whether the relationship $F = km$ is obeyed, where m is the total mass of the block and k is a constant.

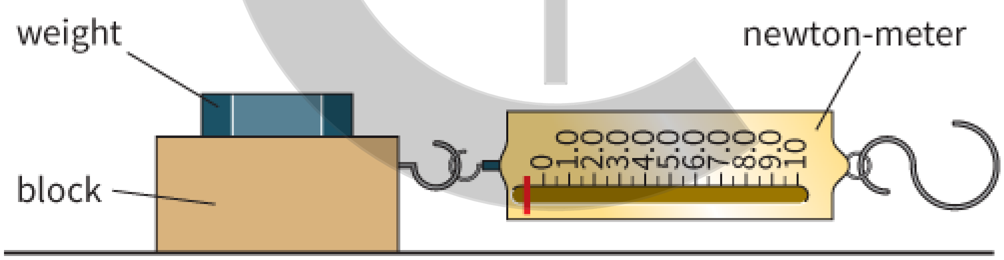


Figure P1.17: A newton-meter, just before it pulls a block along the bench. Look closely at Figure P1.17. When reading the meter, the uncertainty is the smallest scale division on the meter, unless one can reasonably read between the markings. This is difficult and so an uncertainty of 0.5 N, the smallest scale division, is reasonable.

Another problem in using the meter is that it reads less than zero before it is pulled. It needs a small force to bring the meter to zero. This is a zero error and all the actual readings will be too large by the same amount. This is probably because the meter was adjusted to read zero when hanging vertically and it is now being used horizontally.

Fortunately, the meter can be adjusted to read zero before starting to pull.

Table P1.10 describes the problems that may be encountered with this experiment, together with suggested improvements.

Problem	Improvement
'Taking readings for just two masses was not enough.'	'Use more than two masses and plot a graph of the force against the mass.'
'It was difficult to zero the newton-meter used horizontally.'	'Use a force sensor and computer.' 'Use a pulley and string to connect a tray to the block. Then tip sand onto a tray until the block starts to move. The weight of the sand and tray is then the force.'
'The reading of *F* was very low on the scale and gave a large percentage uncertainty.'	'Use heavier masses on top of the block.'
'The block starts to move suddenly and it is difficult to take the reading as this happens.'	'Video the experiment and play back frame by frame to see the largest force.' 'Use a force sensor and computer.'
'Different parts of the board are rougher than others.'	'Mark round the block with a pencil at the start and put it back in the same place each time.'

Table P1.10: Suggestions for improving Experiment 4.

Question

21 If you grip the bulb of a thermometer gently in your fingers, the reading rises to a new value. The reading will be different depending on whether you cover the bulb entirely or only partially with your fingers.

A laboratory thermometer can be used to measure the increase in temperature.

a Suggest a value for the uncertainty in such a reading. (You may need to look at some different thermometers.)

b Describe how you would test whether the temperature rise is proportional to the area of the bulb covered by your fingers You can take the surface area of the bulb to be 1 cm^2 and when you cover half of the bulb the area covered is 0.5 cm^2. The exact value of the surface area is not important; just the ratio is important.

c Suggest difficulties with this experiment, and how it might be improved. One problem with a thermometer is that it takes time for the reading to rise. What can you do about this?

Experiment 5: Electrical measurements

Electrical experiments have their own problems. Figure P1.18 shows an apparatus used to test the hypothesis that the resistance R of a wire is related to its length l by the formula $R = kl$, where k is a constant. The current is kept constant and the voltmeter reading is taken at two different values of l, for l = 0.30 m and 0.50 m.

What problems are likely to arise when using this apparatus? Table P1.11 identifies some possible problems with this experiment, and some suggestions for improvement.

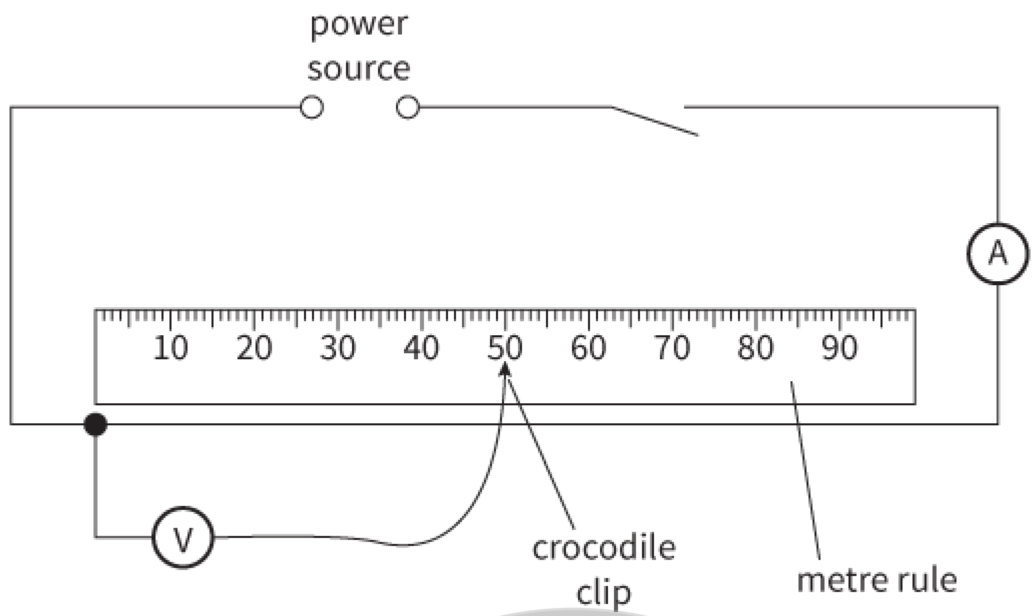


Figure P1.18: Apparatus used to check the hypothesis $R = kl$.

REFLECTION

Without looking at your textbook, produce a list of the problems and improvements that can be encountered in mechanics experiments, light experiments and electrical experiments.

Check your list against someone else's list.

Problem	Improvement
'Taking readings for just two lengths was not enough.'	'Use more than two lengths and plot a graph of the voltmeter reading against the length.' 'Calculate more than just two values of k.'
'Difficult to measure the length of the wire as the clips have width and I don't know where inside they grip the wire.'	'Use narrower clips.' 'Solder the contacts onto the wire.'
'The scale is not sensitive enough and can only measure to 0.05 V.'	'Use a voltmeter that reads to 0.01 V.' 'Use a digital voltmeter.'
'The values of voltage are small, particularly at 0.30 m.'	'Use a larger current so that when l = 0.50 m the voltmeter reading is at the top of the scale.'
'The voltmeter reading fluctuates because of contact resistance.'	'Clean the wires with wire wool first.'
'Other factors may have changed the resistance; for example, the temperature may have increased because of the current.'	'Wait a long time until the wire has reached a constant temperature.' 'Use smaller currents, but with a more sensitive voltmeter.'

Table P1.11: Suggestions for improving Experiment 5.

SUMMARY

A **precise** reading is one in which there is very little spread about the mean value.

The **uncertainty** in a reading is an estimate of the difference between the reading and true value of the quantity being measured.

A **systematic error** cause readings to differ from the true value by a consistent amount each time the reading is made.

Random errors cause readings to vary around the mean value in an unpredictable way from one reading to another.

A **zero error** is caused when an instrument gives a non-zero reading when the true value of the quantity is zero.

Find the uncertainty from the largest of the smallest division on the instrument used or half the range of a number of readings of the same measurement.

Each column of a table must be labelled with a quantity / unit, and, if a reading be given to the precision of the instrument, usually to the same number of decimal places. Calculated quantities may have one more significant figure than the readings used.

The **independent** variable is the one that the experimenter alters or selects.

The **dependent** variable is the quantity that changes as a result of the independent variable being altered by the experimenter.

$$\text{gradient} = \frac{\text{change in y}}{\text{change in x}}$$

Use a large triangle to show the values used in calculating the gradient.

In testing a relationship, write down a criterion. Calculate the percentage difference between two values of the constant. Compare the percentage difference with the percentage uncertainty in one of the variables and write a conclusion as to whether the criterion is obeyed or not

If quantities are added or subtracted, then add absolute uncertainties. If quantities are multiplied or divided, add percentage uncertainties.

A **problem** is a difficulty you experience during the experiment.

An **improvement** is a suggestion that will reduce the problem. You should have experience of a range of these problems and improvements. For more details, consult the Practical Workbook.

EXAM-STYLE QUESTIONS

1 Quantity P has a fractional uncertainty ***p***. Quantity Q has a fractional uncertainty ***q***.

What is the fractional uncertainty in $\frac{P^2}{Q^3}$? **[1]**

A $p - q$

B $p + q$

C $2p - 3q$

D $2p + 3q$

2 The p.d. *V* across a wire of length *l* is given by the formula $V = \frac{4l\rho I}{d^2}$ where *d* is the diameter of the wire, ρ is the resistivity and there is a current *I* in the wire.

Which quantity provides the largest contribution to the percentage uncertainty in ***V***? **[1]**

	Quantity	Value of quantity	Absolute uncertainty
A	*l* / cm	250	±10
B	*d*/mm	1.4	±0.1
C	ρ/Ω m	1.5×10^{-8}	$\pm0.2 \times 10^{-8}$
D	*I* / A	2.0	±0.2

Table P1.12

3 What is the uncertainty in the following sets of readings? All of them are written down to the smallest division on the instrument used in their measurement.

a 24.6, 24.9, 30.2, 23.6 cm **[1]**

b 2.66, 2.73, 3.02 s **[1]**

c 24.0, 24.0, 24.0 g **[1]**

[Total: 3]

4 Electrical experiments usually involve the reading of meters such as the voltmeters shown.

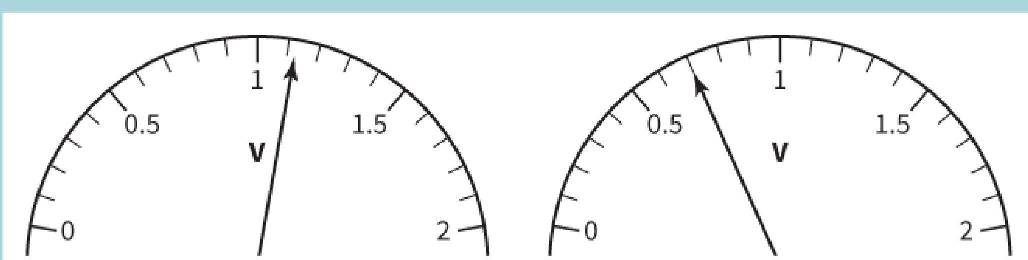


Figure P1.19

a What is the reading shown by each voltmeter, and the uncertainty in each reading? **[2]**

b The voltmeters show the readings obtained when they were connected across two wires that were identical apart from their different lengths. The current in each wire was 0.500 A and the length l of the wire was 30.0 cm in the right diagram and 50.0 cm in the left diagram.

Use the scale readings to test the hypothesis that the resistance R of the wire is proportional to length l. Consider the effect of the uncertainties on your conclusion. **[4]**

[Total: 6]

5 This apparatus can be used to test the hypothesis that T, the time taken for a ball to roll down a plane from rest, is related to the distance s by the formula $T^2 = ks$, where k is a constant.

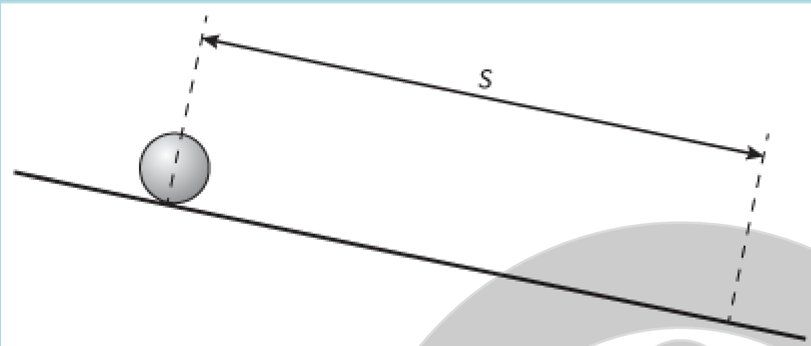


Figure P1.20

The ball is timed using a stopwatch over two different values of s.

Suggest problems with the experiment and how they might be overcome. You should consider problems in measuring the distance as well as the time. Also note what happens to the ball; it may not roll in the way that you expect. **[8]**

Questions 6–8 are designed to illustrate some aspects of practical questions. They are not formal practical questions as, ideally, you should perform the experiment yourself and take some readings. This helps you to see the problems.

6 An experiment explores the relationship between the period of a vibrating spring and the mass m in a pan holder. The student is instructed to set up the apparatus as shown here, with a mass of 200 g in the pan.

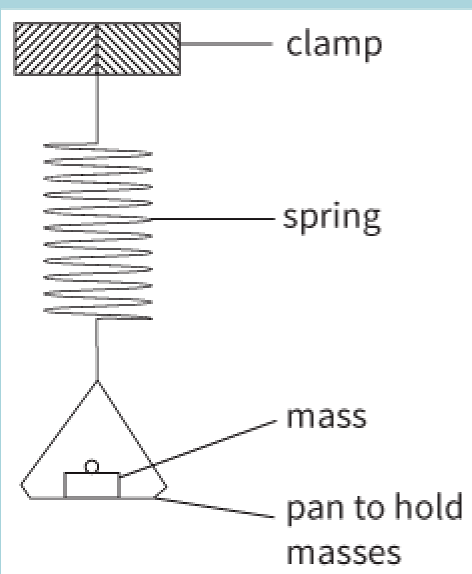


Figure P1.21

The student is then told to move the pan downwards by approximately 1 cm and to release it so that it vibrates in a vertical direction.

The student is asked to record the time taken for 20 oscillations of the spring, and then to repeat the procedure, using masses between 20 g and 200 g until She has six sets of readings. Columns are provided in the table for $\sqrt{m}$ and *T*, the period of the pendulum.

This table shows the readings taken by a student with the different masses.

Mass / g	**Time for 20 oscillations / s**	$\sqrt{m}$	***T***
20	12.2		
50	15.0		
100	18.7		
150	21.8		
200	24.5		
190	24.0		

Table P1.13

a Copy the table and include values for $\sqrt{m}$ and *T*. **[2]**

b Plot a graph of *T* on the *y*-axis against $\sqrt{m}$ on the *x*-axis. Draw the straight line of best fit. **[4]**

c Determine the gradient and y-intercept of this line. **[2]**

d The quantities *T* and *m* are related by the equation:

$$T = C + k\sqrt{m}$$

where C and *k* are constants.

Find the values of the two constants C and *k*. **Give** appropriate units. **[2]**

[Total: 10]

7 A student releases a toy car to roll down a ramp, as shown.

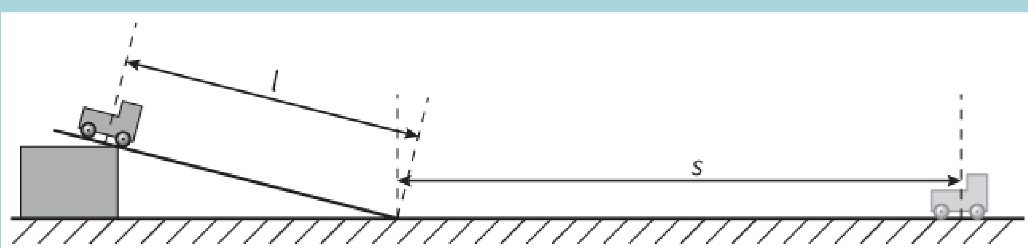


Figure P1.22

The student measures the distance *l* from the middle of the car as it is released to the bottom of the ramp and the distance *s* travelled along the straight section before the car stops. He also measures the time *t* taken to travel the distance *s.* He then repeats the experiment using a different value of *l*.

The student obtained readings with *l* = 40 and 60 cm, taking each reading for *s* and *t* twice. The readings were:

l = 40.0 cm: values for *s* were 124 and 130 cm; values for *t* were 4.6 and 4.8 s

l = 60.0 cm: values for *s* were 186 and 194 cm; values for *t* were 4.9 and 5.2 s.

a For the smaller value of l, obtain a value for:

i the average value of *s* **[1]**

ii the absolute and percentage uncertainty in the value of *s* **[2]**

iii the average value of *t* **[1]**

iv the absolute and percentage uncertainty in the value of *t*. **[2]**

b i For both values of l, calculate the average speed *v* of the car along the straight section of track using the relationship $v = \frac{s}{t}$ **[1]**

ii **Justify** the number of significant figures that you have given for your values of *v*. **[1]**

c i It is suggested that *s* is proportional to l. Explain whether the readings support this relationship. **[2]**

ii (HARDER) It is suggested that v^2 is proportional to l. Explain whether the readings support this relationship. **[2]**

d Describe **four** sources of uncertainty or limitations of the procedure for this experiment. **[4]**

e Describe **four** improvements that could be made to this experiment. You may suggest the use of other apparatus or different procedures. **[4]**

[Total: 20]

8 This apparatus shows a resistor in some water.

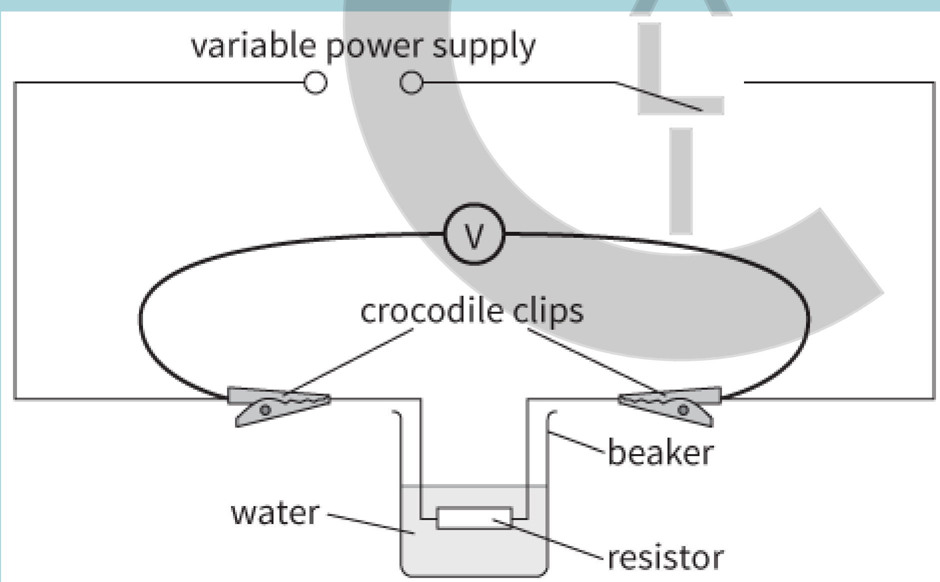


Figure P1.23

A student measures the rise in temperature θ of the water in 100 s using two different values of voltage.

The student wrote:

‘When the voltage was set at 6.0 V, the rise in temperature of the water in 100 s was 14.5 °C. The voltmeter reading decreased by about 0.2 V during the experiment, and so the final voltmeter reading was 5.8 V.

‘The reading fluctuated from time to time by about 0.2 V. The smallest scale division on the thermometer was 1 °C, but I could read it to 0.5 °C. I did not have time to repeat the reading.

'When the voltage was set at 12.0 V, the rise in temperature in 100 s was 51.0 °C and the voltage was almost the same at the end, but fluctuated by about 0.2 V.'

a Estimate the percentage uncertainty in the measurement of the first voltage. **[1]**

b It is suggested that θ is related to V according to the formula $\theta = kV^2$, where k is a constant.

- **i** Calculate **two** values for k. Include the units in your answer. **[2]**
- **ii** Justify the number of significant figures you have given for your value of k. **[1]**
- **iii** Explain whether the results support the suggested relationship. **[1]**

c Describe **four** sources of uncertainty or limitations of the procedure for this experiment. **[4]**

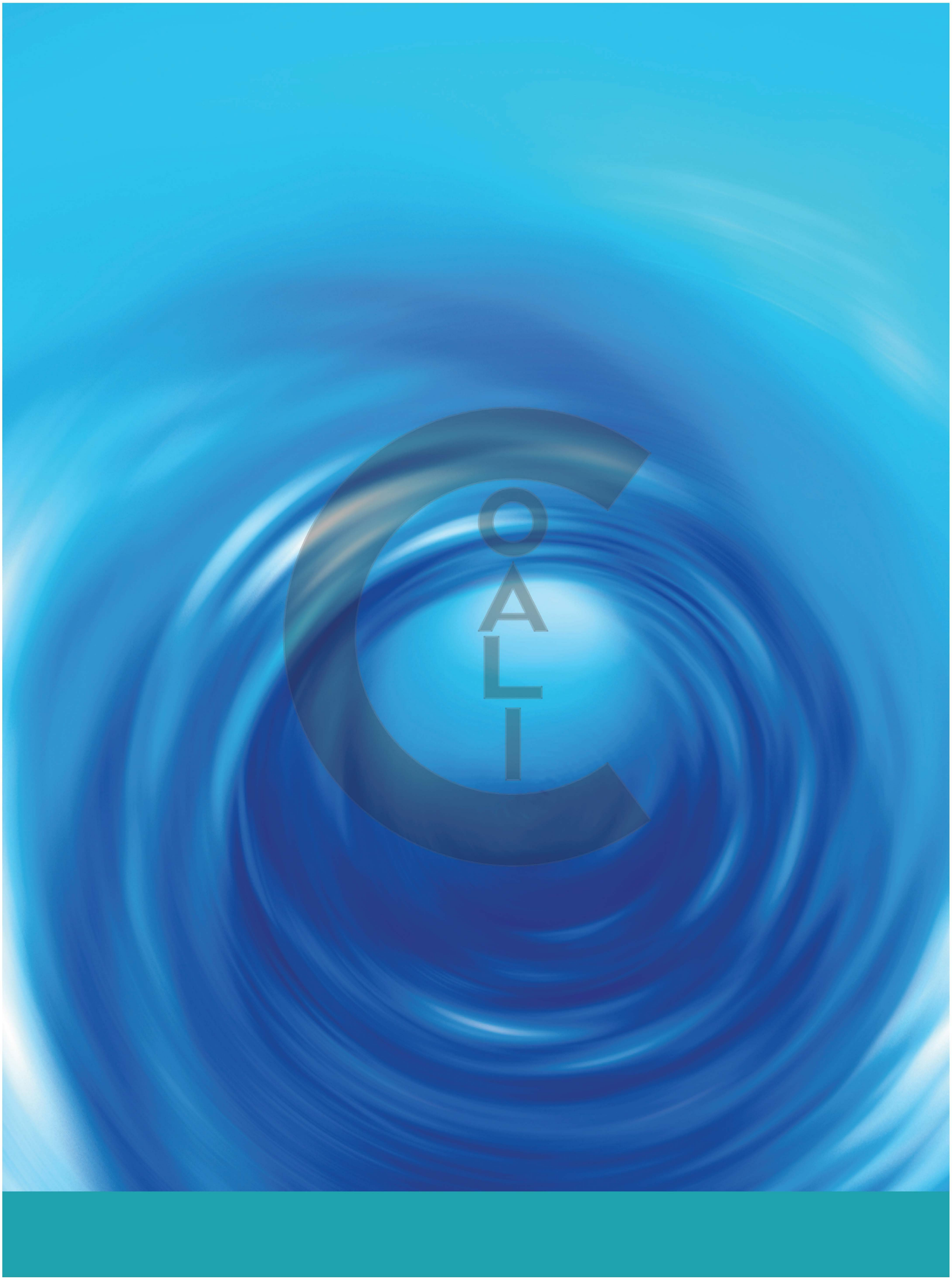
d Describe **four** improvements that could be made to this experiment. You may suggest the use of other apparatus or different procedures. **[4]**

[Total: 13]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
recognise random, systematic and zero errors	P1.2, P1.4			
distinguish between precision and accuracy	P1.4			
estimate absolute uncertainties	P1.10			
combine uncertainties	P1.5, P1.6			
make a variety of measurements and present data in an adequate table, produce best fit straight-line graphs and obtain the intercept and gradient	P1.7			
use readings to draw conclusions and to test a relationship	P1.8			
identify limitations in procedure and the main sources of uncertainty	P1.11			
suggest changes to an experiment to improve accuracy and extend the investigation.	P1.11			


COALI

〉Chapter 16

Circular motion

LEARNING INTENTIONS

In this chapter you will learn how to:

- express angular displacement in radians
- solve problems using the concept of angular speed
- describe motion along a circular path as due to a perpendicular force that causes a centripetal acceleration
- recall and use equations for centripetal acceleration.

BEFORE YOU START

Explain to a friend why a stationary child's ball on the floor of a train starts rolling towards the front of the train when the driver brakes and the train slows down. Write down your explanation and discuss it with a partner.

MOVING IN CIRCLES

The racing car in Figure 16.1 shows two examples of circular motion. The car's wheels spin around the axles, and the car follows a curved path as it speeds round the bend.

Part of the skill of the driver is to judge the maximum speed at which the car can take the corner without the car sliding out of control. Consider the path the car takes around the corner as the car goes round the bend. The driver feels himself thrown towards the outside of the curve. This is caused by his inertia. His body 'wants' to go on in a straight line at the same constant speed, in the same way that the inertia of the child's ball in a braking train 'wants', to carry on in a straight line at constant speed and, therefore, rolls towards the front of the train.

Figure 16.1: Circular motion: the car's wheels go round in circles as the car itself follows a curved path.

16.1 Describing circular motion

Many things move in circles, such as:

- the wheels of a car or a bicycle
- the Earth in its (approximately circular) orbit round the Sun
- the hands of a clock
- a spinning DVD in a laptop
- the drum of a washing machine.

Sometimes, things move along a path that is part of a circle. For example, the car in Figure 16.1 is travelling around a bend in the road that is an arc of a circle.

Circular motion is different from the straight-line motion that we have discussed previously in our study of kinematics and dynamics in Chapters 1 to 6. However, we can extend these ideas of dynamics to build up a picture of circular motion.

Around the clock

The second hand of a clock moves steadily round the clock face. It takes one minute for it to travel all the way round the circle. There are 360° in a complete circle and 60 seconds in a minute. So the hand moves 6° every second. If we know the angle θ through which the hand has moved from the vertical (12 o'clock) position, we can predict the position of the hand.

In the same way, we can describe the position of any object as it moves around a circle simply by stating the angle θ of the arc through which it has moved from its starting position. This is shown in Figure 16.2.

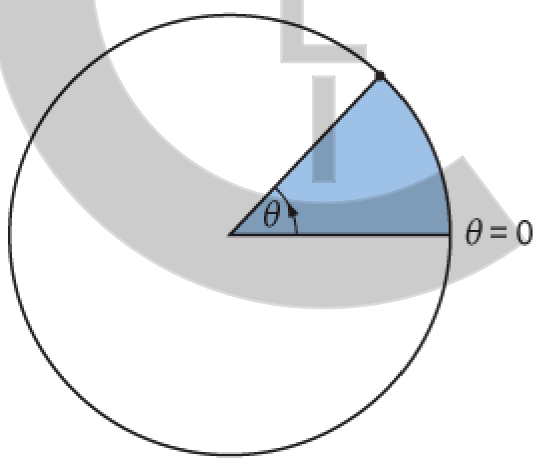


Figure 16.2: To know how far an object has moved round the circle, we need to know the angle θ.

The angle θ through which the object has moved is known as its **angular displacement**. For an object moving in a straight line, its position is defined by its displacement *s*, the **distance** it has travelled from its starting position. The corresponding quantity for circular motion is angular displacement θ, the **angle** of the arc through which the object has moved from its starting position.

Question

1 a By how many degrees does the angular displacement of the hour hand of a clock change each hour?

b A clock is showing 3.30. Calculate the angular displacements in degrees from the 12.00 position of the clock to:

i the minute hand

ii the hour hand.

16.2 Angles in radians

When dealing with circles and circular motion, it is more convenient to measure angles and angular displacements in units called radians rather than in degrees.

If an object moves a distance *s* around a circular path of radius *r* (Figure 16.3a), its angular displacement θ in **radians** is defined as follows:

$$\text{angle (in radians)} = \frac{\text{length of arc}}{\text{radius}}$$

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

Since both *s* and *r* are distances measured in metres, it follows that the angle θ is simply a ratio. It is a dimensionless quantity. If the object moves twice as far around a circle of twice the radius (Figure 16.3b), its angular displacement θ will be the same.

$$\theta = \frac{\text{length of arc}}{\text{radius}} = \frac{2s}{2r} = \frac{s}{r}$$

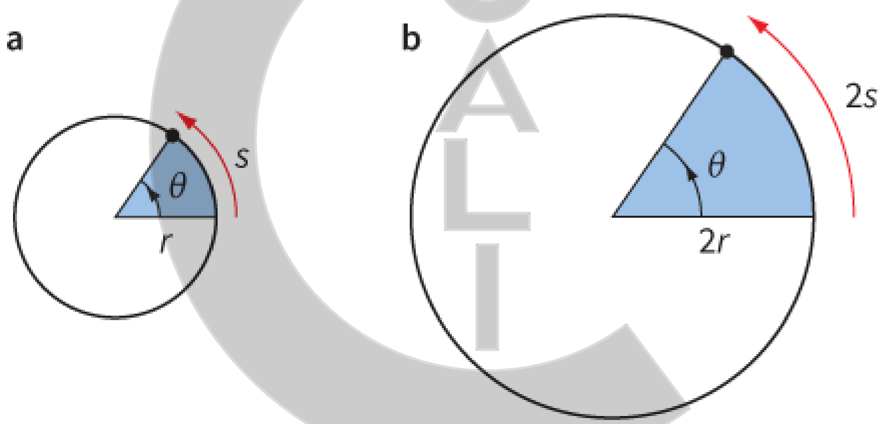


Figure 16.3: The size of an angle depends on the radius and the length of the arc. Doubling both leaves the angle unchanged.

When we define θ in this way, its units are radians rather than degrees. How are radians related to degrees? If an object moves all the way round the circumference of the circle, it moves a distance of $2\pi r$. We can calculate its angular displacement in radians:

$$\theta = \frac{\text{circumference}}{\text{radius}} = \frac{2\pi r}{2r} = 2\pi$$

KEY EQUATION

$$\theta = \frac{\text{arc length}}{\text{radius}}$$

where θ is the angle in radians.

Hence a complete circle contains 2π radians. But we can also say that the object has moved through 360°. Hence:

$$360° = 2\pi \text{ rad}$$

Similarly, we have:

$180° = \pi$ rad $90° = \frac{\pi}{2}$ rad

$45° = \frac{\pi}{4}$ rad and so on

Defining the radian

One **radian** is defined as the angle subtended at the centre of a circle by an arc of length equal to the radius of the circle.

This is illustrated in Figure 16.4.

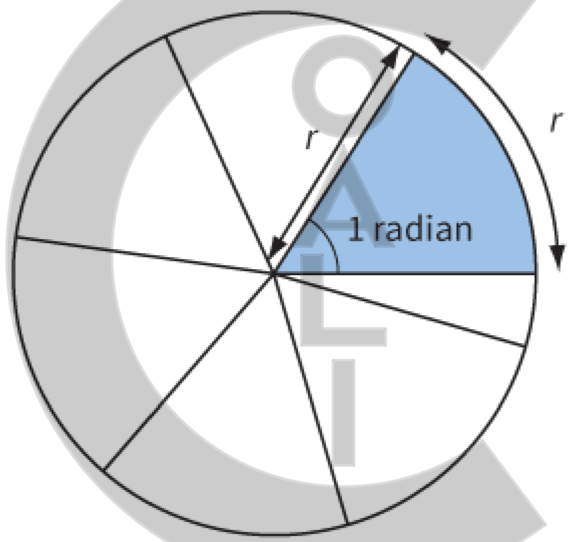


Figure 16.4: The length of the arc is equal to the radius when the angle is 1 radian.

An angle of 360° is equivalent to an angle of 2π radians. We can therefore determine what 1 radian is equivalent to in degrees.

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

If you can remember that there are 2π rad in a full circle, you will be able to convert between radians and degrees:

- to convert from degrees to radians, multiply by $\frac{2\pi}{360°}$ or $\frac{\pi}{180°}$
- to convert from radians to degrees, multiply by $\frac{360°}{2\pi}$ or $\frac{180°}{\pi}$

Now look at Worked example 1.

WORKED EXAMPLE

1 If $\theta = 60°$, what is the value of θ in radians?

The angle θ is 60°. 360° is equivalent to 2π radians. Therefore:

$$\begin{aligned} \theta &= 60° \times \frac{2\pi}{360°} \\ &= \frac{\pi}{3} \\ &= 1.05 \text{ rad} \end{aligned}$$

(Note that it is often useful to express an angle as a multiple of π radians.)

Question

2 **a** Convert the following angles from degrees into radians: 30°, 90°, 105°.

b Convert these angles from radians to degrees: 0.5 rad, 0.75 rad, π rad, $\frac{\pi}{r}$rad

c Express the following angles as multiples of π radians: 30°, 120°, 270°, 720°.

16.3 Steady speed, changing velocity

If we are to use Newton's laws of motion to explain circular motion, we must consider the **velocity** of an object going round in a circle, rather than its **speed**.

There is an important distinction between speed and velocity: **speed** is a scalar quantity that has magnitude only, whereas **velocity** is a vector quantity, with both magnitude and direction.

We need to think about the direction of motion of an orbiting object.

Figure 16.5 shows how we can represent the velocity of an object at various points around its circular path.

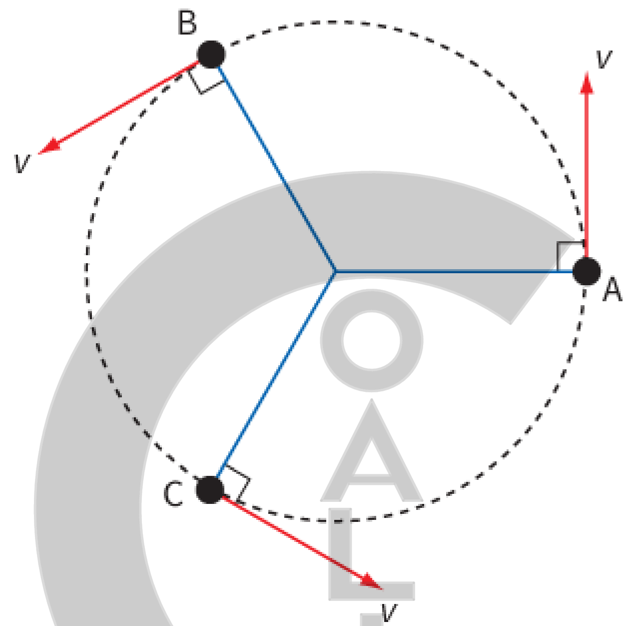


Figure 16.5: The velocity *v* of an object changes direction as it moves along a circular path.

The arrows are straight and show the direction of motion at a particular instant. They are drawn as tangents to the circular path. As the object travels through points A, B, C, etc., its speed remains constant but its direction changes. Since the direction of the velocity *v* is changing, it follows that *v* itself (a vector quantity) is changing as the object moves in a circle.

Questions

3 Explain why all the velocity arrows in Figure 16.5 are drawn the same length.

4 A toy train travels at a steady speed of 0.2 m s^{-1} around a circular track (Figure 16.6). A and B are two points opposite to one another on the track.

- **a** Determine the change in the speed of the train as it travels from A to B.
- **b** Determine the change in the velocity of the train as it travels from A to B.

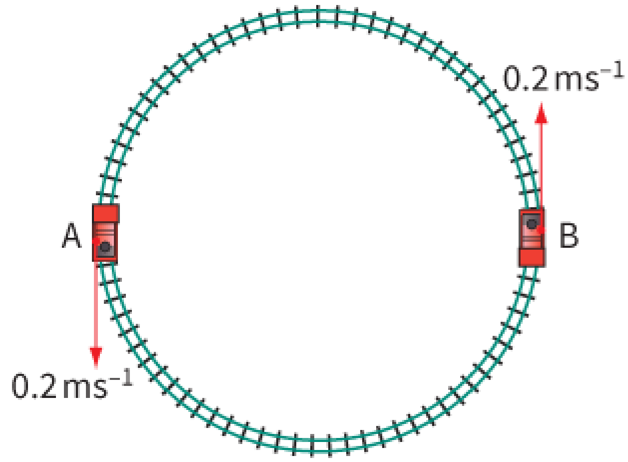


Figure 16.6: A toy train travelling around a circular track.

16.4 Angular speed

As the hands of a clock travel steadily around the clock face, their velocity is constantly changing. The minute hand travels round 360° or 2p radians in 3600 seconds.

Although its velocity is changing, we can say that its **angular speed** is constant, because it moves through the same angle each second:

$$\text{anglular speed} = \frac{\text{angular displacement}}{\text{time taken}}$$

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

where $\Delta\theta$ is the change in angle and Δt is the change in time.

We use the symbol ω (Greek letter omega) for angular velocity, measured in radians per second (rad s^{-1}). For the minute hand of a clock, we have $\omega = \frac{2\pi}{3600} \approx 0.00175 \text{ rad s}^{-1}$.

KEY EQUATION

$$\text{anglular speed} = \frac{\text{angular displacement}}{\text{time taken}}$$

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

A particularly useful example of the equation $\omega = \frac{\Delta\theta}{\Delta t}$ is when a single revolution is considered. The time to make one revolution is referred to as the period (T), the angle through which the object rotates in one revolution is 2π radians. So, substituting in the equation:

$$\omega = \frac{2\pi}{T}$$

KEY EQUATION

$$\omega = \frac{2\pi}{T}$$

Questions

5 Show that the angular speed of the second hand of a clock is about 0.105 rad s^{-1}.

6 In a washing machine, the clothes are held in cylinder called a drum. The drum has holes in it that allow water to enter the drum and also to drain out of the drum.

The drum of a particular washing machine spins at a rate of 1200 rpm (revolutions per minute).

a Determine the number of revolutions per second of the drum.

b Determine the angular speed of the drum.

Relating speed and angular speed

Think again about the second hand of a clock. As the hand goes round, each bit of the hand has the same angular speed. However, different bits of the hand have different velocities. The tip of the hand moves fastest; points closer to the centre of the clock face move more slowly.

This shows that the speed v of an object travelling around a circle depends on two quantities: its angular speed ω and its distance from the centre of the circle r. We can write the relationship as an equation:

$$\text{speed} = \text{angular speed} \times \text{radius}$$
$$v = \omega r$$

KEY EQUATION

$$\text{speed} = \text{angular speed} \times \text{radius}$$
$$v = \omega r$$

Worked example 2 shows how to use this equation.

WORKED EXAMPLE

2 A toy train travels around a circular track of radius 2.5 m in a time of 40 s. What is its speed?

Step 1 Calculate the train's angular speed ω. One circuit of the track is equivalent to 2π radians. The rain travels around in 40 s. Therefore:

$$\omega = \frac{2\pi}{40}$$
$$= 0.157 \text{ rad}^{-1}$$

Step 2 Calculate the train's speed:

$$v = \omega r = 0.157 \times 2.5 = 0.39 \text{m s}^{-1}$$

Hint: *You could have arrived at the same answer by calculating the distance travelled (the circumference of the circle) and dividing by the time taken.*

Questions

7 The angular speed of the second hand of a clock is 0.105 rad s^{-1}. If the length of the hand is 1.8 cm, calculate the speed of the tip of the hand as it moves round.

8 A car travels around a 90° bend in 15 s. The radius of the bend is 50 m.

a Determine the angular speed of the car.

b Determine the speed of the car.

9 A spacecraft orbits the Earth in a circular path of radius 7000 km at a speed of 7800 m s^{-1}. Determine its angular velocity.

16.5 Centripetal forces

When an object's velocity is changing, it has acceleration. In the case of uniform circular motion, the acceleration is rather unusual because, as we have seen, the object's speed does not change but its velocity does. How can an object accelerate and at the same time have a steady speed?

One way to understand this is to think about what Newton's laws of motion can tell us about this situation. **Newton's first law** states that an object remains at rest or in a state of uniform velocity (at constant speed in a straight line) unless it is acted on by an external force. In the case of an object moving at steady speed in a circle, we have a body whose velocity is not constant; therefore, there must be a resultant (unbalanced) force acting on it.

Now we can think about different situations where objects are going round in a circle and try to find the force that is acting on them.

- Consider a rubber bung on the end of a string. Imagine whirling it in a horizontal circle above your head (Figure 16.7). To make it go round in a circle, you have to pull on the string. The pull of the string on the bung is the unbalanced force, which is constantly acting to change the bung's velocity as it orbits your head. If you let go of the string, suddenly there is no tension in the string and the bung will fly off at a tangent to the circle.
- Similarly, as the Earth orbits the Sun, it has a constantly changing velocity. Newton's first law suggests that there must be an unbalanced force acting on it. That force is the gravitational pull of the Sun. If the force disappeared, the Earth would travel off in a straight line.

In both of these cases, you should be able to see why the direction of the force is as shown in Figure 16.8. The force on the object is directed towards the centre of the circle. We describe each of these resultant forces as a **centripetal force** – that is, directed towards the centre.

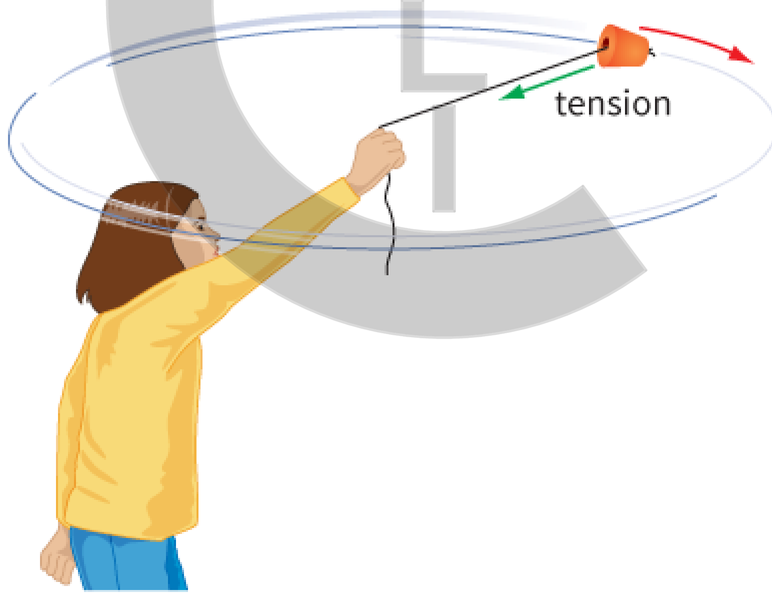


Figure 16.7: Whirling a rubber bung.

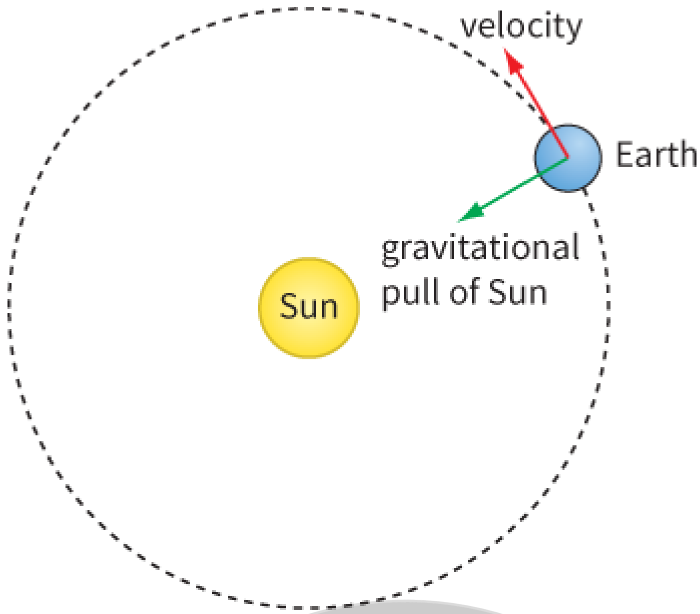


Figure 16.8: The gravitational pull of the Sun provides the centripetal force that keeps the Earth in its orbit.

It is important to note that the word centripetal is an adjective. We use it to describe a force that is making something travel along a circular path. It does not tell us what causes this force, which might be gravitational, electrostatic, magnetic, frictional or whatever.

Questions

10 In each of the following cases, state what provides the resultant force causing centripetal acceleration:

- **a** the Moon orbiting the Earth
- **b** a car going round a bend on a flat, rough road
- **c** the weight on the end of a swinging pendulum.

11 A car is travelling along a flat road in winter. The car approaches a patch of ice on a bend. Explain why the car cannot go around the perfectly smooth, icy bend. Suggest what might happen if the driver tries turning the steering wheel when the car is on the ice.

Vector diagrams

Figure 16.9a shows an object travelling along a circular path, at two positions in its orbit. It reaches position B a short time after A. How has its velocity changed between these two positions?

The change in the velocity of the object can be determined using a vector triangle. The vector triangle in Figure 16.9b shows the difference between the final velocity v_B and initial velocity v_A. The change in the velocity of the object between the points B and A is shown by the smaller arrow labelled Δv. Note that the change in the velocity of the object is (more or less):

- at right angles to the velocity at A
- directed towards the centre of the circle.

The object is accelerating because its velocity changes. Since acceleration is the rate of change of velocity, it follows that the acceleration of the object must be in the same direction as the change in the velocity – towards the centre of the circle. This is not surprising because, according to $F = ma$, the acceleration a of the object is in the same direction as the centripetal force F:

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

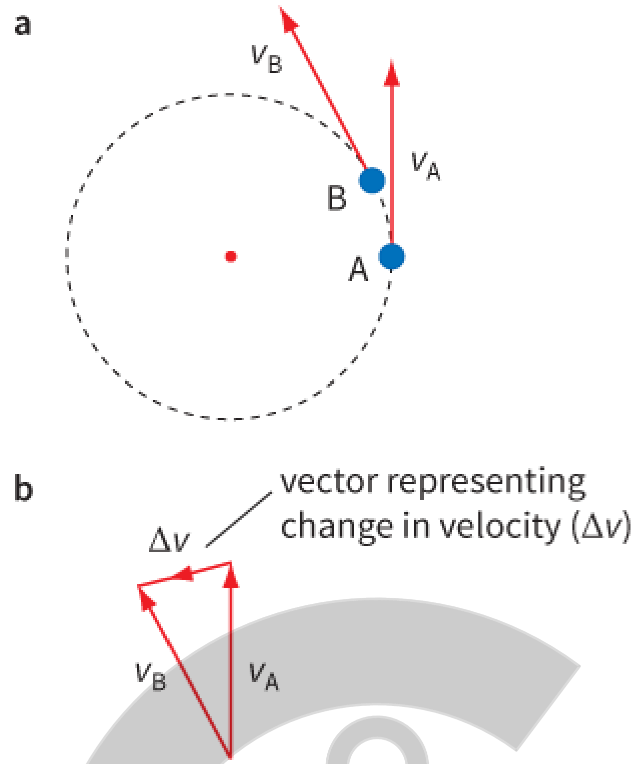


Figure 16.9: Changes in the velocity vector.

Acceleration at steady speed

Now that we know that the centripetal force *F* and acceleration are always at right angles to the object's velocity, we can explain why its speed remains constant. If the force is to make the object change its speed, it must have a component in the direction of the object's velocity; it must provide a push in the direction in which the object is already travelling. However, here we have a force at 90° to the velocity, so it has no component in the required direction. (Its component in the direction of the velocity is $F \cos 90° = 0$.) It acts to pull the object around the circle, without ever making it speed up or slow down.

You can also use the idea of work done to show that the speed of the object moving in a circle remains the same. The work done by a force is equal to the product of the force and the distance moved by the object in the direction of the force. The distance moved by the object in the direction of the centripetal force is zero; hence the work done is zero. If no work is done on the object, its kinetic energy must remain the same and hence its speed is unchanged.

Question

12 An object follows a circular path at a steady speed. Describe how each of the following quantities changes as it follows this path: speed, velocity, kinetic energy, momentum, centripetal force, centripetal acceleration. (Refer to both magnitude and direction, as appropriate.)

Understanding circular motion

Isaac Newton devised an ingenious thought experiment that allows us to think about how an object can remain in a circular orbit around the Earth. Consider a large cannon on some high point on the Earth's surface, capable of firing objects horizontally. Figure 16.10 shows what will happen if we fire them at different speeds.

If the object is fired too slowly, gravity will pull it down towards the ground and it will land at some distance from the cannon. A faster initial speed results in the object landing further from the cannon.

Now, if we try a bit faster than this, the object will travel all the way round the Earth. We have to get just the right speed to do this. As the object is pulled down towards the Earth, the curved surface of the Earth falls away beneath it. The object follows a circular path, constantly falling under gravity but never getting any closer to the surface.

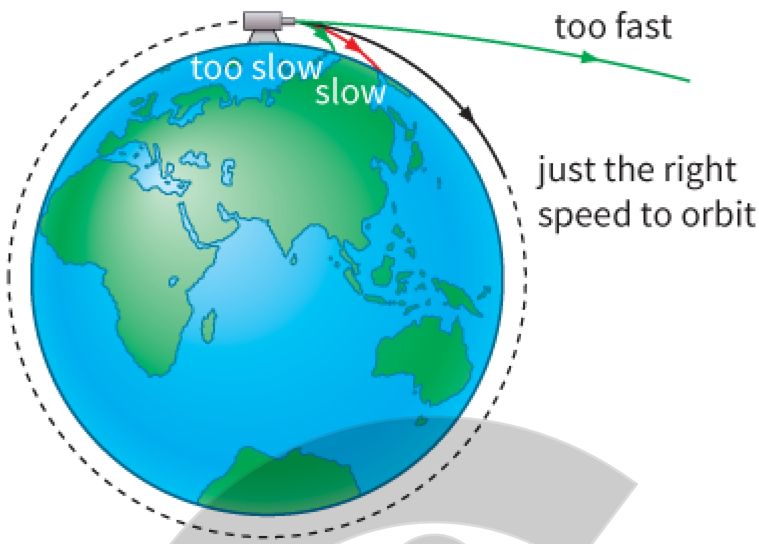


Figure 16.10: Newton's 'thought experiment'.

If the object is fired too fast, it travels off into space, and fails to get into a circular orbit. So we can see that there is just one correct speed to achieve a circular orbit under gravity. (Note that we have ignored the effects of air resistance in this discussion.)

16.6 Calculating acceleration and force

If we spin a bung around in a circle (Figure 16.7), we get a feeling for the factors that determine the centripetal force F required to keep it in its circular orbit. The greater the mass m of the bung and the greater its speed v, the greater is the force F that is required. However, if the radius r of the circle is increased, F is smaller.

Now we will deduce an expression for the **centripetal acceleration** of an object moving around a circle with a constant speed.

Figure 16.11 shows a particle moving round a circle. In time Δt it moves through an angle $\Delta\theta$ from A to B. Its speed remains constant but its velocity changes by Δv, as shown in the vector diagram. Since the narrow angle in this triangle is also $\Delta\theta$, we can say that:

$$\Delta\theta = \frac{\Delta v}{v}$$

Dividing both sides of this equation by Δt and rearranging gives:

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

The quantity on the left is $\frac{\Delta v}{\Delta t} = a$ the particle's acceleration.

The quantity on the right is $\frac{\Delta\theta}{\Delta t} = \omega$ the angular velocity.

Substituting for these gives:

$$a = v\omega$$

Using $v = \omega r$ we can eliminate ω from this equation:

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

where a is the centripetal acceleration, v is the speed and r is the radius of the circle.

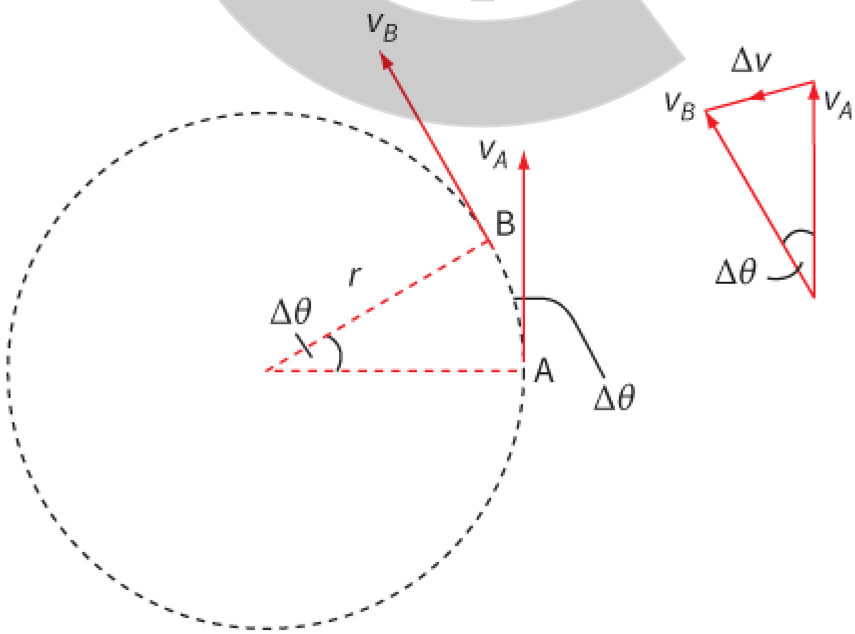


Figure 16.11: Deducing an expression for centripetal acceleration.

Question

13 Show that an alternative equation for the centripetal acceleration is $a = \omega^2 r$.

KEY EQUATION

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

Newton's second law of motion

Now that we have an equation for centripetal acceleration, we can use **Newton's second law** of motion to deduce an equation for centripetal force. If we write this law as $F = ma$, we find:

$$\text{centripetal force } F = \frac{mv^2}{r} = mr\omega^2$$

Remembering that an object accelerates in the direction of the resultant force on it, it follows that both F and a are in the same direction, towards the centre of the circle.

Questions

14 Calculate how long it would take a ball to orbit the Earth once, just above the surface, at a speed of 7920 m s^{-1}. (The radius of the Earth is 6400 km.)

15 A stone of mass 0.20 kg is whirled round on the end of a string in a vertical circle of radius 30 cm. The string will break when the tension in it exceeds 8.0 N. Calculate the maximum speed at which the stone can be whirled without the string breaking.

Figure 16.12: The view from the International Space Station, orbiting the Earth over Australia.

16 The International Space Station (Figure 16.12) has a mass of 350 tonnes, and orbits the Earth at an average height of 340 km where the gravitational acceleration is 8.8 m s^{-2}. The radius of the Earth is 6400 km. Calculate:

a the centripetal force on the space station

b the speed at which it orbits

c the time taken for each orbit

d the number of times it orbits the Earth each day.

17 An toy truck of mass 0.40 kg travels round a horizontal circular track of radius 0.50 m. It makes three complete revolutions every 10 seconds.

Calculate:

a its speed

b its centripetal acceleration

c the centripetal force.

18 Mars orbits the Sun once every 687 days at a distance of 2.3×10^{11} m. The mass of Mars is 6.4×10^{23} kg. Calculate:

a the average speed in metres per second

b its centripetal acceleration

c the gravitational force exerted on Mars by the Sun.

Calculating orbital speed

We can use the force equation to calculate the speed that an object must have to orbit the Earth under gravity, as in Newton's thought experiment. The necessary centripetal force $\frac{mv^2}{r}$ is provided by the Earth's gravitational pull mg.

Hence:

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

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

where $g = 9.81$ m s^{-2} is the acceleration of free fall close to the Earth's surface. The radius of its orbit is equal to the Earth's radius, approximately 6400 km. Hence, we have:

$$9.81 = \frac{v^2}{(6.4 \times 10^6)}$$

$$v^2 = 9.81 \times (6.4 \times 10^6)$$

$$v = \sqrt{9.81(6.4 \times 10^6)}$$

$$v \approx 7.92 \times 10^3 \text{ ms}^{-1}$$

Thus, if you were to throw or hit a ball horizontally at almost 8 km s^{-1}, it would go into orbit around the Earth.

16.7 The origins of centripetal forces

It is useful to look at one or two situations where the physical origin of the centripetal force may not be immediately obvious. In each case, you will notice that the forces acting on the moving object are not balanced – there is a resultant force. An object moving along a circular path is not in equilibrium and the resultant force acting on it is the centripetal force.

1 A car cornering on a level road (Figure 16.13). Here, the road provides two forces. The force *N* is the normal contact force that balances the weight *mg* of the car–the car has no acceleration in the vertical direction.

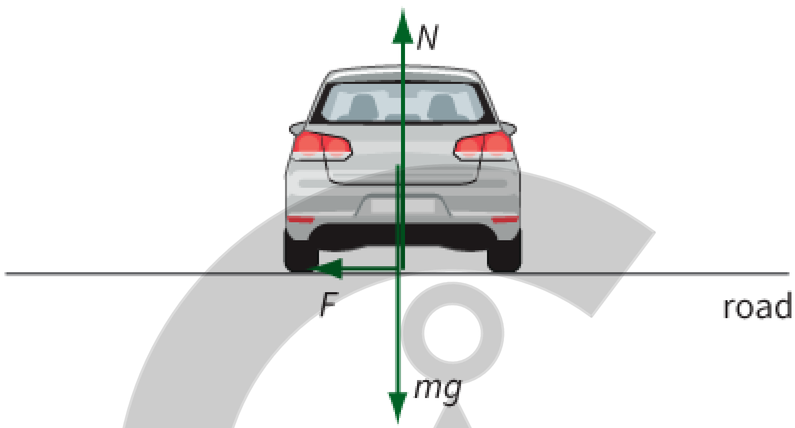


Figure 16.13: This car is moving away from us and turning to the left. Friction provides the centripetal force. *N* and *F* are the total normal contact and friction forces (respectively) provided by the contact of all four tyres with the road.

The second force is the force of friction *F* between the tyres and the road surface. This is the unbalanced, centripetal force. If the road or tyres do not provide enough friction, the car will not go round the bend along the desired path. The friction between the tyres and the road provides the centripetal force necessary for the car's circular motion.

2 A car cornering on a banked road (Figure 16.14a). Here, the normal contact force *N* has a horizontal component that can provide the centripetal force. The vertical component of *N* balances the car's weight. Therefore:

$$\text{vertically} \quad N\cos\theta = mg$$

$$\text{horizontally} \quad N\sin\theta = \frac{mv^2}{r}$$

where *r* is the radius of the circular corner and *v* is the car's speed.

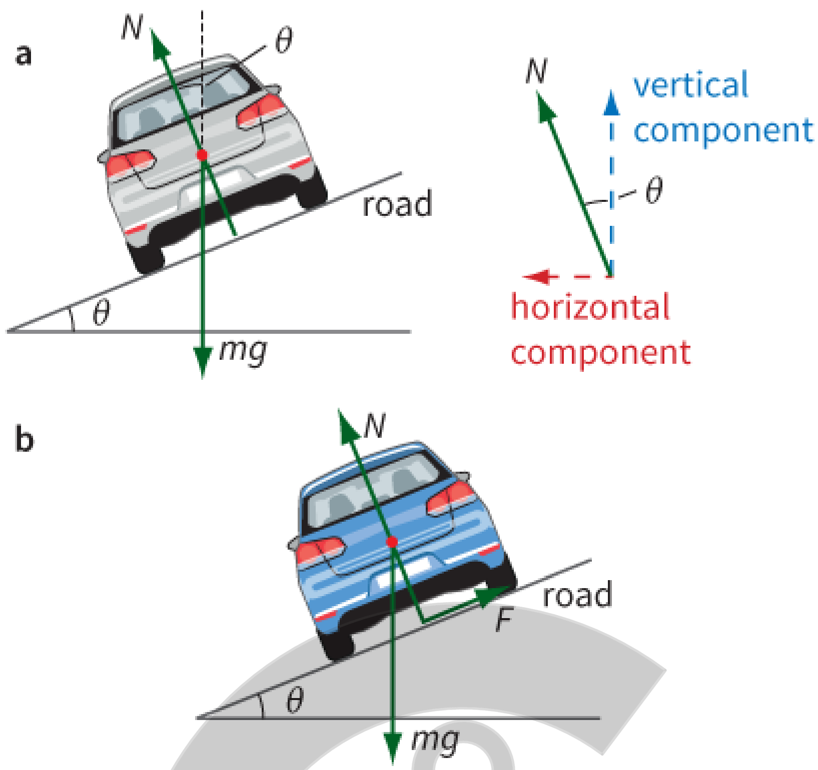


Figure 16.14: **a** On a banked road, the horizontal component of the normal contact force from the road can provide the centripetal force needed for cornering. **b** For a slow car, friction acts up the slope to stop it from sliding down.

If a car travels around the bend too slowly, it will tend to slide down the slope and friction will act up the slope to keep it on course (Figure 16.14b). If it travels too fast, it will tend to slide up the slope. If friction is insufficient, it will move up the slope and come off the road.

3 An aircraft banking (Figure 16.15a). To change direction, the pilot tips the aircraft's wings. The vertical component of the lift force *L* on the wings balances the weight. The horizontal component of *L* provides the centripetal force.

4 A stone being whirled in a horizontal circle on the end of a string – this arrangement is known as a conical pendulum (Figure 16.15b). The vertical component of the tension *T* is equal to the weight of the stone. The horizontal component of the tension provides the centripetal force for the circular motion.

5 At the fairground (Figure 16.15c). As the cylinder spins, the floor drops away. Friction balances your weight. The normal contact force of the wall provides the centripetal force. You feel as though you are being pushed back against the wall; what you are feeling is the push of the wall on your back.

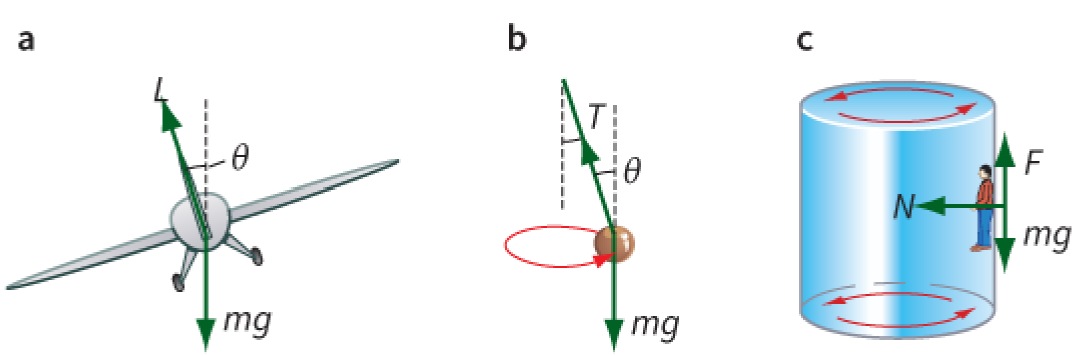


Figure 16.15: Three more ways of providing a centripetal force.

Note that the three situations shown in Figures 16.14a, 16.15a and 16.15b are equivalent. The moving object's weight acts downwards. The second force has a vertical component, which balances the weight, and a horizontal component, which provides the centripetal force.

Questions

19 Explain why it is impossible to whirl a bung around on the end of a string in such a way that the string remains perfectly horizontal.

20 Explain why an aircraft will tend to lose height when banking, unless the pilot increases its speed to provide more lift.

21 If you have ever been down a water-slide (a flume) (Figure 16.16) you will know that you tend to slide up the side as you go around a bend. Explain how this provides the centripetal force needed to push you around the bend. Explain why you slide higher if you are going faster.

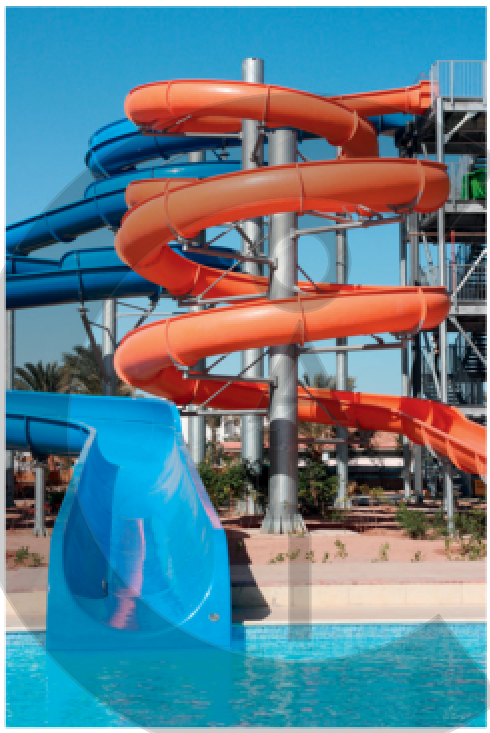


Figure 16.16: A water-slide is a good place to experience centripetal forces.

REFLECTION

In order to increase the proportion of heavy water (deuterium oxide) in a sample of water, the scientists in the 'Manhattan Project' used a centrifuge. Prepare a short talk on the aim of the Manhattan Project and to explain how a centrifuge works.

Think about an astronaut far away from any planets. Discuss with a partner whether the string on a conical pendulum could rotate in a horizontal plane, and what speed it would start to rotate in this manner. What conclusions did you each come to? Did discussing this question with a partner help you to understand the concepts?

SUMMARY

Angles can be measured in radians. An angle of 2π rad is equal to 360°.

An object moving at a steady speed along a circular path has uniform circular motion.

The angular displacement θ is a measure of the angle through which an object moves in a circle.

The angular velocity ω is the rate at which the angular displacement changes:

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

For an object moving with uniform circular motion, speed and angular velocity are related by $v = \omega r$.

An object moving in a circle is not in equilibrium; it has a resultant force acting on it.

The resultant force acting on an object moving in a circle is directed towards the centre of the circle and is at right angles to the velocity of the object.

An object moving in a circle has a centripetal acceleration a given by:

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

The magnitude of the force F acting on an object of mass m moving at a speed v in a circle of radius r is given by:

$$F = \frac{mv^2}{r} = mr\omega^2$$

EXAM-STYLE QUESTIONS

1 Which statement is correct? **[1]**

A There is a resultant force on an object moving along a circular path at constant speed away from the centre of the circle causing it to be thrown outwards.

B There is a resultant force on an object moving along a circular path at constant speed towards the centre of the circle causing it to be thrown outwards.

C There is a resultant force on an object moving along a circular path at constant speed towards the centre of the circle causing it to move in the circle.

D There is zero resultant force on an object moving along a circular path at constant speed because it is in equilibrium.

2 When ice-dancers spin, as shown in the diagram, the first dancer's hand applies a centripetal force to the second dancer's hand.

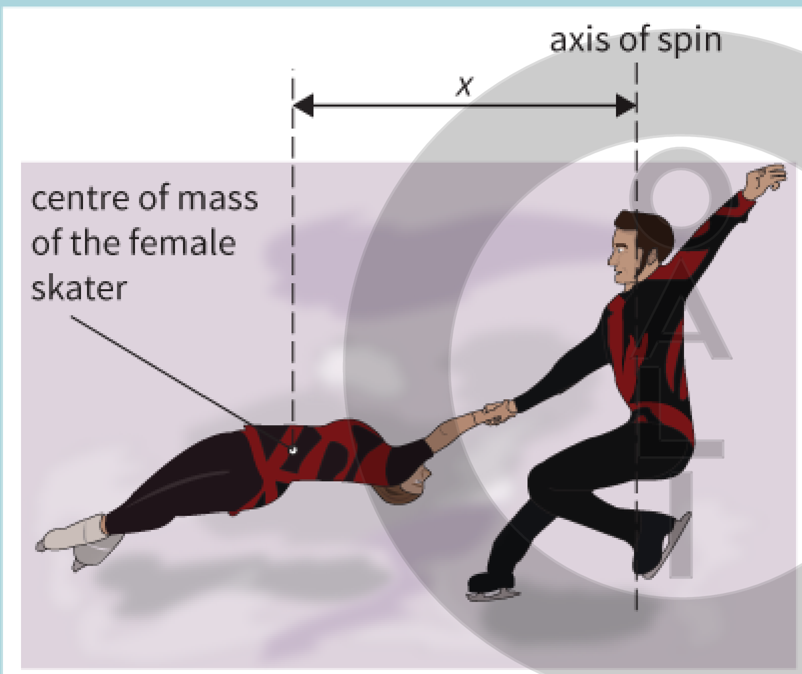


Figure 16.17

In which case is the centripetal force the greatest? **[1]**

	x / m	**Speed of the female skater's centre of mass / m s^{-1}**
A	0.45	9.0
B	0.45	10.0
C	0.50	9.0
D	0.50	10.0

Table 16.1

3 **a** Explain what is meant by a **radian**. **[1]**

b A body moves round a circle at a constant speed and completes one revolution in 15 s. Calculate the angular speed of the body. **[2]**

[Total: 3]

4 This diagram shows part of the track of a roller-coaster ride in which a truck loops the loop. When the truck is at the position shown, there is no reaction force between the wheels of the truck and the track. The diameter of the loop in the track is 8.0 m.

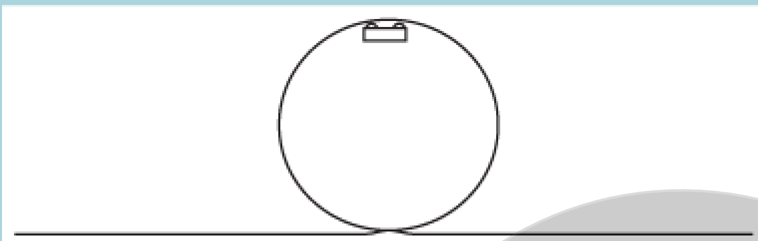


Figure 16.18

a Explain what provides the centripetal force to keep the truck moving in a circle. **[1]**

b Given that the acceleration due to gravity g is 9.8 m s^{-2}, calculate the speed of the truck. **[3]**

[Total: 4]

5 This diagram shows a toy of mass 60 g placed on the edge of a rotating turntable. **[1]**

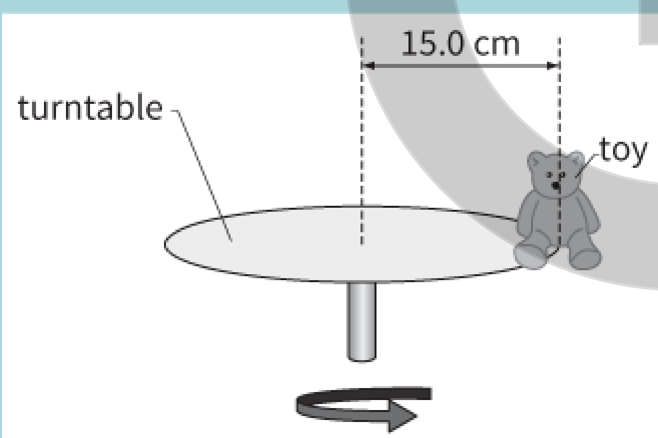


Figure 16.19

a The radius of the turntable is 15.0 cm. The turntable rotates, making 20 revolutions every minute. Calculate the resultant force acting on the toy. **[3]**

b Explain why the toy falls off when the speed of the turntable is increased. **[2]**

[Total: 6]

6 One end of a string is secured to the ceiling and a metal ball of mass 50 g is tied to its other end. The ball is initially at rest in the vertical position. The ball is raised through a vertical height of 70 cm, as shown. The ball is then released. It describes a circular arc as it passes through the vertical position.

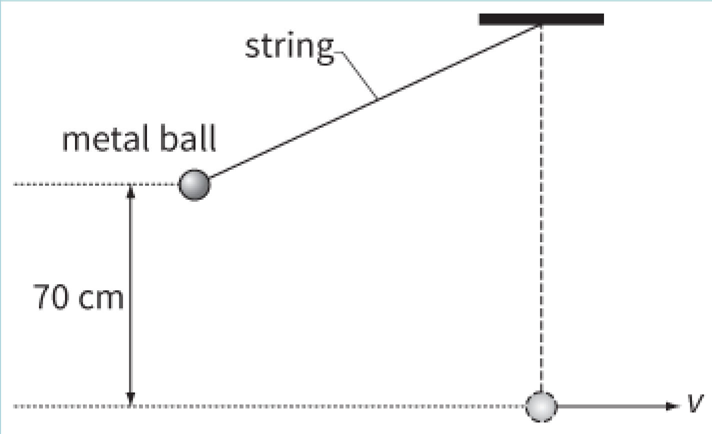


Figure 16.20

The length of the string is 1.50 m.

a Ignoring the effects of air resistance, determine the speed v of the ball as it passes through the vertical position. **[2]**

b Calculate the tension T in the string when the string is vertical. **[3]**

c Explain why your answer to part b is not equal to the weight of the ball. **[2]**

[Total: 7]

7 A car is travelling round a bend when it hits a patch of oil. The car slides off the road onto the grass verge. Explain, using your understanding of circular motion, why the car came off the road. **[2]**

8 This diagram shows an aeroplane banking to make a horizontal turn. The aeroplane is travelling at a speed of 75 m s^{-1} and the radius of the turning circle is 800 m.

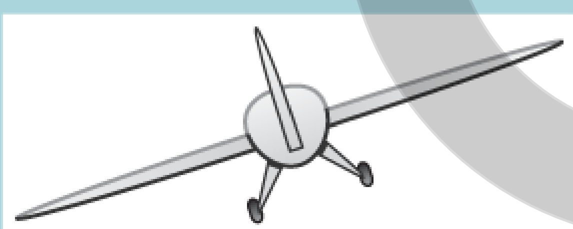


Figure 16.21

a Copy the diagram. On your copy, draw and label the forces acting on the aeroplane. **[2]**

b Calculate the angle that the aeroplane makes with the horizontal. **[4]**

[Total: 6]

9 **a** Explain what is meant by the term **angular speed**. **[2]**

b This diagram shows a rubber bung, of mass 200 g, on the end of a length of string being swung in a horizontal circle of radius 40 cm. The string makes an angle of 56° with the vertical.

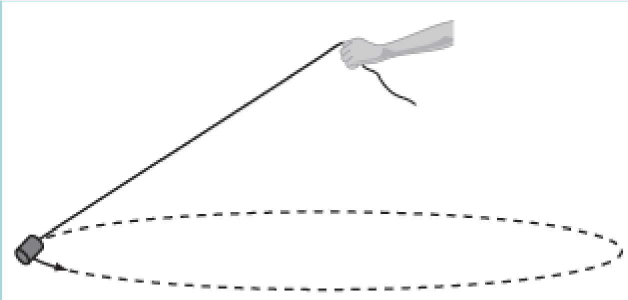


Figure 16.22

Calculate:

i the tension in the string **[2]**

ii the angular speed of the bung **[3]**

iii the time it takes to make one complete revolution. **[1]**

[Total: 8]

10 a Explain what is meant by a **centripetal acceleration**. **[2]**

b A teacher swings a bucket of water, of total mass 5.4 kg, round in a vertical circle of diameter 1.8 m.

i Calculate the minimum speed that the bucket must be swung at so that the water remains in the bucket at the top of the circle. **[3]**

ii Assuming that the speed remains constant, what will be the force on the teacher's hand when the bucket is at the bottom of the circle? **[2]**

[Total: 7]

11 In training, military pilots are given various tests. One test puts them in a seat on the end of a large arm that is then spun round at a high speed, as shown.

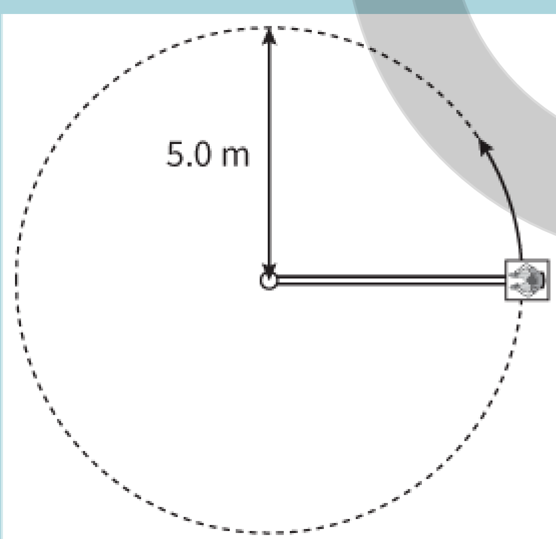


Figure 16.23

a Describe what the pilot will feel and relate this to the centripetal force. **[3]**

b At top speed the pilot will experience a centripetal force equivalent to six times his own weight (6 *mg*).

i Calculate the speed of the pilot in this test. **[3]**

ii Calculate the number of revolutions of the pilot per minute. **[2]**

c Suggest why it is necessary for pilots to be able to be able to withstand forces of this type. **[2]**

[Total: 10]

12 a Show that in one revolution there are 2π radians. **[2]**

b This diagram shows a centrifuge used to separate solid particles suspended in a liquid of lower density. The container is spun at a rate of 540 revolutions per minute.

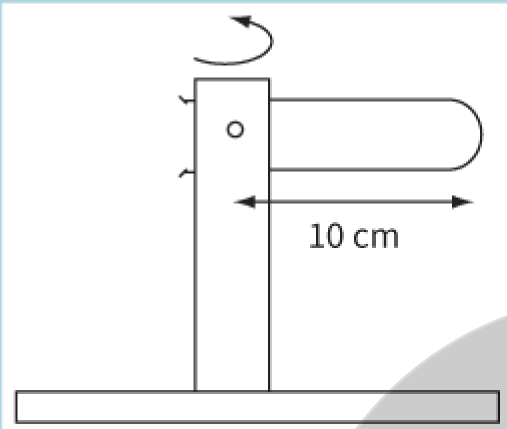


Figure 16.24

i Calculate the angular velocity of the container. **[2]**

ii Calculate the centripetal force on a particle of mass 20 mg at the end of the test tube. **[2]**

c An alternative method of separating the particles from the liquid is to allow them to settle to the bottom of a stationary container under gravity.

By comparing the forces involved, explain why the centrifuge is a more effective method of separating the mixture. **[2]**

[Total: 8]

SELF-EVALUATION CHECKLIST

After studying this chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define the radian and use it as the unit of angular displacement	16.2			
understand the concept of angular speed	16.4			
recall and use the relationship angular speed $\omega = \frac{2\pi}{T}$ where T is the time for one complete revolution	16.4			
recall and use the relationship angular speed $v = \omega r$	16.4			
understand that the force on an object rotating round a circle is towards the centre of the circle and is called a centripetal force	16.5			
recognise that the centripetal force is at right angles to the velocity of the object	16.5			
recognise that the centripetal force causes centripetal acceleration	16.5			
recognise that a constant centripetal force causes circular motion with constant angular speed	16.5			
recall and use the formula: $a = \frac{v^2}{r} = r\omega^2$	16.6			
recall and use the formula: $F = \frac{mv^2}{r} = mr\omega^2$	16.6			

〉 Chapter 17

Gravitational fields

LEARNING INTENTIONS

In this chapter you will learn how to:

- describe a gravitational field as a field of force and define gravitational field strength *g*
- represent a gravitational field using field lines
- understand the meaning of centre of mass and use the concept in problems involving uniform spheres
- recall and use Newton's law of gravitation
- solve problems involving the gravitational field strength of a uniform field and the field of a point mass
- understand how the gravitational potential energy, $E = -\frac{Gm_1m_2}{r}$, of two point masses is a consequence of gravitational potential
- define and solve problems involving gravitational potential
- analyse circular orbits in an inverse square law field, including geostationary orbits.

BEFORE YOU START

We have all experienced gravity and the effects it has on us in everyday life.

Imagine what would happen if gravity was 'switched off'. Discuss the effects that it would have, not only on everyday life, but on a much bigger scale. For example, what would happen to the objects on your desk? Can you predict what would happen to the orbit of the Earth?

GRAVITATIONAL FORCES AND FIELDS

Gravity is amazing! It is the first interaction that we experience, and it is an interaction that we take for granted. But without it, there would be no Earth, no Sun, no galaxies, no us!

You have probably seen pictures like Figure 17.1 before – astronauts can float in mid-air in the International Space Station (ISS). This is not because there is no gravity at this altitude, but because the ISS is effectively in freefall as it orbits around the Earth. The only astronauts who have experienced zero gravity are those who experienced it on their way to and from the Moon. Even then, there would be a small gravitational effect due to the pull of the Earth, the Sun and the Moon itself. This is sometimes referred to as microgravity.

All life on Earth evolved in the Earth's gravitational field. The bodies of all animals evolved so that they could cope with the strains and stresses of the forces inherent in living in a gravitational field. Astronauts spending long periods of time in microgravity, for example in the proposed trips to Mars, would find their bodies losing calcium and their muscles, which are used to supporting their weight, getting weaker.

One of the most amazing things about gravity is its role in the development of the Universe. Nebulae are great clouds of dust and (mostly) hydrogen gas in outer space. Irregularities in the density of the cloud lead to more material being attracted to the higher density areas, so this area becomes more and more concentrated as more material is attracted. The process gathers pace. The vast quantity of gravitational potential energy of the spread-out hydrogen becomes kinetic energy of hydrogen atoms that, once great enough, allows nuclear fusion to occur … and a star is born.

Why is this so amazing? Gravity is a very weak interaction; the electromagnetic interaction is far, far stronger. The electric repulsion between two protons in the nucleus of an atom is about 10^{36} times bigger (that is, 10 followed by 35 zeroes!) than the gravitational attraction.

Figure 17.1: An astronaut on board a space shuttle in Earth-orbit.

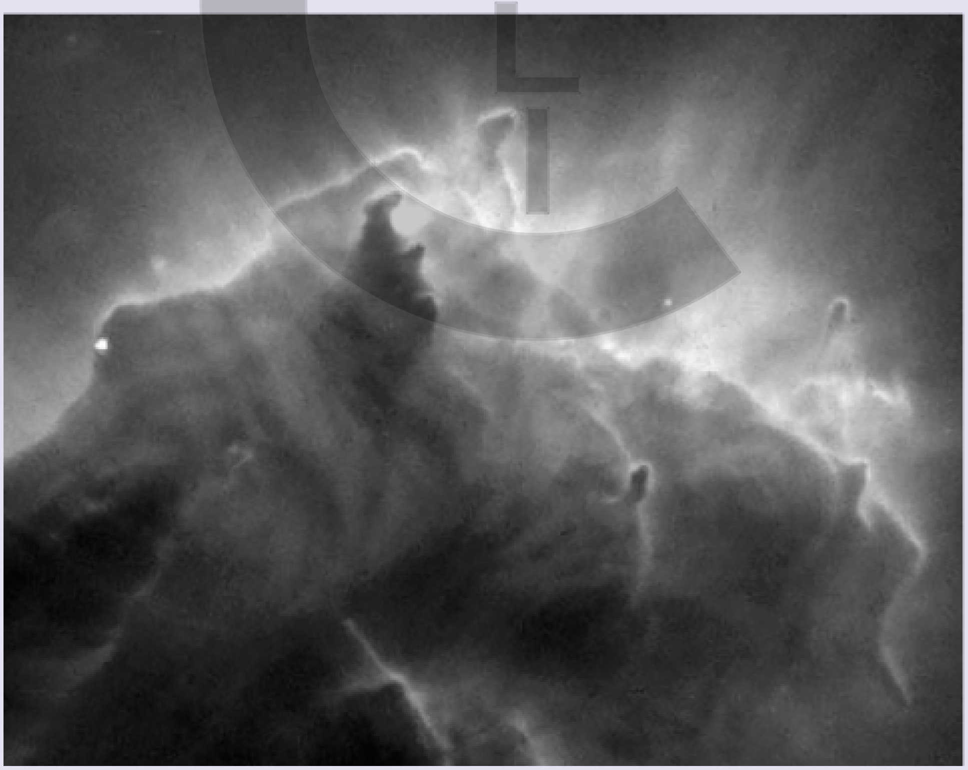


Figure 17.2: Gaseous star nurseries.

Why does the gravitational interaction rule? The gravitational force is always attractive, whereas there are two types of charge (positive and negative) so the electromagnetic interaction can either be attractive or repulsive. In general, the negative and positive charges tend to cancel out, making any large scale object nearly electrically neutral.

17.1 Representing a gravitational field

We can represent the Earth's **gravitational field** by drawing field lines, as shown in Figure 17.3.

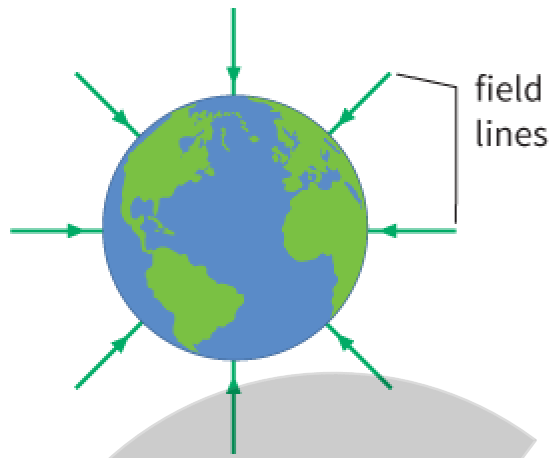


Figure 17.3: The Earth's gravitational field is represented by field lines.

The field lines show two things:

- The arrows on the field lines show us the direction of the gravitational force on a mass placed in the field.
- The spacing of the field lines indicates the strength of the gravitational field–the further apart they are, the weaker the field.

The drawing of the Earth's gravitational field shows that all objects are attracted towards the centre of the Earth. This is true even if they are below the surface of the Earth. The gravitational force gets weaker as you get further away from the Earth's surface – this is shown by the greater separation between the field lines. The Earth is almost a uniform spherical mass, although it does bulge a bit at the equator. The gravitational field of the Earth is as if its entire mass was concentrated at its centre; this is known as its **centre of mass**. As far as any object beyond the Earth's surface is concerned, the Earth behaves as a point mass.

Figure 17.4 shows the Earth's gravitational field closer to its surface. The gravitational field in and around a building on the Earth's surface shows that the gravitational force is directed downwards everywhere and (because the field lines are very nearly parallel and evenly spaced) the strength of the gravitational field is virtually the same at all points in and around the building. This means that your weight is virtually the same everywhere in this gravitational field. Your weight does not become much less when you go upstairs.

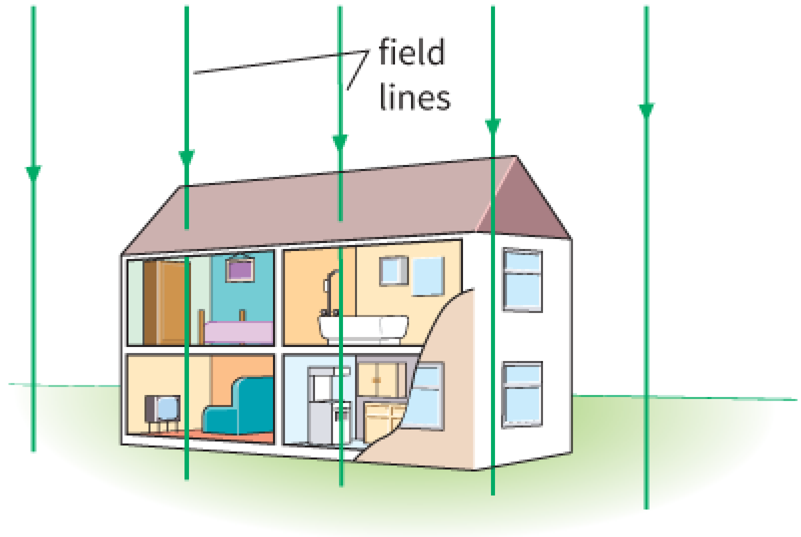


Figure 17.4: The Earth's gravitational field is uniform on the scale of a building.

We describe the Earth's gravitational field as **radial**, since the field lines diverge (spread out) radially from the centre of the Earth. However, on the scale of a building, the gravitational field is **uniform**, since the field lines are equally spaced. Jupiter is a more massive planet than the Earth and so we would represent its gravitational field by showing more closely spaced field lines.

Newton's law of gravitation

Newton used his ideas about mass and gravity to suggest a law of gravitation for two point masses (Figure 17.5).

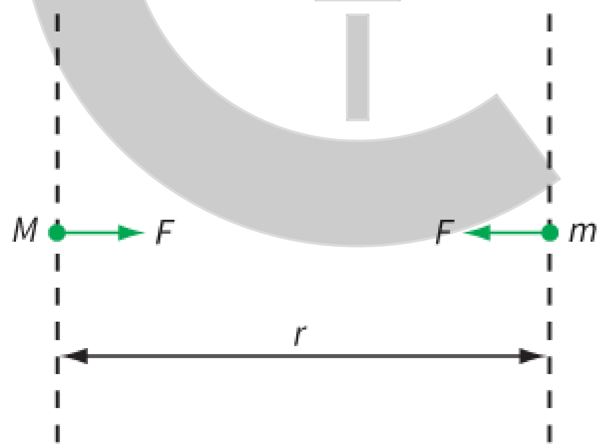


Figure 18.5: Two point masses separated by distance *r*.

Newton considered two point masses *M* and *m* separated by a distance *r*. Each point mass attracts the other with a force *F*. (According to Newton's third law of motion, the point masses interact with each other and therefore exert equal but opposite forces on each other.)

Newton's law of gravitation states that any two point masses attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of their separation.

Note that the law refers to 'point masses' – you can alternatively use the term 'particles'. Things are more complicated if we think about solid bodies that occupy a volume of space. Each particle of one body attracts every particle of the other body and we would have to add all these forces together to work out the force each body has on the other. Newton was able to show that two uniform spheres attract one another with a force that is the same as if their masses were concentrated at their centres (provided their centre-to-centre distance is greater than the sum of their radii).

According to Newton's law of gravitation, we have:

$$\text{force} \propto \text{product of the masses, or } F \propto Mm$$

$$\text{force} \propto \frac{1}{\text{distance}^2} \text{ or } F \propto \frac{1}{r^2}$$

Therefore:

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

To make this into an equation, we introduce the gravitational constant G:

$$F = \frac{GMm}{r^2}$$

(The force is attractive, so F is in the opposite direction to r.)

The gravitational constant G is sometimes referred to as the universal gravitational constant because it is believed to have the same value, 6.67×10^{-11} N m^2 kg^{-2}, throughout the Universe. This is important for our understanding of the history and likely long-term future of the Universe.

The equation can also be applied to spherical objects (such as the Earth and the Moon) provided we remember to measure the separation r between the centres of the objects. You may also come across the equation in the form:

$$F = \frac{Gm_1m_2}{r^2}$$

where m_1 and m_2 are the masses of the two bodies.

KEY EQUATION

Newton's law of gravitation:

$$F = \frac{Gm_1m_2}{r^2}$$

Let us examine this equation to see why it seems reasonable. First, each of the two masses is important. Your weight (the gravitational force on you) depends on your mass and on the mass of the planet you happen to be standing on.

Second, the further away you are from the planet, the weaker its pull. Twice as far away gives one-quarter of the force. This can be seen from the diagram of the field lines in Figure 17.6. If the distance is doubled, the lines are spread out over four times the surface area, so their concentration is reduced to one-quarter. This is called an inverse square law. Inverse square laws are common in physics, light or γ-rays spreading out uniformly from a point source also follow an inverse square law.

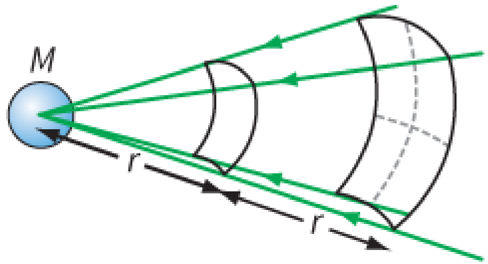


Figure 17.6: Field lines are spread out over a greater surface area at greater distances, so the strength of the field is weaker.

We measure distances from the centre of mass of one body to the centre of mass of the other (Figure 17.7). We treat each body as if its mass were concentrated at one point. The two bodies attract each other with equal and opposite forces, as required by Newton's third law of motion. The Earth pulls on you with a force (your weight) directed towards the centre of the Earth; you attract the Earth with an equal force, directed away from its centre and towards you. Your pull on an object as massive as the Earth has little effect on it. The Sun's pull on the Earth, however, has a very significant effect.

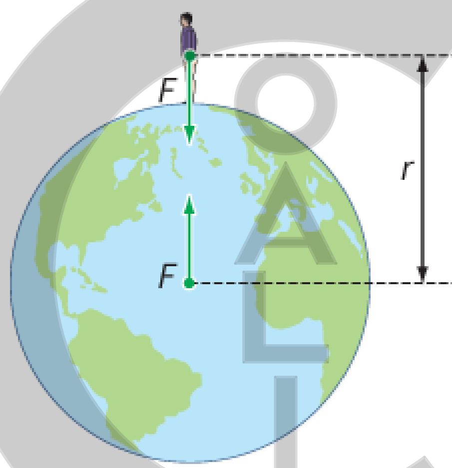


Figure 17.7: A person and the Earth exert equal and opposite attractive forces on each other.

Questions

1 Calculate the gravitational force of attraction between:

- **a** two objects separated by a distance of 1.0 cm and each having a mass of 100 g
- **b** two asteroids separated by a distance of 4.0 10^9 m and each having a mass of 5.0×10^{10} kg
- **c** a satellite of mass 1.4×10^4 kg orbiting the Earth at a distance of 6800 km from the Earth's centre. (The mass of the Earth is 6.0×10^{24} kg.)

2 Estimate the gravitational force of attraction between two people sitting side by side on a park bench. How does this force compare with the gravitational force exerted on each of them by the Earth (in other words, their weight)?

17.2 Gravitational field strength *g*

We can describe how strong or weak a gravitational field is by stating its **gravitational field strength**. We are used to this idea for objects on or near the Earth's surface. The gravitational field strength is the familiar quantity g. Its value is approximately 9.8 m s^{-2}. The weight of a body of mass m is mg.

To make the meaning of g clearer, we should write it as 9.8 N kg^{-1}. That is, each 1 kg of mass experiences a gravitational force of 9.8 N.

The gravitational field strength g at any point in a gravitational field is defined as follows:

The gravitational field strength at a point is the gravitational force exerted per unit mass on a small object placed at that point.

This can be written as an equation:

$$g = \frac{F}{m}$$

where F is the gravitational force on the object and m is the mass of the object. Gravitational field strength has units of N kg^{-1}. This is equivalent to m s^{-2}.

We can use the definition to determine the gravitational field strength for a point (or spherical) mass. The force between two point masses is given by:

$$F = \frac{GMm}{r^2}$$

So, the gravitational field strength g due to the mass M at a distance of r from its centre is:

$$\begin{aligned} g &= \frac{F}{m} \\ &= \frac{GMm}{r^2 m} \\ &\Rightarrow \frac{GM\cancel{m}}{r^2\cancel{m}} \\ &= \frac{GM}{r^2} \end{aligned}$$

KEY EQUATION

The gravitation field g due to a point mass is:

$$g = \frac{GM}{r^2}$$

where G is the universal gravitational constant, M is the mass and r is the distance from the mass.

You must learn how to derive this equation using Newton's law of gravitation and your understanding of a gravitational field.

Since force is a vector quantity, it follows that gravitational field strength is also a vector. We need to give its direction as well as its magnitude in order to specify it completely. The field strength g is not a constant; it decreases as the distance r increases. The field strength obeys an inverse square law with distance. The field strength will decrease by a factor of four when the distance from the centre is doubled. Close to the Earth's surface, the magnitude of g is about 9.81 N kg^{-1}. Even if you climbed Mount Everest, which is 8.85 km high, the field strength will only decrease by 0.3%.

So the gravitational field strength g at a point depends on the mass M of the body causing the field, and the distance r from its centre (see Worked example 1).

Gravitational field strength g also has units m s^{-2}; it is an acceleration. Another name for g is 'acceleration of free fall'. Any object that falls freely in a gravitational field has this acceleration, approximately 9.8 m s^{-2} near the Earth's surface. In Chapter 2, you learned about different ways to determine an experimental value for g, the local gravitational field strength.

WORKED EXAMPLE

1 The Earth has radius 6400 km. The gravitational field strength on the Earth's surface is 9.81 N kg^{-1}. Use this information to determine the mass of the Earth and its mean density.

Step 1 Write down the quantities given:

$r = 6.4 \times 10^6$ m $g = 9.81$ N kg^{-1}

Step 2 Use the equation $g = \frac{GM}{r^2}$ to determine the mass of the Earth M.

$$g = \frac{GM}{r^2}$$

$$9.8 = \frac{6.67\times10^{11}M}{(6.4\times10^6)^2}$$

$$M = 9.8 \times \frac{(6.4\times10^6)^2}{6.67\times10^{-11}}$$

$$= 6.0 \times 10^{24}\,\text{kg}$$

Step 3 Use the equation $\text{density} = \frac{\text{mass}}{\text{volume}}$ to determine the density ρ of the Earth.

The Earth is a spherical mass. Its volume can be calculated using $\frac{4}{3}\pi r^3$

$$\rho = \frac{M}{V}$$

$$= \frac{6.0\times10^{24}}{\frac{4}{3}\times\pi\times(6.4\times10^6)^3}$$

$$= 5500\ \text{kg m}^{-3}$$

We can now see in more detail why the Earth's gravitational field may be considered to be uniform near a planet's surface. Consider a planet of radius R, then, at its surface, the gravitational field strength, $g = \frac{GM}{R^2}$ where G is the universal gravitational constant.

If we move up from the surface, a distance ΔR, the new gravitational field strength, $g_{new} = \frac{GM}{(R+\Delta R)^2}$

The percentage change in R is $\frac{\Delta R}{R} \times 100\%$ and, using a similar logic to that used when dealing with uncertainties, the percentage change in the gravitational field strength is $2 \times \frac{\Delta R}{R} \times 100\%$

WORKED EXAMPLE

2 Calculate the change in the gravitational field strength between the Earth's equator and at a height of 10 km above the equator.

(Earth's radius at the equator is 6357 km, mass of the Earth = 5.974 × 10^{24} kg

G = 6.67 × 10^{-11} Nm^{2}kg^{-2})

Gravitational field strength at the equator =

$$g = \frac{GM}{R^2} = \frac{6.67\times10^{-11}\times5.974\times10^{24}}{(6.357\times10^6)^2} = 9.860\ \text{Nkg}^{-1}$$

percentage change in the field in moving 10 km from the surface

$= \frac{\Delta R}{R} \times 100\% = \frac{10}{5974} \times 100\% = 0.167\%$

The change in the field strength = 0.167% of g = 0.167% of 9.860 = 0.01650 Nkg^{-1}

Generally the field strength is quoted to two or three significant figures (9.8 or 9.81 Nkg^{-1}). 10 km is roughly the height at which airliners cruise and even at this height there is no significant change in the field strength, indeed local variation due to different densities of the Earth's crust and the fact that the Earth is not a perfect sphere have a much larger effect.

Looking at this from a different viewpoint, if we refer back to Figure 17.6 and the spreading of the field lines, as we go further from the centre of mass we see that a change in height increases the area subtended by those lines. The area subtended increases by the square of the distance and thus moving a distance ΔR from the surface of the Earth would give a percentage increase in the area subtended by the field lines of $2 \times \frac{\Delta R}{R} \times 100\%$. Following similar logic to the worked example, we can see that a 10 km vertical movement from the Earth's surface leads to a percentage increase in area of 0.16%, which shows that over this distance the lines in Figure 17.4 (a rise of less than 10 metres) are very nearly parallel and the field is very nearly uniform.

Questions

You will need the data in Table 17.1 to answer these questions.

Body	Mass / kg	Radius / km	Distance from Earth / km
Earth	6.0×10^{24}	6 400	–
Moon	7.4×10^{22}	1 740	3.8×10^{5}
Sun	2.0×10^{30}	700 000	1.5×10^{8}

Table 17.1: Data for Questions 3 to 8.

3 Mount Everest is approximately 9.0 km high. Estimate how much less a mountaineer of mass 100 kg (including backpack) would weigh at its summit, compared to her weight at sea level. Would this difference be measurable with bathroom scales?

4 **a** Calculate the gravitational field strength:

i close to the surface of the Moon

ii close to the surface of the Sun.

b Suggest how your answers help to explain why the Moon has only a thin atmosphere, while the Sun has a dense atmosphere.

5 **a** Calculate the Earth's gravitational field strength at the position of the Moon.

b Calculate the force the Earth exerts on the Moon. Hence, determine the Moon's acceleration towards the Earth.

6 Jupiter's mass is 320 times that of the Earth and its radius is 11.2 times the Earth's. The Earth's surface gravitational field strength is 9.81 N kg^{-1}. Calculate the gravitational field strength close to the surface of Jupiter.

7 The Moon and the Sun both contribute to the tides on the Earth's oceans. Which has a bigger pull on each kilogram of seawater, the Sun or the Moon?

8 Astrologers believe that the planets exert an influence on us, particularly at the moment of birth. (They don't necessarily believe that this is an effect of gravity!)

a Calculate the gravitational force on a 4.0 kg baby caused by Mars when the planet is at its closest to the Earth at a distance of 100 000 000 km. Mars has a mass 6.4×10^{23} kg.

b Calculate the gravitational force on the same baby due to its 50 kg mother at a distance of 0.40 m.

COALI

17.3 Energy in a gravitational field

As well as the force on a mass in a gravitational field, we can think about its energy. If you lift an object from the ground, you increase its gravitational potential energy (g.p.e.). The higher you lift it, the more work you do on it and so the greater its g.p.e. The object's change in g.p.e. can be calculated as $mg\Delta h$, where Δh is the change in its height (as we saw in Chapter 5).

This approach is satisfactory when we are considering objects close to the Earth's surface. However, we need a more general approach to calculating gravitational energy, for two reasons:

- If we use g.p.e. = $mg\Delta h$, we are assuming that an object's g.p.e. is zero on the Earth's surface. This is fine for many practical purposes but not, for example, if we are considering objects moving through space, far from Earth. For these, there is nothing special about the Earth's surface.
- If we lift an object to a great height, g decreases and we would need to take this into account when calculating g.p.e.

For these reasons, we need to set up a different way of thinking about gravitational potential energy. We start by picturing a mass at infinity, that is, at an infinite distance from all other masses. We say that here the mass has zero potential energy. This is a more convenient way of defining the zero of g.p.e. than using the surface of the Earth.

Now we picture moving the mass to the point where we want to know its g.p.e. As with lifting an object from the ground, we determine the work done to move the mass to the point. The work done on it is equal to the energy transferred to it; that is, its g.p.e., and that is how we can determine the g.p.e. of a particular mass.

17.4 Gravitational potential

In practice, it is more useful to talk about the gravitational potential at a point. This tells us the g.p.e. per unit mass at the point (just as field strength g tells us the force per unit mass at a point in a field). The symbol used for potential is ϕ (Greek letter phi), and unit mass means one kilogram. **Gravitational potential** at a point is defined as the work done per unit mass bringing a unit mass from infinity to the point.

For a point mass M, we can write an equation for ϕ at a distance r from M:

$$\phi = -\frac{GM}{r}$$

where G is the gravitational constant as before. Notice the minus sign; gravitational potential is always negative. This is because, as a mass is brought towards another mass, its g.p.e. decreases. Since g.p.e. is zero at infinity, it follows that, anywhere else, g.p.e. and potential are less than zero; that is, they are negative.

KEY EQUATION

Gravitational potential:

$$\phi = -\frac{GM}{r}$$

Imagine a spacecraft coming from a distant star to visit the Solar System. The variation of the gravitational potential along its path is shown in Figure 17.8. We will concentrate on three parts of its journey:

1 As the spacecraft approaches the Earth, it is attracted towards it. The closer it gets to Earth, the lower its g.p.e. becomes and so the lower its potential.

2 As the spacecraft moves away from the Earth, it has to work against the pull of the Earth's gravity. Its g.p.e. increases and so we can say that the potential increases. The Earth's gravitational field creates a giant 'potential well' in space. We live at the bottom of that well.

3 As the spacecraft approaches the Sun, it is attracted into a much deeper well. The Sun's mass is much greater than the Earth's and so its pull is much stronger and the potential at its surface is more negative than on the Earth's surface.

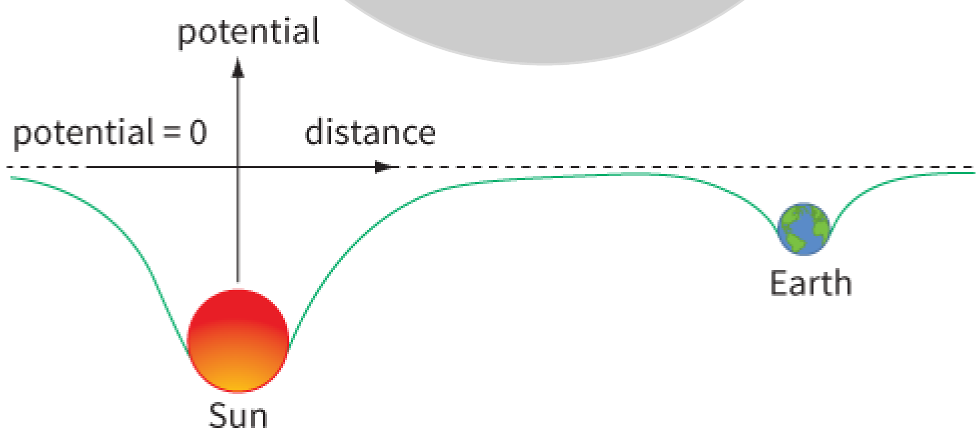


Figure 17.8: The gravitational potential is zero at infinity (far from any mass), and decreases as a mass is approached.

WORKED EXAMPLE

3 A planet has a diameter of 6800 km and a mass of 4.9×10^{23} kg. A rock of mass 200 kg, initially at rest and a long distance from the planet, accelerates towards the planet and hits the surface of the planet.

Calculate the change in potential energy of the rock and its speed when it hits the surface.

Step 1 Write down the quantities given.

$r = 3.4 \times 10^6$ m $M = 4.9 \times 10^{23}$ kg

Step 2 The equation $\phi = -\frac{GM}{r}$ gives the potential at the surface of the planet, that is, the gravitational potential energy per unit mass at that point. So the gravitational potential energy of the rock of mass *m* at that point is given by:

$$\text{g.p.e.} = -\frac{GMm}{r}$$

The g.p.e. of the rock when it is far away is zero, so the value we calculate using this equation gives the decrease in the rock's g.p.e. during its fall to hit the planet.

$$\text{change in g.p.e} = \frac{6.67\times10^{-11}\times4.9\times10^{23}\times200}{3.4\times10^{6}}$$

$$= 1.92 \times 10^9 \text{ J}$$

$$\approx 1.9 \times 10^9 \text{ J}$$

Step 3 In the absence of an atmosphere, all of the g.p.e. becomes kinetic energy of the rock, and so:

$$\tfrac{1}{2}mv^2 = 1.92 \times 10^9 \text{ J}$$

$$v = \sqrt{\frac{1.92\times10^9\times2}{200}}$$

$$= 440 \text{ m s}^{-1}$$

Note that the rock's final speed when it hits the planet does not depend on the mass of the rock. This is because, if you equate the two equations for k.e. and the change in g.p.e., the mass *m* of the rock cancels.

Questions

You will need the data for the mass and radius of the Earth and the Moon from Table 17.1 to answer this question.

Gravitational constant $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

9 **a** Determine the gravitational potential at the surface of the Earth.

b Determine the gravitational potential at the surface of the Moon.

c Which has the shallower 'potential well', the Earth or the Moon? Draw a diagram similar to Figure 17.8 to compare the 'potential wells' of the Earth and the Moon.

d Use your diagram to explain why a large rocket is needed to lift a spacecraft from the surface of the Earth but a much smaller rocket can be used to launch from the Moon's surface.

Gravitational potential difference

Very often, we consider problems where it is useful to know how much energy is needed to lift a satellite from the surface of a planet or moon to a height where the satellite can be put into orbit. The equation for the change in potential, $\phi = \frac{GM}{r}$ can be used twice, once to find the potential at the surface and once to find the potential at the orbital height. However, it is much easier to combine the two operations and use the equation:

$$\Delta\phi = GM\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

Question

10 During the manned Moon landings in the 1960s, the command module orbited the Moon in an elliptic orbit with a maximum height of 310 km above the surface of the Moon, whilst the lunar module descended and landed on the Moon's surface.

a Explain why the potential energy of the command module varied during its orbit.

b Calculate the maximum gravitational potential difference between the lunar surface and the position of the command module.

Fields: terminology

The words used to describe gravitational (and other) fields can be confusing. Remember:

- **Field strength** tells us about the **force** on unit mass at a point;
- **Potential** tells us about **potential energy** of unit mass at a point.

You will meet the idea of electric field strength in Chapter 21, where it is the force on unit charge. Similarly, when we talk about the potential difference between two points in electricity, we are talking about the difference in electrical potential energy per unit charge. In that chapter, you will meet repulsive fields as well as attractive fields and this should develop your understanding as to why the choice of infinity for the zero of potential is the only sensible choice.

17.5 Orbiting under gravity

For an object orbiting a planet, such as an artificial satellite orbiting the Earth, gravity provides the centripetal force that keeps it in orbit (Figure 17.9). This is a simple situation as there is only one force acting on the satellite–the gravitational attraction of the Earth. The satellite follows a circular path because the gravitational force is at right angles to its velocity.

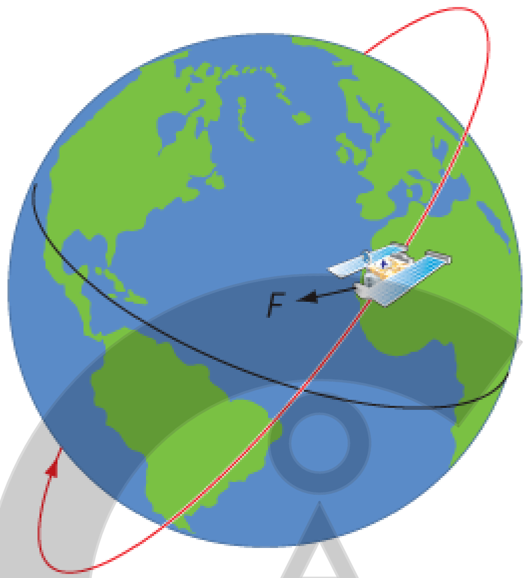


Figure 17.9: The gravitational attraction of the Earth provides the centripetal force on an orbiting satellite.

From Chapter 16, you know that the centripetal force *F* on a body is given by:

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

Consider a satellite of mass *m* orbiting the Earth at a distance *r* from the Earth's centre at a constant speed *v*. Since it is the gravitational force between the Earth and the satellite that provides this centripetal force, we can write:

$$\frac{GMm}{r^2} = \frac{mv^2}{r}$$

where *M* is the mass of the Earth. (There is no need for a minus sign here as the gravitational force and the centripetal force are both directed towards the centre of the circle.)

Rearranging gives:

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

This equation allows us to calculate, for example, the speed at which a satellite must travel to stay in a circular orbit. Notice that the mass of the satellite *m* has cancelled out. The implication of this is that all satellites, whatever their masses, will travel at the same speed in a particular orbit. You would find this very reassuring if you were an astronaut on a space walk outside your spacecraft (Figure 17.10). You would travel at the same speed as your craft, despite the fact that your mass is a lot less than its mass. The equation can be applied to the planets of our solar system – *M* becomes the mass of the Sun.

Figure 17.10: During this space walk, both the astronaut and the spacecraft travel through space at over 8 km s^{-1}.

Now look at Worked example 3.

WORKED EXAMPLE

4 The Moon orbits the Earth at an average distance of 384 000 km from the centre of the Earth. Calculate its orbital speed. (The mass of the Earth is 6.0×10^{24} kg.)

Step 1 Write down the known quantities.

$r = 3.84 \times 10^8$ m $\quad M = 6.0 \times 10^{24}$ kg $v = ?$

Step 2 Use the equation $v^2 = \frac{GM}{r}$ to determine the orbital speed v.

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

$$v^2 = \frac{6.67\times10^{-11} \times 6.0\times10^{24}}{3.84\times10^{8}}$$

$$v^2 = 1.04 \times 10^6$$

$$v = \sqrt{1.04 \times 10^6}$$

$$= 1020 \text{ m s}^{-1}$$

$$\approx 1.0 \times 10^3 \text{ m s}^{-1}$$

So, the Moon travels around its orbit at a speed of roughly 1 km s^{-1}.

Question

11 Calculate the orbital speed of an artificial satellite travelling 200 km above the Earth's surface. (The radius of Earth is 6.4×10^6 m and its mass is 6.0×10^{24} kg.)

17.6 The orbital period

It is often more useful to consider the time taken for a complete orbit, the **orbital period** T.

Since the distance around an orbit is equal to the circumference $2\pi r$, it follows that:

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

We can substitute this in the equation for v^2.

This gives:

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

and rearranging this equation gives:

$$T^2 = \left(\frac{4\pi^2}{GM}\right) r^3$$

$$\text{or } \sqrt{\frac{4\pi r^3}{GM}}$$

This equation shows that the orbital period T is related to the radius r of the orbit. The square of the period is directly proportional to the cube of the radius ($T^2 \propto r^3$). This is an important result. It was first discovered by Johannes Kepler, who analysed the available data for the planets of the Solar System. It was an empirical law (one based solely on experiment) since he had no theory to explain why there should be this relationship between T and r. It was not until Isaac Newton formulated his law of gravitation that it was possible to explain this fact.

17.7 Orbiting the Earth

The Earth has one natural satellite – the Moon – and many thousands of artificial satellites – some spacecraft and a lot of debris. Each of these satellites uses the Earth's gravitational field to provide the centripetal force that keeps it in orbit. In order for a satellite to maintain a particular orbit, it must travel at the correct speed. This is given by the equation in topic 17.5 Orbiting under gravity:

$$V^2 = \frac{GM}{r}$$

It follows from this equation that, the closer the satellite is to the Earth, the faster it must travel. If it travels too slowly, it will fall down towards the Earth's surface. If it travels too quickly, it will move out into a higher orbit.

Question

12 A satellite orbiting a few hundred kilometres above the Earth's surface will experience a slight frictional drag from the Earth's (very thin) atmosphere. Draw a diagram to show how you would expect the satellite's orbit to change as a result. How can this problem be overcome if it is desired to keep a satellite at a particular height above the Earth?

Observing the Earth

Artificial satellites have a variety of uses. Many are used for making observations of the Earth's surface for commercial, environmental, meteorological or military purposes. Others are used for astronomical observations, benefiting greatly from being above the Earth's atmosphere. Still others are used for navigation, telecommunications and broadcasting.

Figure 17.11 shows two typical orbits. A satellite in a circular orbit close to the Earth's surface, and passing over the poles, completes about 16 orbits in 24 hours. As the Earth turns below it, the satellite 'sees' a different strip of the Earth's surface during each orbit. A satellite in an elliptical orbit has a more distant view of the Earth.

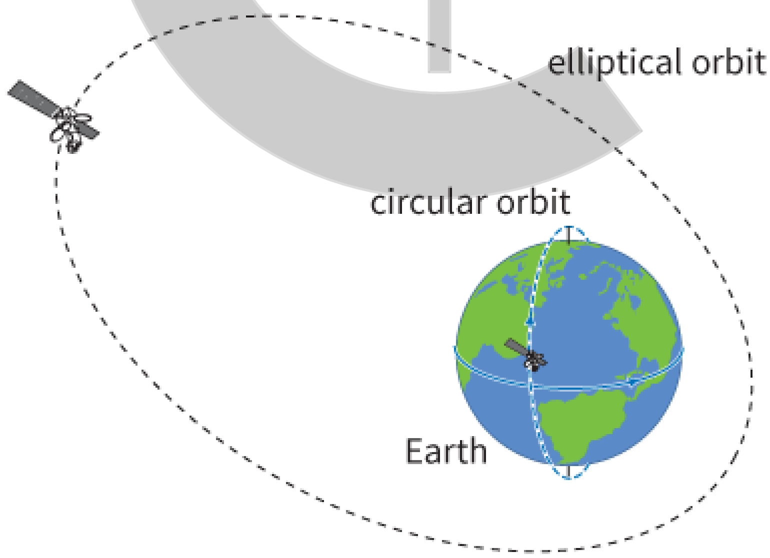


Figure 17.11: Satellites orbiting the Earth.

Geostationary orbits

A special type of orbit is one in which a satellite travels from west to east and is positioned so that, as it orbits, the Earth rotates below it with the same angular speed. The satellite remains above a fixed point on the Earth's equator. This kind of orbit is called a **geostationary orbit**. There are over 300 satellites in such orbits. They are used for telecommunications (transmitting telephone messages around the world) and for satellite television transmissions. A base station on Earth sends the TV signal up to the satellite, where it is amplified and broadcast back to the ground. Satellite receiver dishes are a familiar sight; you will have observed how, in a neighbourhood, they all point towards the same point in the sky. Because the satellite is in a geostationary orbit, the dish can be fixed. Satellites in any other orbits move across the sky so that a tracking system is necessary to communicate with them. Such a system is complex and expensive, and too demanding for the domestic market.

Geostationary satellites have a lifetime of perhaps ten years. They gradually drift out of the correct orbit, so they need a fuel supply for the rocket motors that return them to their geostationary position, and that keep them pointing correctly towards the Earth. Eventually, they run out of fuel and need to be replaced.

We can determine the distance of a satellite in a geostationary orbit using the equation:

$$T^2 = \left(\frac{4\pi^2}{GM}\right)r^3$$

For a satellite to stay above a fixed point on the equator, it must take exactly 24 hours to complete one orbit (Figure 17.12).

We know:

$$G = 6.67 \times 10^{-11}\ \mathrm{N\ m^2\ kg^{-2}}$$
$$T = 24\ \text{hours} = 86\,400\ \mathrm{s}$$
$$M = 6.0 \times 10^{24}\ \mathrm{kg}$$
$$T^2 = \left(\frac{4\pi^2}{GM}\right)r^3$$

$$86\,400^2 = \left(\frac{4\pi^2}{6.67\times10^{-11}\times6.0\times10^{24}}\right)r^3$$
$$r^3 = \frac{86\,400^2}{\left(\frac{4\pi^2}{6.67\times10^{-11}\times6.0\times10^{24}}\right)}$$
$$r^3 = 7.57\times10^{22}\ \mathrm{m^3}$$
$$r = \sqrt[3]{7.57\times10^{22}}$$
$$\approx 4.23\times10^{7}\ \mathrm{m}$$

So, for a satellite to occupy a geostationary orbit, it must be at a distance of 42 300 km from the centre of the Earth and at a point directly above the equator. Note that the radius of the Earth is 6400 km, so the orbital radius is 6.6 Earth radii from the centre of the Earth (or 5.6 Earth radii from its surface). Figure 17.12 has been drawn to give an impression of the size of the orbit.

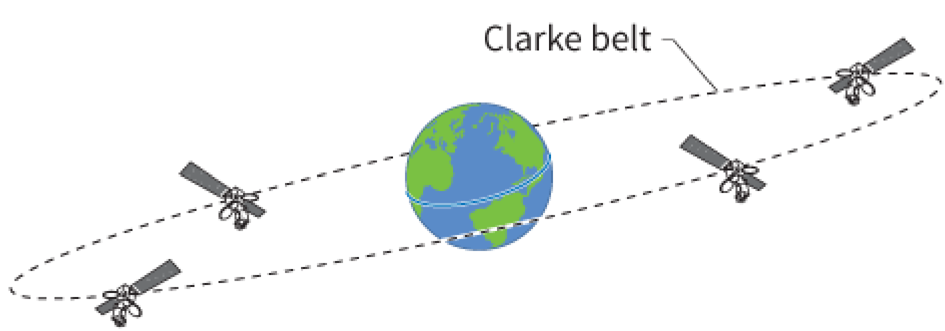


Figure 17.12: Geostationary satellites are parked in the 'Clarke belt', high above the equator. This is a perspective view; the Clarke belt is circular.

Questions

13 For any future mission to Mars, it would be desirable to set up a system of three or four geostationary (or 'martostationary') satellites around Mars to allow communication between the planet and Earth. Calculate the radius of a suitable orbit around Mars.

Mars has mass 6.4×10^{23} kg and a rotational period of 24.6 hours.

14 Although some international telephone signals are sent via satellites in geostationary orbits, most are sent along cables on the Earth's surface. This reduces the time delay between sending and receiving the signal. Estimate this time delay for communication via a satellite, and explain why it is less significant when cables are used.

You will need the following:

- radius of geostationary orbit = 42 300 km
- radius of Earth = 6400 km
- speed of electromagnetic waves in free space $c = 3.0 \times 10^8$ m s^{-1}

REFLECTION

In Question 7, we established that the gravitational force on each kilogram of water on the Earth's surface is greater than that of the Moon; in fact, it is about 175 times greater. However, the Moon affects the tides more than the Sun affects the tides. This seems contradictory.

Find out why and prepare a short paper explaining why.

What were the most interesting discoveries you made while working on this problem?

SUMMARY

The force of gravity is an attractive force between any two objects due to their masses.

The gravitational field strength *g* at a point is the gravitational force exerted per unit mass on a small object placed at that point:

$$g = \frac{F}{m}$$

The external field of a uniform spherical mass is the same as that of an equal point mass at the centre of the sphere.

Newton's law of gravitation states that:

Any two point masses attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of their separation.

The equation for Newton's law of gravitation is:

$$F = \frac{GMm}{r^2}$$

The gravitational field strength at a point is the gravitational force exerted per unit mass on a small object placed at that point:

$$g = \frac{GM}{r^2}$$

On or near the surface of the Earth, the gravitational field is uniform, so the value of *g* is approximately constant. Its value is equal to the acceleration of free fall.

The gravitational potential at a point is the work done in bringing unit mass from infinity to that point.

The gravitational potential of a point mass is given by:

$$\phi = -\frac{GM}{r}$$

The orbital period of a satellite is the time taken for one orbit.

The orbital period can be found by equating the gravitational force $\frac{GM}{r^2}$ to the centripetal force $\frac{mv^2}{r}$.

The orbital speed of a planet or satellite can be determined using the equation:

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

Geostationary satellites have an orbital period of 24 hours and are used for telecommunications transmissions and for television broadcasting.

EXAM-STYLE QUESTIONS

1 An astronaut is on a planet of mass $0.50M_E$ and radius $0.75r_E$, where M_E is the mass of the Earth and r_E is the radius of the Earth.

What is the gravitational field strength at the surface of the planet? **[1]**

A 6.5 N kg^{-1}

B 8.7 N kg^{-1}

C 11 N kg^{-1}

D 12 N kg^{-1}

2 Consider the dwarf planet Pluto to be an isolated sphere of radius 1.2×10^6 m and mass of 1.27×10^{22} kg.

What is the gravitational potential at the surface of Pluto? **[1]**

A −0.59 J kg^{-1}

B -7.1×10^5 J kg^{-1}

C 0.59 J kg^{-1}

D 7.1×10^5 J kg^{-1}

3 Two small spheres each of mass 20 g hang side by side with their centres 5.00 mm apart. Calculate the gravitational attraction between the two spheres. **[3]**

4 It is suggested that the mass of a mountain could be measured by the deflection from the vertical of a suspended mass. This diagram shows the principle.

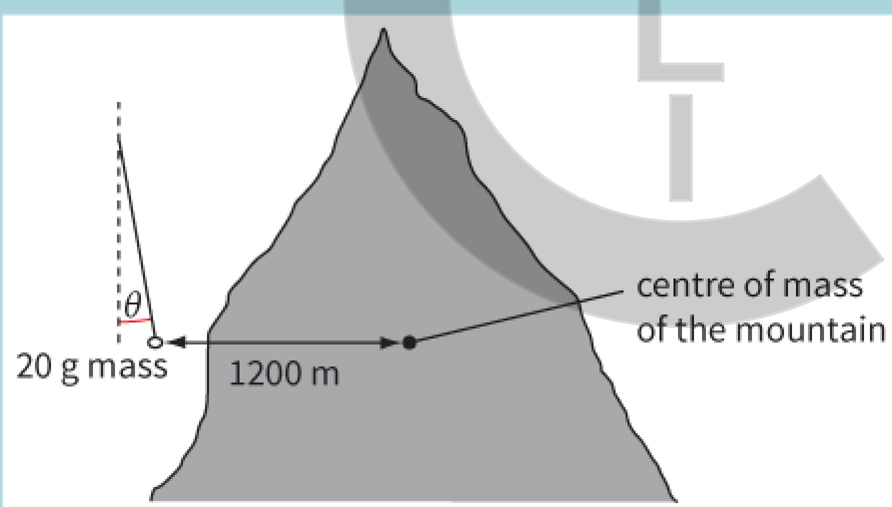


Figure 17.13

a Copy the diagram and draw arrows to represent the forces acting on the mass. Label the arrows. **[3]**

b The whole mass of the mountain, 3.8×10^{12} kg, may be considered to act at its centre of mass. Calculate the horizontal force on the mass due to the mountain. **[2]**

c Compare the force calculated in part **b** with the Earth's gravitational force on the mass. **[2]**

[Total: 7]

5 This diagram shows the Earth's gravitational field.

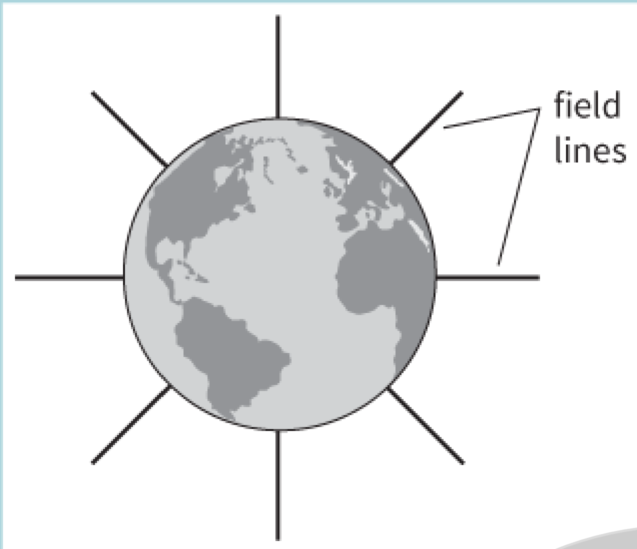


Figure 17.14

a Copy the diagram and add arrows to show the direction of the field. **[1]**

b Explain why the formula for potential energy gained ($mg\Delta h$) can be used to find the increase in potential energy when an aircraft climbs to a height of 10 000 m, but cannot be used to calculate the increase in potential energy when a spacecraft travels from the Earth's surface to a height of 10 000 km. **[2]**

[Total: 3]

6 Mercury, the smallest of the eight recognised planets, has a diameter of 4.88×10^6 m and a mean density of 5.4×10^3 kg m^{-3}.

a Calculate the gravitational field at its surface. **[5]**

b A man has a weight of 900 N on the Earth's surface. What would his weight be on the surface of Mercury? **[2]**

[Total: 7]

7 Calculate the potential energy of a spacecraft of mass 250 kg when it is 20 000 km from the planet Mars. (Mass of Mars = 6.4×10^{23} kg, radius of Mars = 3.4×10^6 m.) **[3]**

8 Ganymede is the largest of Jupiter's moons, with a mass of 1.48×10^{23} kg. It orbits Jupiter with an orbital radius of 1.07×10^6 km and it rotates on its own axis with a period of 7.15 days. It has been suggested that to monitor an unmanned landing craft on the surface of Ganymede a geostationary satellite should be placed in orbit around Ganymede.

a Calculate the orbital radius of the proposed geostationary satellite. **[2]**

b Suggest a difficulty that might be encountered in achieving a geostationary orbit for this moon. **[1]**

[Total: 3]

9 The Earth orbits the Sun with a period of 1 year at an orbital radius of 1.50×10^{11} m. Calculate:

a the orbital speed of the Earth **[3]**

b the centripetal acceleration of the Earth **[2]**

c the Sun's gravitational field strength at the Earth. **[1]**

[Total: 6]

10 The planet Mars has a mass of 6.4×10^{23} kg and a diameter of 6790 km.

a **i** Calculate the acceleration due to gravity at the planet's surface. **[2]**

ii Calculate the gravitational potential at the surface of the planet. **[2]**

b A rocket is to return some samples of Martian material to Earth. Write down how much energy each kilogram of matter must be given to escape completely from Mars' gravitational field. **[1]**

c Use your answer to part **b** to show that the minimum speed that the rocket must reach to escape from the gravitational field is 5000 m s^{-1}. **[2]**

d Suggest why it has been proposed that, for a successful mission to Mars, the craft that takes the astronauts to Mars will be assembled at a space station in Earth orbit and launched from there, rather than from the Earth's surface. [2]

[Total: 9]

11 **a** Explain what is meant by the **gravitational potential at a point**. **[2]**

b This diagram shows the gravitational potential near a planet of mass *M* and radius *R*.

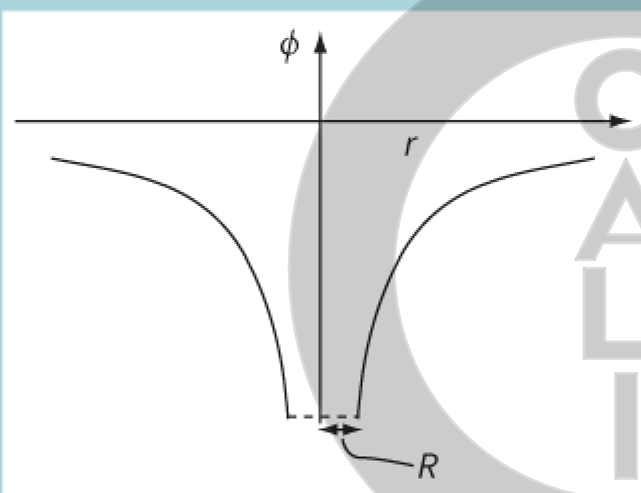


Figure 17.15

On a copy of the diagram, draw similar curves:

i for a planet of the same radius but of mass 2*M*–label this i. **[2]**

ii for a planet of the same mass but of radius 2*R*–label this ii. **[2]**

c Use the graphs to explain from which of these three planets it would require the least energy to escape. **[2]**

d Venus has a diameter of 12 100 km and a mass of 4.87×10^{24} kg.

Calculate the energy needed to lift one kilogram from the surface of Venus to a space station in orbit 900 km from the surface. **[4]**

[Total: 12]

12 **a** Explain what is meant by the gravitational field strength at a point. **[2]**

This diagram shows the dwarf planet, Pluto, and its moon, Charon. These can be considered to be a double planetary system orbiting each other about their joint centre of mass.

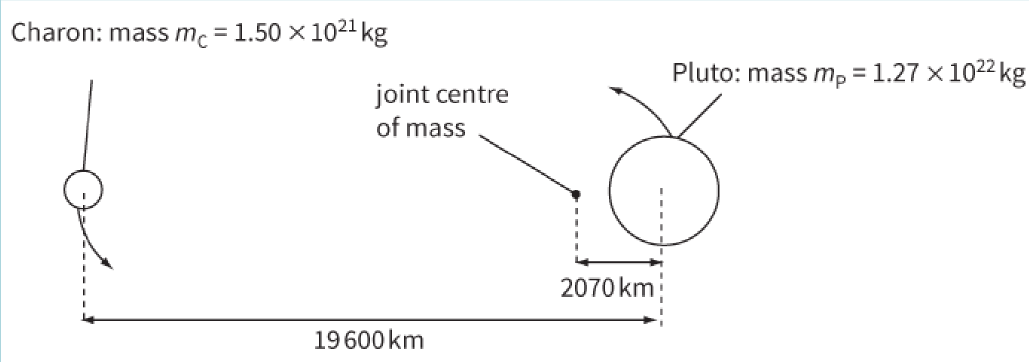


Figure 17.16

b Calculate the gravitational pull on Charon due to Pluto. **[3]**

c Use your result to part **b** to calculate Charon's orbital period. **[3]**

d Explain why Pluto's orbital period must be the same as Charon's. **[1]**

[Total: 9]

13 This diagram shows the variation of the Earth's gravitational field strength with distance from its centre.

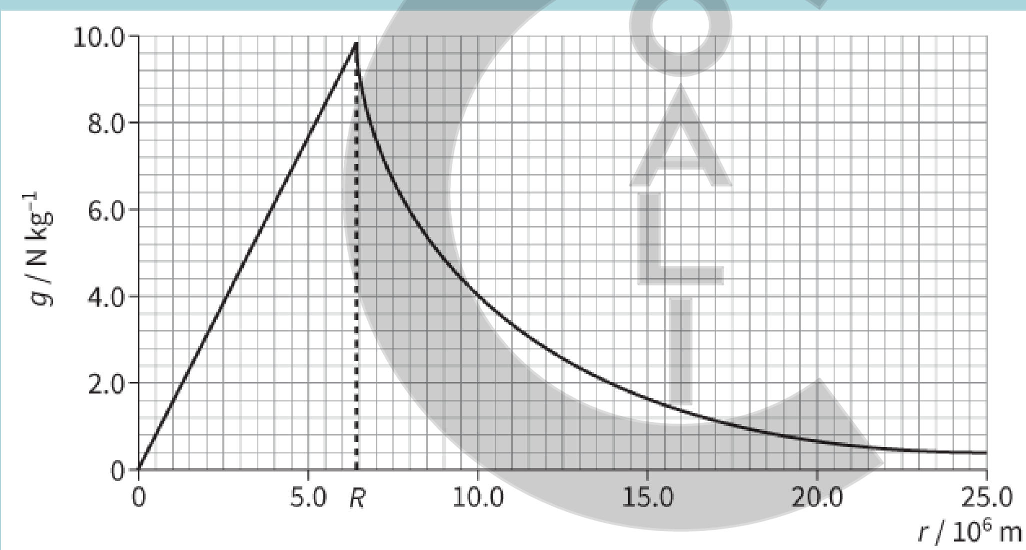


Figure 17.17

a Determine the gravitational field strength at a height equal to 2*R* above the Earth's surface, where *R* is the radius of the Earth. **[1]**

b A satellite is put into an orbit at this height. State the centripetal acceleration of the satellite. **[1]**

c Calculate the speed at which the satellite must travel to remain in this orbit. **[2]**

d **i** Frictional forces mean that the satellite gradually slows down after it has achieved a circular orbit. Draw a diagram of the initial circular orbital path of the satellite, and show the resulting orbit as frictional forces slow the satellite down. **[1]**

ii Suggest and explain why there is not a continuous bombardment of old satellites colliding with the Earth. **[2]**

[Total: 7]

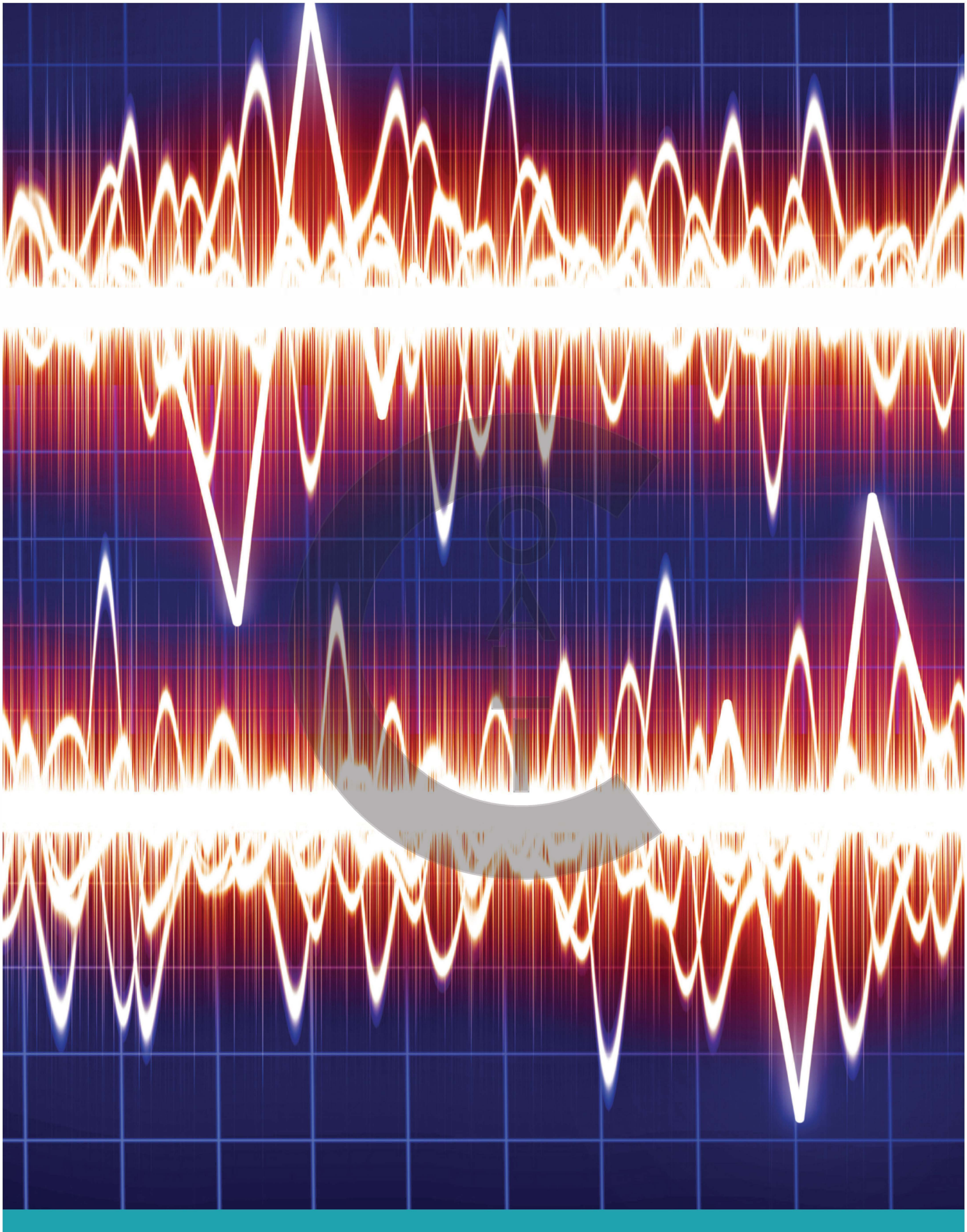
COALI

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the nature of the gravitational field	17.1			
represent and interpret a gravitational field using field lines	17.1			
recall and use Newton's law of gravitation: $F = \frac{Gm_1m_2}{r^2}$	17.1			
understand why *g* is approximately constant near the Earth's surface	17.2			
derive from Newton's law of gravitation: $g = \frac{GM}{r^2}$	17.2			
recall and use the equation: $g = \frac{GM}{r^2}$	17.2			
understand that the gravitational potential at infinity is zero	17.3			
define gravitational potential at a point, φ, as the work done in bringing unit mass from infinity to that point	17.4			
understand that the gravitational potential decreases, being more negative, as an object moves closer to a second object	17.4			
recall and use the formula that the gravitational potential $\phi = -\frac{GM}{r}$	17.4			
use the formula: $\phi = GM\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$	17.4			
understand that the potential energy of two point masses is equal to: $F = -\frac{Gm_1m_2}{r}$	17.1			
solve problems involving circular orbits of satellites by relating the gravitational force to the centripetal acceleration of the satellite	17.5			
understand that a satellite in a geostationary satellite remains above the same point on the Earth's surface	17.7			

I can	See topic…	Needs more work	Almost there	Ready to move on
understand that a geostationary satellite has an orbital period of 24 hours and travels from west to east.	17.7			



〉Chapter 18

Oscillations

LEARNING INTENTIONS

In this chapter you will learn how to:

- give examples of free and forced oscillations
- use appropriate terminology to describe oscillations
- use the equation $a = -\omega^2 x$ to define simple harmonic motion (s.h.m.)
- recall and use equations for displacement and velocity in s.h.m.
- draw and use graphical representations of s.h.m.
- describe energy changes during s.h.m.
- recall and use $E = \frac{1}{2}m\omega^2 x_0$ where E is the total energy of a system undergoing simple harmonic motion
- describe the effects of damping on oscillations and draw graphs showing these effects
- understand that resonance involves a maximum amplitude of oscillation
- understand that resonance occurs when an oscillating system is forced to oscillate at its natural frequency.

BEFORE YOU START

- Look at objects that vibrate or move in repetitive motion, such as the pendulum of a clock, a mass on the end of a spring, a toy yo-yo on a string, a branch of a tree in a breeze or an insect's wings. Write down three examples and be ready to share these with the class.
- What does their motion have in common? What differences are there between them? Discuss with a partner.

OSCILLATIONS AND ENGINEERING

When designing new products, designers need to consider unwanted oscillations in the product, whether that be a new vacuum cleaner, a new bridge, a new electric toothbrush or a new aircraft.

Figure 18.1 shows the Millennium Bridge over the river Thames in London. The bridge was revolutionary: a suspension bridge without the large supporting towers that are generally integral to the design of suspension bridges. It was built to celebrate the new millennium and it opened on 10 June 2000; the first new crossing built over the Thames for over 100 years.

Unfortunately, it had to be closed two days later when engineers detected that, when there were a lot of people walking across the bridge, it started to sway and twist. The effect was made worse because people automatically adjusted their walking so that each pace coincided with the movements of the bridge.

Figure 18.1: The Millenium Bridge across the Thames, London.

It took nearly two years before the engineers were able to fix the problem and for the bridge to be reopened – at a cost of nearly five million pounds!

Why do you think a designer, designing a new electric toothbrush, should be aware of the effects of oscillations?

It is not only large oscillations that are dangerous; very small, repeated vibrations will cause cracks to form in metals. You will be aware that if you bend a thin sheet of metal back and forth a few times it becomes easier to bend – a few more times and it will break. This is an extreme example of metal fatigue. Much smaller vibrations repeated often enough will cause microscopic cracks to form at points of high stress. These microscopic cracks will widen and, eventually, the structure will fail.

The first passenger jet airliner, the de Haviland Comet, suffered from this problem and after two of the aircraft disintegrated in mid-air, all Comets were grounded. After meticulous investigation, engineers concluded that the most likely cause of the accidents was metal fatigue at the high stress points near the corners of the 'almost', square windows. You may have noticed that the windows in more modern airliners are more oval in shape that avoids the high stress points found at the corners of a square shape.

18.1 Free and forced oscillations

Oscillations and vibrations are everywhere. A bird in flight flaps its wings up and down. An aircraft's wings also vibrate up and down, but this is not how it flies. The wings are long and thin, and they vibrate slightly because they are not perfectly rigid. Many other structures vibrate – bridges when traffic flows across, buildings in high winds.

A more specific term than vibration is **oscillation**. An object **oscillates** when it moves back and forth repeatedly, on either side of some equilibrium position. If we stop the object from oscillating, it returns to the equilibrium position.

We make use of oscillations in many different ways – for pleasure (a child on a swing), for music (the vibrations of a guitar string), for timing (the movement of a pendulum or the vibrations of a quartz crystal). Whenever we make a sound, the molecules of the air oscillate, passing the sound energy along. The atoms of a solid vibrate more and more as the temperature rises.

These examples of oscillations and vibrations may seem very different from one another. In this chapter, we will look at the characteristics that are shared by many oscillations.

Free or forced?

Free

The easiest oscillations to understand are free oscillations. If you pluck a guitar string, it continues to vibrate for some time after you have released it. The guitar string vibrates at a particular frequency (the number of vibrations per unit time). This is called its **natural frequency** of vibration, and it gives rise to the particular note that you hear. Change the length of the string, and you change the natural frequency. Every oscillator has a natural frequency of vibration, the frequency with which it vibrates freely after an initial disturbance.

Forced

Many objects can be forced to vibrate. If you sit on a bus, you may notice that the vibrations from the engine are transmitted to your body, causing you to vibrate with the same frequency. These are not free vibrations of your body; they are forced vibrations. Their frequency is not the natural frequency of vibration of your body, but the forcing frequency of the bus.

In the same way, you can force a metre ruler to oscillate by waving it up and down; however, its natural frequency of vibration will be much greater than this, as you will discover if you hold one end down on the bench and then quickly push down and let go of the other end (Figure 18.2).

Figure 18.2: A ruler vibrating freely at its natural frequency.

Question

1 State which of the following are free oscillations, and which are forced:

- **a** the wing beat of a mosquito
- **b** the movement of the pendulum in a upright clock
- **c** the vibrations of a cymbal after it has been struck
- **d** the shaking of a building during an earthquake.

18.2 Observing oscillations

Many oscillations are too rapid or too small for us to observe. Our eyes cannot respond rapidly enough if the frequency of oscillation is more than about 5 Hz (five oscillations per second); anything faster than this appears as a blur. In order to see the general characteristics of oscillating systems, we need to find suitable systems that oscillate slowly. Practical Activity 18.1 describes three suitable situations to investigate.

PRACTICAL ACTIVITY 18.1

Observing slow oscillations

A mass–spring system

A trolley, loaded with extra masses, is tethered by identical springs in between two clamps (Figure 18.3). Move the trolley to one side and it will oscillate back and forth along the bench. Listen to the sound of the trolley moving. Where is it moving fastest? What happens to its speed as it reaches the ends of its oscillation? What is happening to the springs as the trolley oscillates?

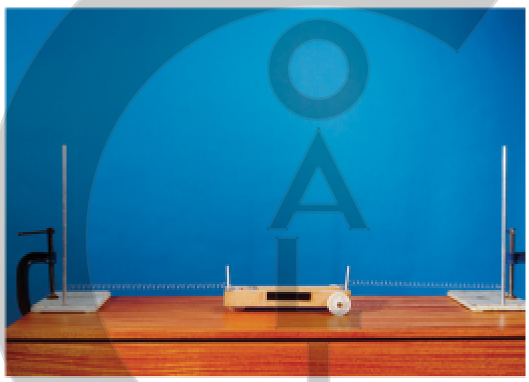


Figure 18.3: A trolley tethered between springs will oscillate freely from side to side.

A long pendulum

A string, at least 2 m long, hangs from the ceiling with a large mass fixed at the end (Figure 18.4). Pull the mass some distance to one side, and let go. The pendulum will swing back and forth at its natural frequency of oscillation. Try to note the characteristics of its motion. In what ways is it similar to the motion of the oscillating trolley? In what ways is it different?

A loudspeaker cone

A signal generator, set to a low frequency (say, 1 Hz), drives a loudspeaker so that it vibrates (Figure 18.5). You need to be able to see the cone of the loudspeaker.

How does this motion compare with that of the pendulum and the mass–spring system? Try using a higher frequency (say, 100 Hz). Use an electronic stroboscope flashing at a similar frequency to show up the movement of the cone. (It may help to paint a white spot on the centre of the cone.) Do you observe the same pattern of movement?

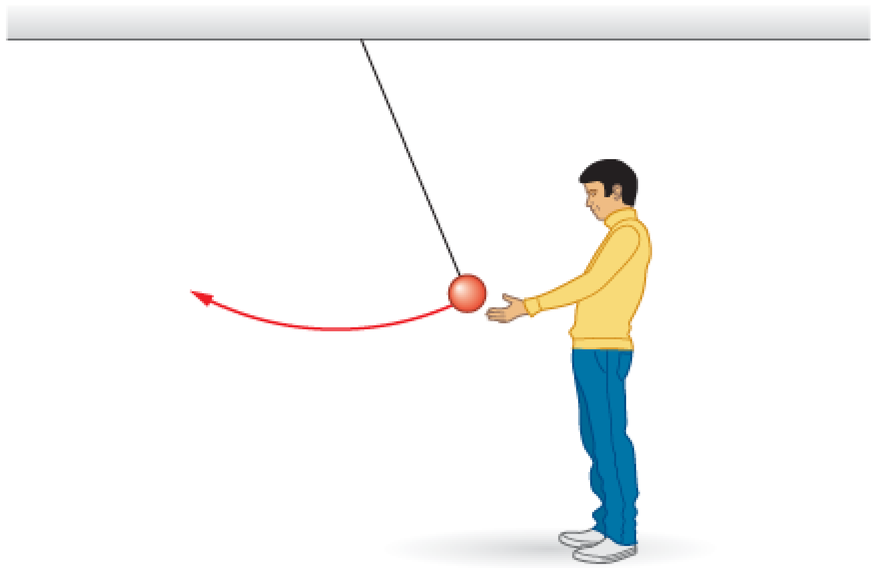


Figure 18.4: A long pendulum oscillates back and forth.

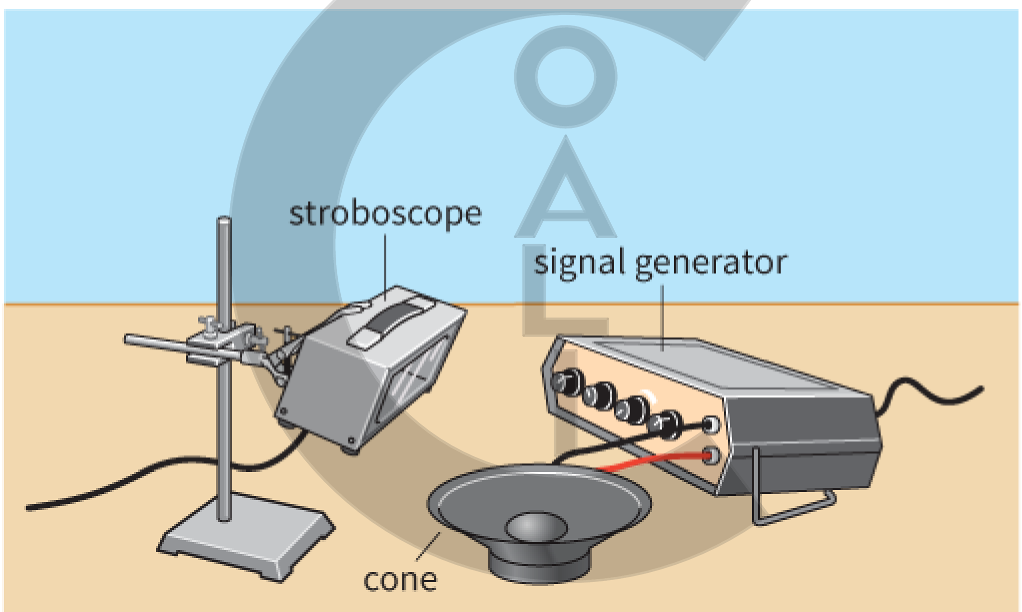


Figure 18.5: A loudspeaker cone forced to vibrate up and down.

Question

2 If you could draw a velocity–time graph for any of the oscillators described in Practical Activity 18.1, what would it look like? Would it be a curve like the one shown in Figure 18.6a, or triangular (saw-toothed) like the one shown in Figure 18.6b?

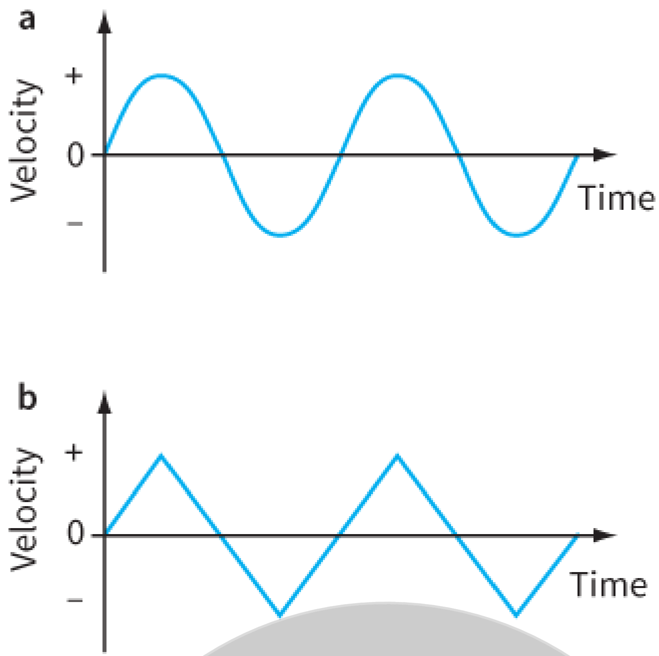


Figure 18.6: Two possible velocity–time graphs for vibrating objects.

18.3 Describing oscillations

All of the examples discussed so far show the same pattern of movement. The trolley accelerates as it moves towards the centre of the oscillation. It is moving fastest at the centre. It decelerates as it moves towards the end of the oscillation. At the extreme position, it stops momentarily, reverses its direction and accelerates back towards the centre again.

Amplitude, period and frequency

Many oscillating systems can be represented by a displacement–time graph like that shown in Figure 18.7. The displacement *x* varies in a smooth way on either side of the midpoint. The shape of this graph is a sine curve, and the motion is described as **sinusoidal**.

Notice that the displacement changes between positive and negative values, as the object moves through the equilibrium position. The maximum displacement from the equilibrium position is called the **amplitude** x_0 of the oscillation.

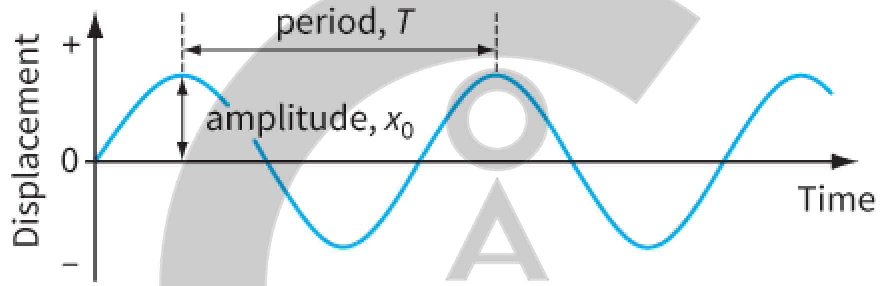


Figure 18.7: A displacement–time graph to show the meaning of amplitude and period.

The displacement–time graph can also be used to show the **period** and frequency of the oscillation. The period *T* is the time for one complete oscillation. Note that the oscillating object must go from one side to the other and back again (or the equivalent). The frequency *f* is the number of oscillations per unit time, and so *f* is the reciprocal of *T*:

$$\text{frequency} = \frac{1}{\text{period}} \equiv f = \frac{1}{T}$$

The equation can also be written as:

$$\text{period} = \frac{1}{\text{frequency}} \equiv T = \frac{1}{f}$$

Question

3 From the displacement–time graph shown in Figure 18.8, determine the amplitude, period and frequency of the oscillations represented.

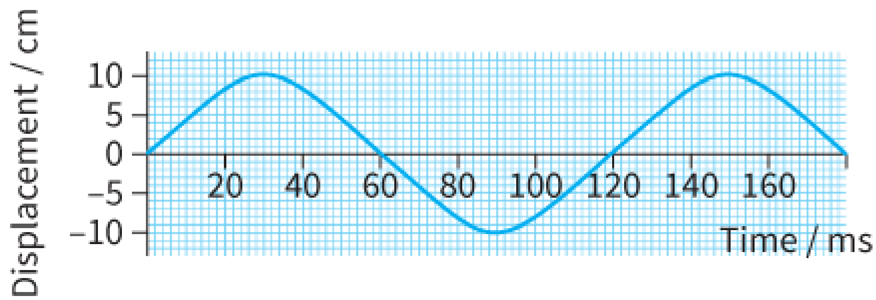


Figure 18.8: A displacement–time graph for an oscillator.

Phase

The term **phase** describes the point that an oscillating mass has reached within the complete cycle of an oscillation. It is often important to describe the **phase difference** between two oscillations. The graph of Figure 18.9a shows two oscillations that are identical except for their phase difference. They are out of step with one another. In this example, they have a phase difference of one-quarter of an oscillation. Phase difference can be measured as a fraction of an oscillation, in degrees or in radians (see Worked example 1).

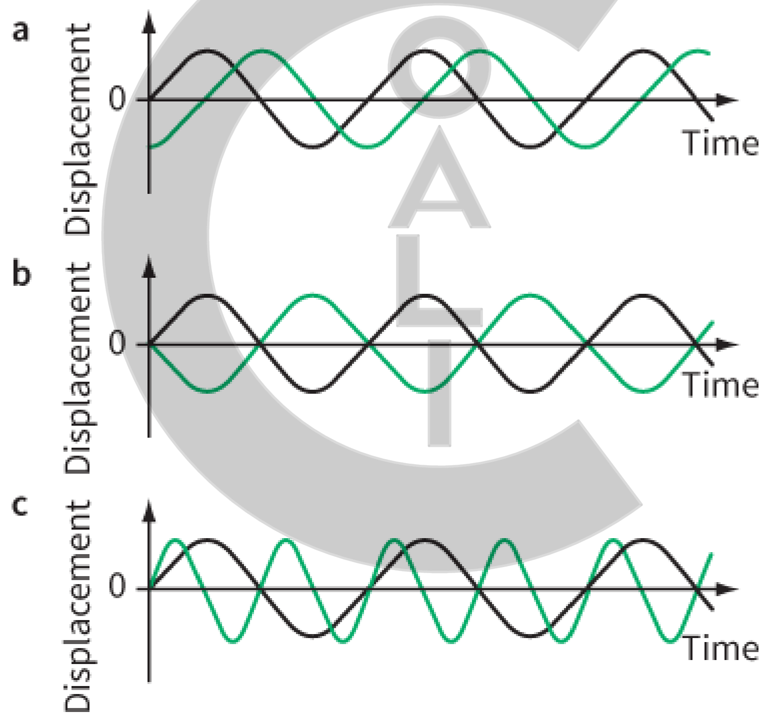


Figure 18.9: Illustrating the idea of phase difference.

WORKED EXAMPLE

1 Figure 18.10 shows displacement–time graphs for two identical oscillators. Calculate the phase difference between the two oscillations. Give your answer in degrees and in radians.

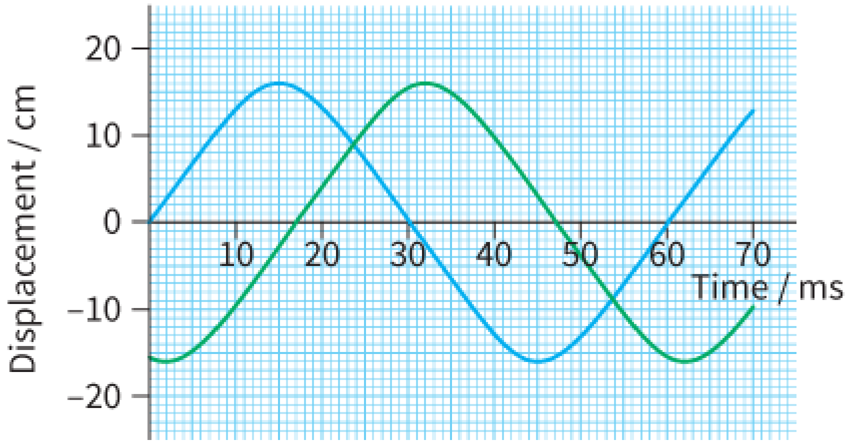


Figure 18.10: The displacement–time graphs of two oscillators with the same period.

Step 1 Measure the time interval t between two corresponding points on the graphs.

t = 17 ms

Step 2 Determine the period T for one complete oscillation.

T = 60 ms

Hint: *Remember that a complete oscillation is when the object goes from one side to the other and back again.*

Step 3 Now you can calculate the phase difference as a fraction of an oscillation.

phase difference = fraction of an oscillation

Therefore:

$$\text{phase difference} = \frac{t}{T} = \frac{17}{60} = 0.283 \text{ oscillations}$$

Step 4 Convert to degrees and radians. There are 360° and 2π rad in one oscillation.

$$\text{phase difference} = 0.283 \times 360° = 102° \approx 100°$$

$$\text{phase difference} = 0.283 \times 2\pi \text{ rad} = 1.78 \text{ rad} \approx 1.8 \text{ rad}$$

Question

4 **a** Figure 18.9b shows two oscillations that are out of phase. By what fraction of an oscillation are they out of phase?

b Why would it not make sense to ask the same question about Figure 18.9c?

18.4 Simple harmonic motion

There are many situations where we can observe the special kind of oscillations called **simple harmonic motion (s.h.m.)**. Some are more obvious than others. For example, the vibrating strings of a musical instrument show s.h.m. When plucked or bowed, the strings move back and forth about the equilibrium position of their oscillation. The motion of the tethered trolley in Figure 18.3 and that of the pendulum in Figure 18.4 are also s.h.m. (Simple harmonic motion is defined in terms of the acceleration and displacement of an oscillator – see topic 18.5 Representing s.h.m. graphically.)

Here are some other, less obvious, situations where simple harmonic motion can be found:

- When a pure (single tone) sound wave travels through air, the molecules of the air vibrate with s.h.m.
- When an alternating current flows in a wire, the electrons in the wire vibrate with s.h.m.
- There is a small alternating electric current in a radio or television aerial when it is tuned to a signal in the form of electrons moving with s.h.m.
- The atoms that make up a molecule vibrate with s.h.m. (see, for example, the hydrogen molecule in Figure 18.11a).

Oscillations can be very complex, with many different frequencies of oscillation occurring at the same time. Examples include the vibrations of machinery, the motion of waves on the sea and the vibration of a solid crystal formed when atoms, ions or molecules bond together (Figure 18.11b). It is possible to break down a complex oscillation into a sum of simple oscillations, and so we will focus our attention in this chapter on s.h.m. with only one frequency. We will also concentrate on large-scale mechanical oscillations, but you should bear in mind that this analysis can be extended to the situations already mentioned, and many more besides.

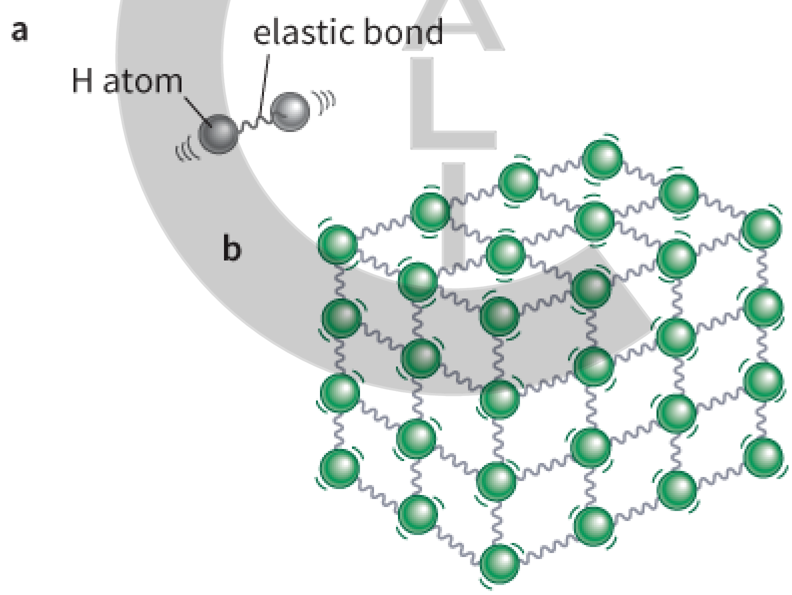


Figure 18.11: We can think of the bonds between atoms as being springy; this leads to vibrations, **a** in a molecule of hydrogen and **b** in a solid crystal.

The requirements for s.h.m.

If a simple pendulum is undisturbed, it is in equilibrium. The string and the mass will hang vertically. To start the pendulum swinging (Figure 18.12), the mass must be pulled to one side of its equilibrium position. The forces on the mass are unbalanced and so it moves back towards its equilibrium position. The mass swings past this point and continues until it comes to rest momentarily at the other side; the process is then repeated in the opposite direction. Note that a complete oscillation in Figure 18.12 is from right to left and back again.

The three requirements for s.h.m. of a mechanical system are:

- a mass that oscillates
- a position where the mass is in equilibrium
- a restoring force that acts to return the mass to its equilibrium position; the restoring force *F* is directly proportional to the displacement *x* of the mass from its equilibrium position and is directed towards that point.

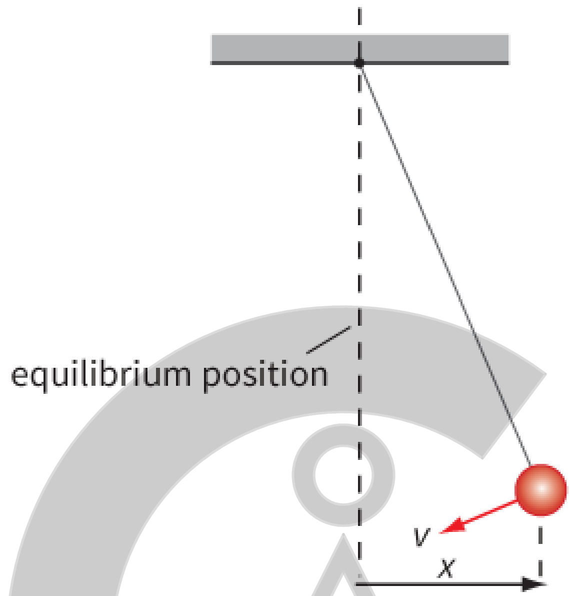


Figure 18.12: This swinging pendulum has positive displacement *x* and negative velocity *v*.

The changes of velocity in s.h.m.

As the pendulum swings back and forth, its velocity is constantly changing. As it swings from right to left (as shown in Figure 18.12) its velocity is negative. It accelerates towards the equilibrium position and then decelerates as it approaches the other end of the oscillation. It has positive velocity as it swings back from left to right. Again, it has maximum speed as it travels through the equilibrium position and decelerates as it swings up to its starting position.

This pattern of acceleration–deceleration–changing direction–acceleration again is characteristic of simple harmonic motion. There are no sudden changes of velocity. In the next topic, we will see how we can observe these changes and how we can represent them graphically.

Questions

5 Identify the features of the motion of the trolley in Figure 18.3 that satisfy the three requirements for s.h.m.

6 Explain why the motion of someone jumping up and down on a trampoline is not simple harmonic motion. (Their feet lose contact with the trampoline during each bounce.)

18.5 Representing s.h.m. graphically

If you set up a trolley tethered between springs (Figure 18.13) you can hear the characteristic rhythm of s.h.m. as the trolley oscillates back and forth. By adjusting the mass carried by the trolley, you can achieve oscillations with a period of about two seconds.

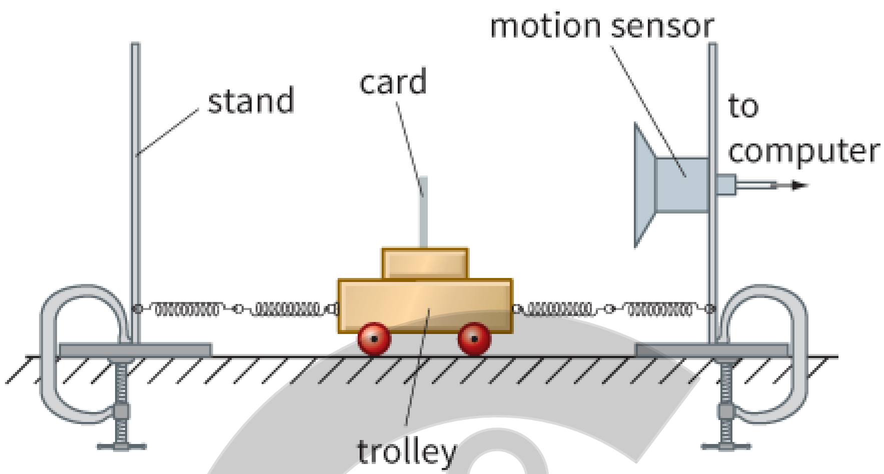


Figure 18.13: A motion sensor can be used to investigate s.h.m. of a spring–trolley system.

The motion sensor allows you to record how the displacement of the trolley varies with time. Ultrasonic pulses from the sensor are reflected by the card on the trolley and the reflected pulses are detected. This 'sonar' technique allows the sensor to determine the displacement of the trolley. A typical screen display is shown in Figure 18.14.

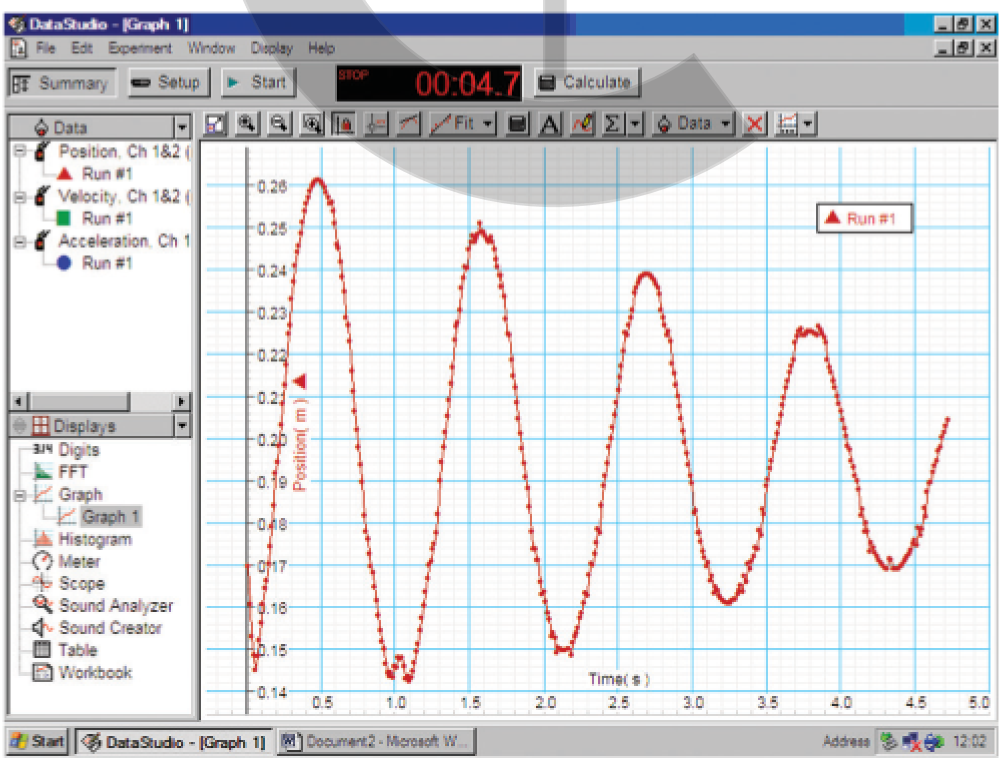


Figure 18.14: A typical displacement–time graph generated by a motion sensor.

The computer can then determine the velocity of the trolley by calculating the rate of change of displacement. Similarly, it can calculate the rate of change of velocity to determine the acceleration.

Idealised graphs of displacement, velocity and acceleration against time are shown in Figure 18.15. We will examine these graphs in sequence to see what they tell us about s.h.m. and how the three graphs are related to one another.

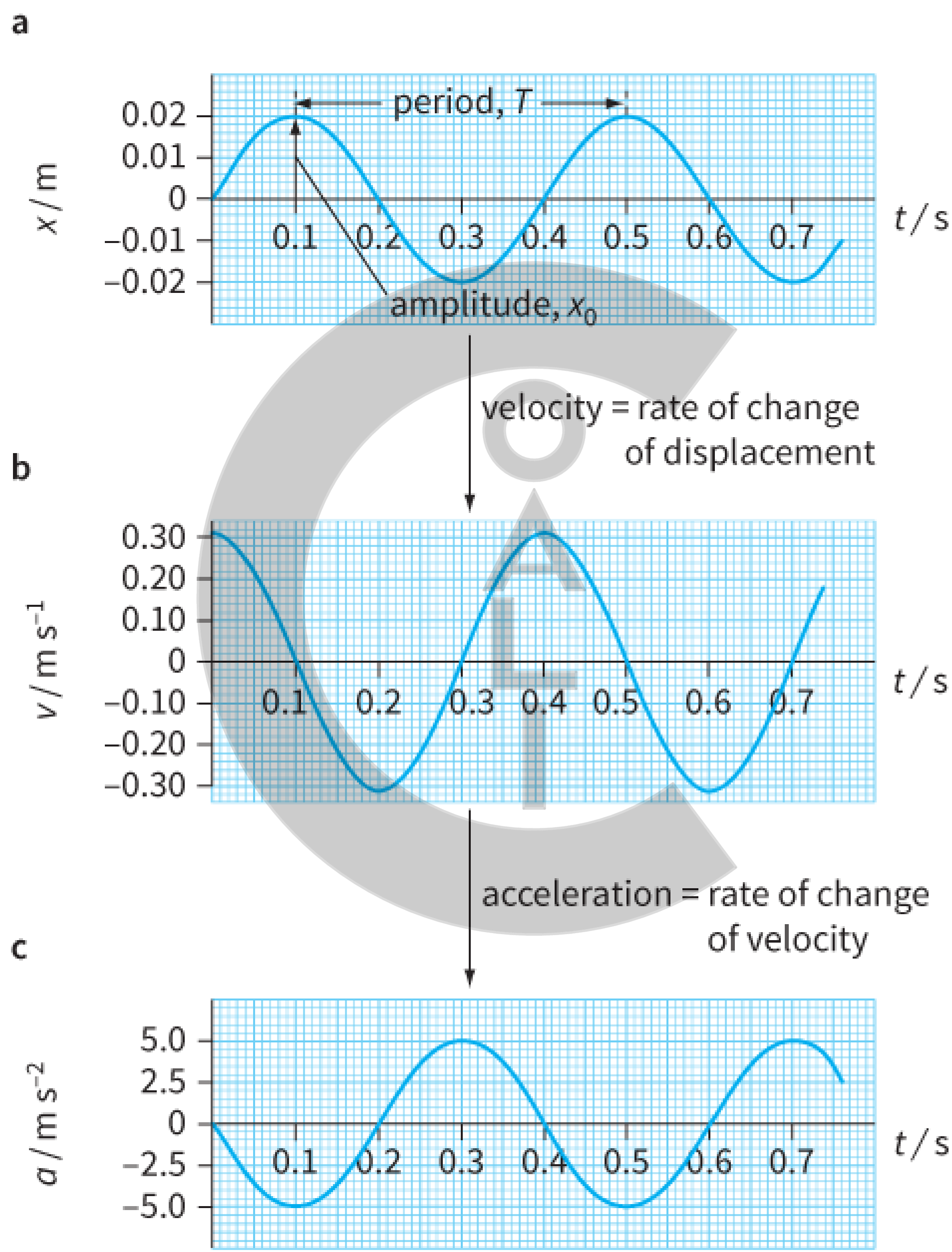


Figure 18.15: Graphs of displacement *x*, velocity *v* and acceleration *a* against time *t* for s.h.m.

Displacement–time (*x*–*t*) graph

The displacement of the oscillating mass varies according to the smooth curve shown in Figure 18.15a. Mathematically, this is a sine curve; its variation is described as sinusoidal. Note that this graph allows us to determine the amplitude x_0 and the period T of the oscillations. In this graph, the displacement x of the oscillation is shown as zero at the start, when t is zero. We have chosen to consider the motion to start when the mass is at the midpoint of its oscillation (equilibrium position) and is moving to the right. We could have chosen any other point in the cycle as the starting point, but it is conventional to start as shown here.

Velocity–time (*v*–*t*) graph

The velocity v of the oscillator at any time can be determined from the gradient of the displacement–time graph:

$$v = \frac{\Delta x}{\Delta t}\bigg|$$

Again, we have a smooth curve (Figure 18.15b), which shows how the velocity v depends on time t. The shape of the curve is the same as for the displacement–time graph, but it starts at a different point in the cycle.

When time t = 0, the mass is at the equilibrium position and this is where it is moving fastest. Hence, the velocity has its maximum value at this point. Its value is positive because at time t = 0 it is moving towards the right.

Acceleration–time (*a*–*t*) graph

Finally, the acceleration a of the oscillator at any time can be determined from the gradient of the velocity–time graph:

$$a = \frac{\Delta v}{\Delta t}\bigg|$$

This gives a third curve of the same general form (Figure 18.15c), which shows how the acceleration a depends on time t. At the start of the oscillation, the mass is at its equilibrium position. There is no resultant force acting on it so its acceleration is zero. As it moves to the right, the restoring force acts towards the left, giving it a negative acceleration. The acceleration has its greatest value when the mass is displaced farthest from the equilibrium position. Notice that the acceleration graph is 'upside-down' compared with the displacement graph. This shows that:

$$\text{acceleration} \propto -\text{displacement}$$

or

$$a \propto -x$$

In other words, whenever the mass has a positive displacement (to the right), its acceleration is to the left, and vice versa.

18.6 Frequency and angular frequency

The frequency *f* of s.h.m. is equal to the number of oscillations per unit time. As we saw earlier, *f* is related to the period *T* by:

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

We can think of a complete oscillation of an oscillator or a cycle of s.h.m. as being represented by 2π radians. (This is similar to a complete cycle of circular motion, where an object moves round through 2π radians.) The phase of the oscillation changes by 2π rad during one oscillation. Hence, if there are *f* oscillations in unit time, there must be $2\pi f$ radians in unit time. This quantity is the **angular frequency** of the s.h.m. and it is represented by the Greek letter ω (omega).

The angular frequency ω is related to frequency *f* by the equation:

$$\omega = 2\pi f$$

KEY EQUATION

Relationship of angular frequency ω to frequency *f*:

$$\omega = 2\pi f$$

Since $f = \frac{1}{T}$, the angular frequency ω is related to the period *T* of the oscillator by the equation:

$$\omega = \frac{2\pi}{T} \text{ or } T = \frac{2\pi}{\omega}$$

Questions

7 Use the graphs shown in Figure 18.15 to determine the values of the following quantities:

- **a** amplitude
- **b** time period
- **c** maximum velocity
- **d** maximum acceleration.

8 State at what point in an oscillation the oscillator has zero velocity but acceleration towards the right.

9 Look at the *x*–*t* graph of Figure 18.15a. When $t = 0.1$ s, what is the gradient of the graph? State the velocity at this instant.

10 Figure 18.16 shows the displacement–time (*x*–*t*) graph for an oscillating mass. Use the graph to determine the following quantities:

- **a** the velocity in cm s^{-1} when $t = 0$ s
- **b** the maximum velocity in cm s^{-1}
- **c** the acceleration in cm s^{-2} when $t = 1.0$ s.

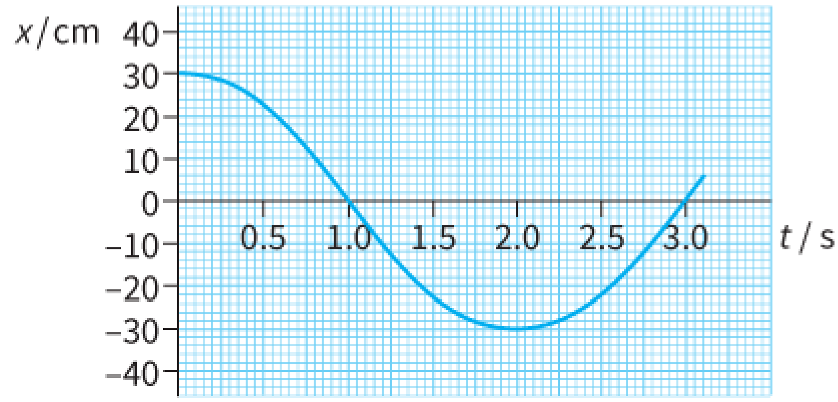


Figure 18.16: A displacement–time graph for an oscillating mass. For Question 10.

In Figure 18.17, a single cycle of s.h.m. is shown, but with the *x*-axis marked with the phase of the motion in radians.

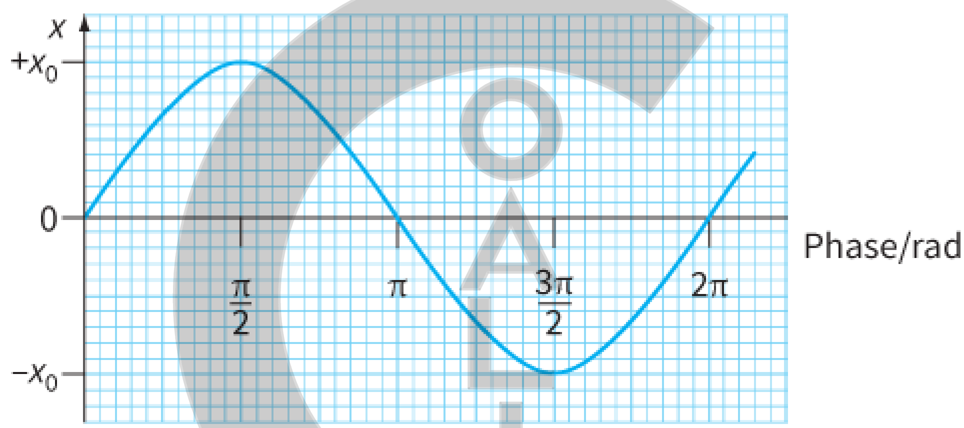


Figure 18.17: The phase of an oscillation varies from 0 to 2π during one cycle.

Questions

11 An object moving with s.h.m. goes through two complete cycles in 1.0 s. Calculate:

a the period T

b the frequency f

c the angular frequency ω.

12 Figure 18.18 shows the displacement–time graph for an oscillating mass. Use the graph to determine the following:

a amplitude

b period

c frequency

d angular frequency

e displacement at A

f velocity at B

g velocity at C.

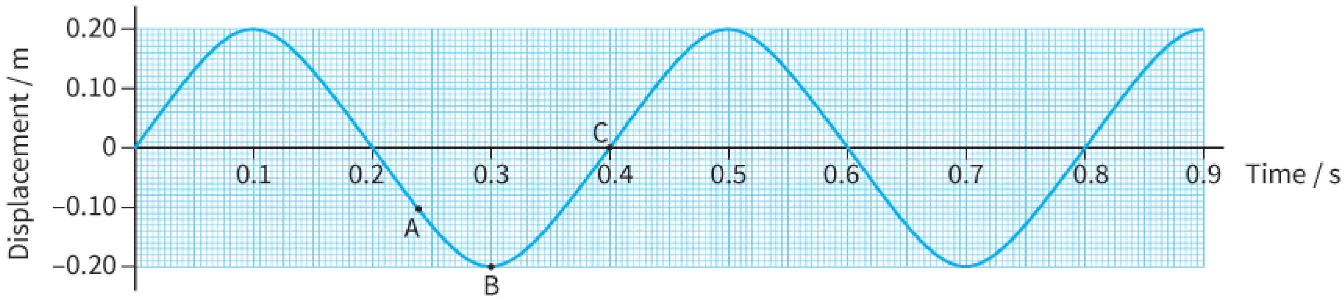


Figure 18.18: A displacement–time graph. For Question 12.

13 An atom in a crystal vibrates with s.h.m. with a frequency of 10^{14} Hz. The amplitude of its motion is 2.0×10^{-12} m.

a Sketch a graph to show how the displacement of the atom varies during one cycle.

b Use your graph to estimate the maximum velocity of the atom.

18.7 Equations of s.h.m.

The graph of Figure 18.15a, shown earlier, represents how the displacement of an oscillator varies during s.h.m. We have already mentioned that this is a sine curve. We can present the same information in the form of an equation. The relationship between the displacement *x* and the time *t* is as follows:

$$x = x_0 \sin \omega t$$

where x_0 is the amplitude of the motion and ω is its frequency. Sometimes, the same motion is represented using a cosine function, rather than a sine function:

$$x = x_0 \cos \omega t$$

KEY EQUATIONS

Equations of simple harmonic motion:

$$x = x_0 \sin \omega t$$
$$x = x_0 \cos \omega t$$

The difference between these two equations is illustrated in Figure 18.19. The sine version starts at $x = 0$; that is, the oscillating mass is at its equilibrium position when $t = 0$.

The cosine version starts at $x = x_0$, so that the mass is at its maximum displacement when $t = 0$.

Note that, in calculations using these equations, the quantity ωt is in radians. Make sure that your calculator is in radian mode for any calculation (see Worked example 2). The presence of the π in the equation should remind you of this.

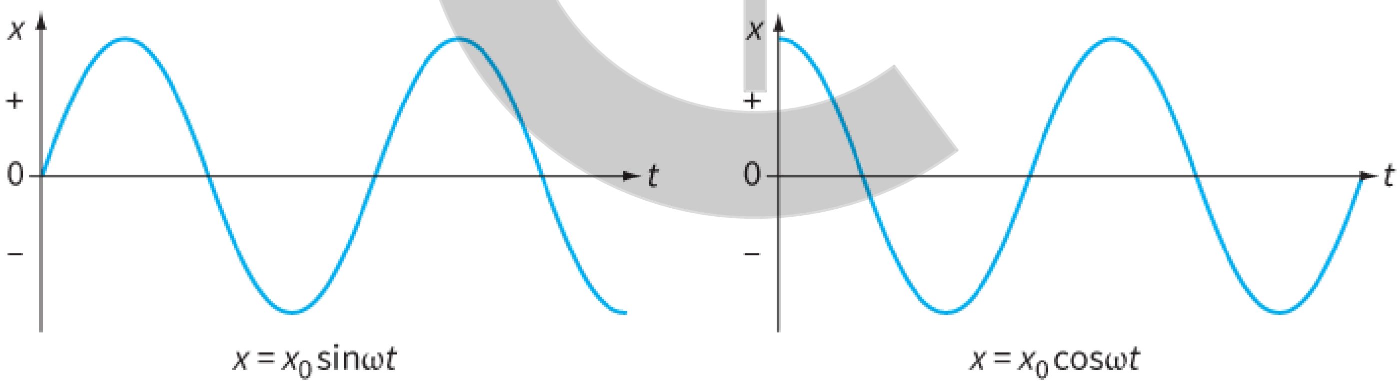


Figure 18.19: These two graphs represent the same simple harmonic motion. The difference in starting positions is related to the sine and cosine forms of the equation for *x* as a function of *t*.

Questions

14 The vibration of a component in a machine is represented by the equation:

$x = 3.0 \times 10^{-4} \sin(240\pi t)$

where the displacement *x* is in metres.

Determine the:

a amplitude

b frequency

c period

of the vibration.

15 A trolley is at rest, tethered between two springs. It is pulled 0.15 m to one side and, when time $t = 0$, it is released so that it oscillates back and forth with s.h.m. The period of its motion is 2.0 s.

a Write an equation for its displacement x at any time t (assume that the motion is not damped by frictional forces).

b Sketch a displacement–time graph to show two cycles of the motion, giving values where appropriate.

Acceleration and displacement

In s.h.m., an object's acceleration depends on how far it is displaced from its equilibrium position and on the magnitude of the restoring force. The greater the displacement *x*, the greater the acceleration *a*. In fact, *a* is proportional to *x*. We can write the following equation to represent this:

$$a = -\omega^2 x$$

where a = the acceleration of an object vibrating in s.h.m., ω is the angular frequency of the object, x = displacement

KEY EQUATION

$$a = -\omega^2 x$$

Acceleration of an object vibrating in simple harmonic motion.

This equation shows that a is proportional to x; the constant of proportionality is ω^2. The minus sign shows that, when the object is displaced to the **right**, the direction of its acceleration is to the **left**.

The acceleration is always directed towards the equilibrium position, in the opposite direction to the displacement.

It should not be surprising that angular frequency ω appears in this equation. Imagine a mass hanging on a spring, so that it can vibrate up and down. If the spring is stiff, the force on the mass will be greater; it will be accelerated more for a given displacement and its frequency of oscillation will be higher.

The equation $a = -\omega^2 x$ helps us to define simple harmonic motion. The acceleration a is directly proportional to displacement x; and the minus sign shows that it is in the opposite direction.

An object vibrates in simple harmonic motion if its acceleration is directly proportional to its displacement from its equilibrium position and is in the opposite direction to the displacement.

If a and x were in the same direction (no minus sign), the body's acceleration would increase as it moved away from the fixed point and it would move away faster and faster, never to return.

Figure 18.20 shows the acceleration–displacement ($a - x$) graph for an oscillator executing s.h.m. Note the following:

- The graph is a straight line through the origin ($a \propto x$).
- It has a negative slope (the minus sign in the equation $a = -\omega^2 x$). This means that the acceleration is always directed towards the equilibrium position.
- The magnitude of the gradient of the graph is ω^2.

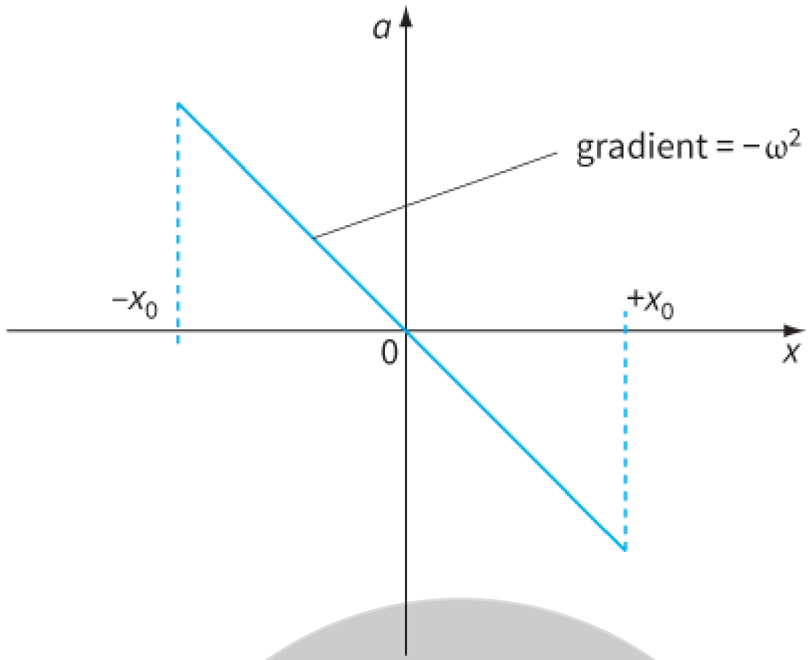


Figure 18.20: Graph of acceleration a against displacement x for an oscillator executing s.h.m.

- The gradient is independent of the amplitude of the motion. This means that the frequency f or the period T of the oscillator is independent of the amplitude and so a simple harmonic oscillator keeps steady time.

If you have studied calculus, you may be able to differentiate the equation for x twice with respect to time to obtain an equation for acceleration and thereby show that the defining equation $a = -\omega^2 x$ is satisfied.

KEY IDEA

We say that the equation $a = -\omega^2 x$ defines simple harmonic motion–it tells us what is required if a body is to perform s.h.m. The equation $x = x_0 \sin \omega t$ is then described as a **solution** to the equation, since it tells us how the displacement of the body varies with time.

WORKED EXAMPLE

2 A pendulum oscillates with frequency 1.5 Hz and amplitude 0.10 m. If it is passing through its equilibrium position when $t = 0$, write an equation to represent its displacement x in terms of amplitude x_0, angular frequency ω and time t. Determine its displacement when $t = 0.50$ s.

Step 1 Select the correct equation. In this case, the displacement is zero when $t = 0$, so we use the sine form:

$$x = x_0 \sin \omega t$$

Step 2 From the frequency f, calculate the angular frequency ω:

$$\omega = 2\pi f$$
$$= 2 \times \pi \times 1.5$$
$$= 3.0\pi$$

Step 3 Substitute values in the equation: $x_0 = 0.10$ m, so:

$x = 0.10 \sin (3.0\pi t)$

Hint: *Remember to put your calculator into radian mode.*

Step 4 To find x when t = 0.50 s, substitute for t and calculate the answer:

$$\begin{aligned} x &= 0.10 \sin(2\pi \times 1.5 \times 0.50) \\ &= 0.10 \sin(4.713) \\ &= -0.10 \text{ m} \end{aligned}$$

This means that the pendulum is at the extreme end of its oscillation; the minus sign means that it is at the negative or left-hand end, assuming you have chosen to consider displacements to the right as positive.

(If your calculation went like this:

$$\begin{aligned} x &= 0.10 \sin(2\pi \times 1.5 \times 0.50) \\ &= 0.10 \sin(4.713) \\ &= -8.2 \times 10^{-3} \text{ m} \end{aligned}$$

then your calculator was incorrectly set to work in degrees, not radians.)

Equations for velocity

The velocity v of an oscillator varies as it moves back and forth. It has its greatest speed when it passes through the equilibrium position in the middle of the oscillation. If we take time t = 0 when the oscillator passes through the middle of the oscillation with its greatest speed v_0, then we can represent the changing velocity as an equation:

$$x = x_0 \cos \omega t$$

We use the cosine function to represent the velocity since it has its maximum value when t = 0.

The equation $v = v_0 \cos \omega t$ tells us how v depends on t. We can write another equation to show how the velocity depends on the oscillator's displacement x:

$$v = \pm \omega \sqrt{x_0^2 - x^2}$$

This equation can be used to deduce the speed of an oscillator at any point in an oscillation, including its maximum speed.

Maximum speed of an oscillator

If an oscillator is executing simple harmonic motion, it has maximum speed when it passes through its equilibrium position. This is when its displacement x is zero. The maximum speed v_0 of the oscillator depends on the frequency f of the motion and on the amplitude x_0. Substituting x = 0 in the equation:

$$v = \omega \sqrt{x_0^2 - x^2}$$

x = 0 when the speed is at a maximum:

$$v_0 = \omega x_0$$

According to this equation, for a given oscillation:

$$v_0 \propto x_0$$

KEY EQUATION

$$v = \omega\sqrt{x_0^2 - x^2}$$

Speed of an oscillator.

A simple harmonic oscillator has a period that is independent of the amplitude. A greater amplitude means that the oscillator has to travel a greater distance in the same time–hence it has a greater speed.

The equation also shows that the maximum speed is proportional to the frequency. Increasing the frequency means a shorter period. A given distance is covered in a shorter time–hence it has a greater speed.

Have another look at Figure 18.15. The period of the motion is 0.40 s and the amplitude of the motion is 0.02 m. The frequency *f* can be calculated as follows:

$$\begin{aligned} f &= \frac{1}{t} \\ &= \frac{1}{0.40} \\ &= 2.5\text{Hz} \end{aligned}$$

We can now use the equation $v_0 = (2\pi f)x_0$ to determine the maximum speed v_0:

$$v_0 = (2\pi f)x_0 = (2\pi \times 2.5) \times 2.0 \times 10^{-2}$$

$$v_0 \approx 0.31 \text{ m s}^{-1}$$

This is how the values on Figure 18.15b were calculated.

Questions

16 A mass secured at the end of a spring moves with s.h.m. The frequency of its motion is 1.4 Hz.

- **a** Write an equation of the form $a = -\omega^2 x$ to show how the acceleration of the mass depends on its displacement.
- **b** Calculate the acceleration of the mass when it is displaced 0.050 m from its equilibrium position.

17 A short pendulum oscillates with s.h.m. such that its acceleration *a* (in m s^{-2}) is related to its displacement *x* (in m) by the equation $a = -300x$. Determine the frequency of the oscillations.

18 The pendulum of a grandfather clock swings from one side to the other in 1.00 s. The amplitude of the oscillation is 12 cm.

- **a** Calculate:
  - **i** the period of its motion
  - **ii** the frequency
  - **iii** the angular frequency.
- **b** Write an equation of the form $a = -\omega^2 x$ to show how the acceleration of the pendulum bob depends on its displacement.
- **c** Calculate the maximum speed of the pendulum bob.
- **d** Calculate the speed of the bob when its displacement is 6 cm.

19 A trolley of mass *m* is fixed to the end of a spring. The spring can be compressed and extended. The spring has a force constant *k*. The other end of the spring is attached to a vertical wall. The trolley lies on a smooth horizontal table. The trolley oscillates when it is displaced from its equilibrium position.

- **a** Show that the motion of the oscillating trolley is s.h.m.
- **b** Show that the period *T* of the trolley is given by the equation:

$$T = 2\pi\sqrt{\frac{m}{k}}$$

18.8 Energy changes in s.h.m.

During simple harmonic motion, there is a constant exchange of energy between two forms: potential and kinetic. We can see this by considering the mass–spring system shown in Figure 18.21.

When the mass is pulled to one side (to start the oscillations), one spring is compressed and the other is stretched. The springs store elastic potential energy. When the mass is released, it moves back towards the equilibrium position, accelerating as it goes. It has increasing kinetic energy. The potential energy stored in the springs decreases while the kinetic energy of the mass increases by the same amount (as long as there are no heat losses due to frictional forces). Once the mass has passed the equilibrium position, its kinetic energy decreases and the energy is transferred back to the springs. Provided the oscillations are undamped, the total energy in the system remains constant.

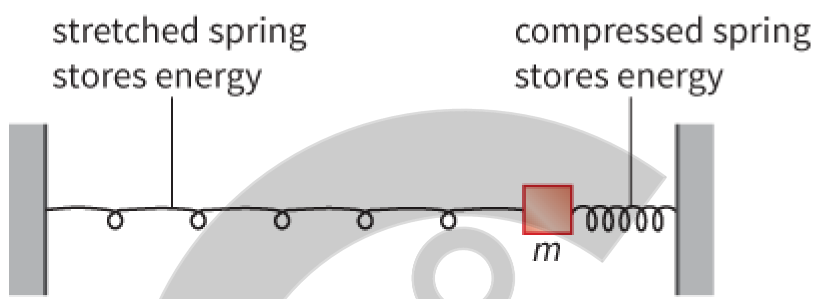


Figure 18.21: The elastic potential energy stored in the springs is converted to kinetic energy when the mass is released.

Energy graphs

We can represent these energy changes in two ways. Figure 18.22 shows how the kinetic energy and elastic potential energy change with time. Potential energy is maximum when displacement is maximum (positive or negative). Kinetic energy is maximum when displacement is zero. The total energy remains constant throughout. Note that both kinetic energy and potential energy go through two complete cycles during one period of the oscillation. This is because kinetic energy is maximum when the mass is passing through the equilibrium position moving to the left and again moving to the right. The potential energy is maximum at both ends of the oscillation.

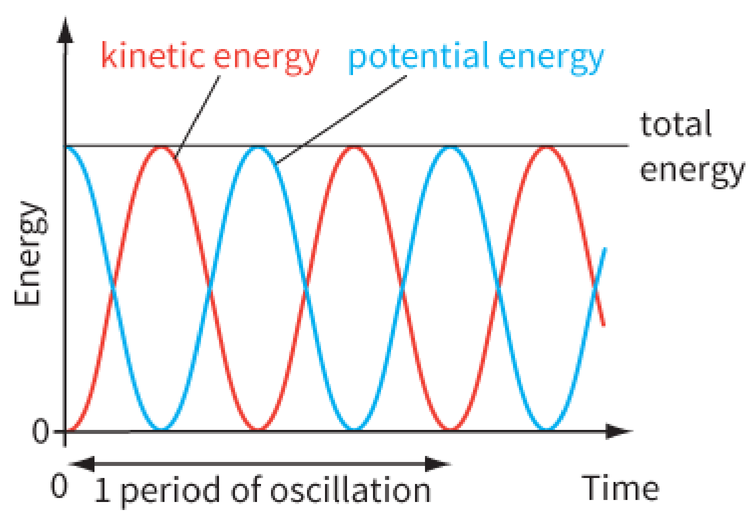


Figure 18.22: The kinetic energy and potential energy of an oscillator vary periodically, but the total energy remains constant if the system is undamped.

A second way to show this is to draw a graph of how potential energy and kinetic energy vary with displacement (Figure 18.23).

The graph shows that:

- kinetic energy is maximum when displacement $x = 0$
- potential energy is maximum when $x = \pm x_0$
- at any point on this graph, the total energy (k.e. + p.e.) has the same value.

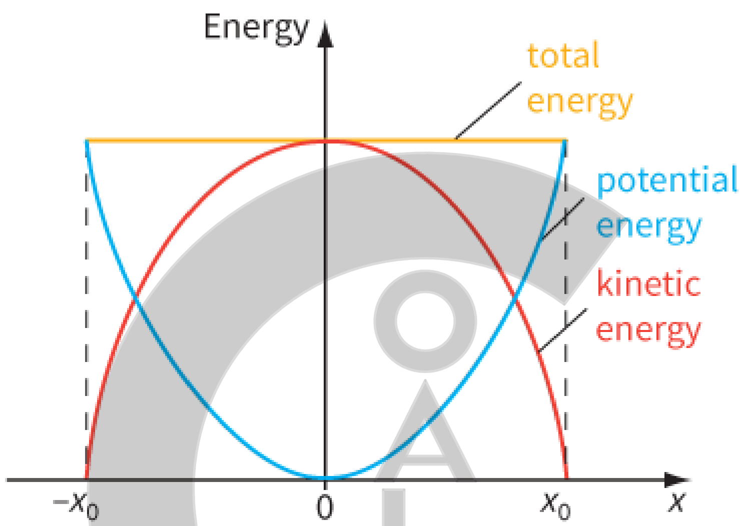


Figure 18.23: The kinetic energy is maximum at zero displacement; the potential energy is maximum at maximum displacement (x_0 and $-x_0$).

It follows that if the maximum speed is v_0 then maximum kinetic energy = $\frac{1}{2}mv_0^2$

At this point in the cycle, all the energy is in the form of kinetic energy, so the total energy of the system is:

$$E_0 = \frac{1}{2}mv_0^2$$

Since:

$$v_0 = \omega x_0$$

Then:

$$E_0 = \frac{1}{2}m\omega^2 x_0^2$$

KEY EQUATION

$$E_0 = \frac{1}{2}m\omega^2 x_0^2$$

Total energy of a system undergoing simple harmonic motion.

Questions

20 To start a pendulum swinging, you pull it slightly to one side.

- **a** What kind of energy does this transfer to the mass?
- **b** Describe the energy changes that occur when the mass is released.

21 Figure 18.23 shows how the different forms of energy change with displacement during s.h.m. Copy the graph, and show how the graph would differ if the oscillating mass were given only half the initial input of energy.

22 Figure 18.24 shows how the velocity v of a 2.0 kg mass was found to vary with time t during an investigation of the s.h.m. of a pendulum. Use the graph to estimate the following for the mass:

- **a** its maximum velocity
- **b** its maximum kinetic energy
- **c** its maximum potential energy
- **d** its maximum acceleration
- **e** the maximum restoring force that acted on it.

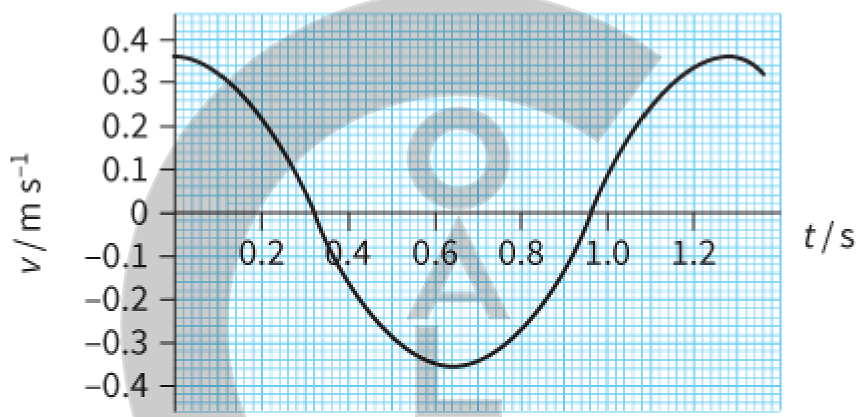


Figure 18.24: A velocity–time graph for a pendulum. For Question 22.

18.9 Damped oscillations

In principle, oscillations can go on for ever. In practice, however, the oscillations we observe around us do not. They die out, either rapidly or gradually. A child on a swing knows that the amplitude of her swinging will decline until eventually she will come to rest, unless she can put some more energy into the swinging to keep it going.

This happens because of friction. On a swing, there is friction where the swing is attached to the frame and there is friction with the air. The amplitude of the child's oscillations decreases as the friction transfers energy away from her to the surroundings.

We describe these oscillations as **damped**. Their amplitude decreases according to a particular pattern. This is shown in Figure 18.25.

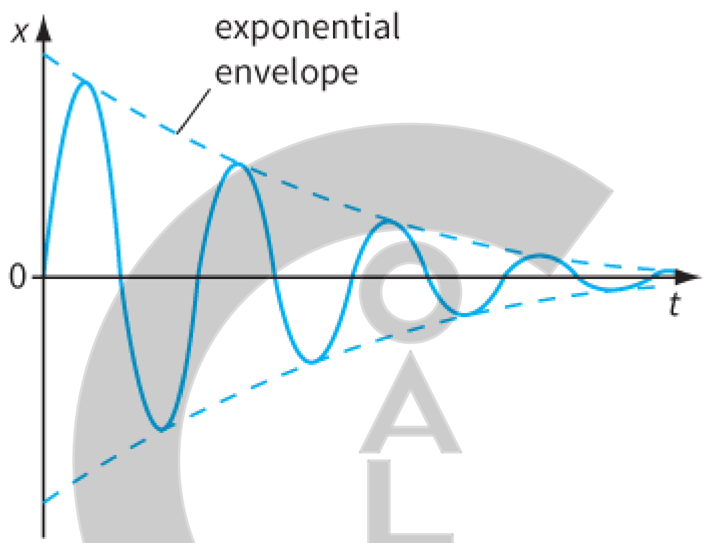


Figure 18.25: Damped oscillations.

The amplitude of damped oscillations does not decrease linearly. It decays exponentially with time. An exponential decay is a particular mathematical pattern that arises as follows. At first, the swing moves rapidly. There is a lot of air resistance to overcome, so the swing loses energy quickly and its amplitude decreases at a high rate. Later, it is moving more slowly. There is less air resistance and so energy is lost more slowly–the amplitude decreases at a lower rate. Hence, we get the characteristic curved shape, which is the 'envelope' of the graph in Figure 18.25.

Notice that the frequency of the oscillations does not change as the amplitude decreases. This is a characteristic of simple harmonic motion. The child may, for example, swing back and forth once every two seconds, and this stays the same whether the amplitude is large or small.

PRACTICAL ACTIVITY 18.2

Investigating damping

You can investigate the exponential decrease in the amplitude of oscillations using a simple laboratory arrangement (Figure 18.26). A hacksaw blade or other springy metal strip is clamped (vertically or horizontally) to the bench. A mass is attached to the free end. This will oscillate freely if you displace it to one side.

A card is attached to the mass so that there is significant air resistance as the mass oscillates. The amplitude of the oscillations decreases and can be measured every five oscillations by judging the position of the blade against a ruler fixed alongside.

A graph of amplitude against time will show the characteristic exponential decrease. You can find the 'half-life' of this exponential decay graph by determining the time it takes to decrease to half its initial amplitude (Figure 18.27).

By changing the size of the card, it is possible to change the degree of damping, and hence alter the half-life of the motion.

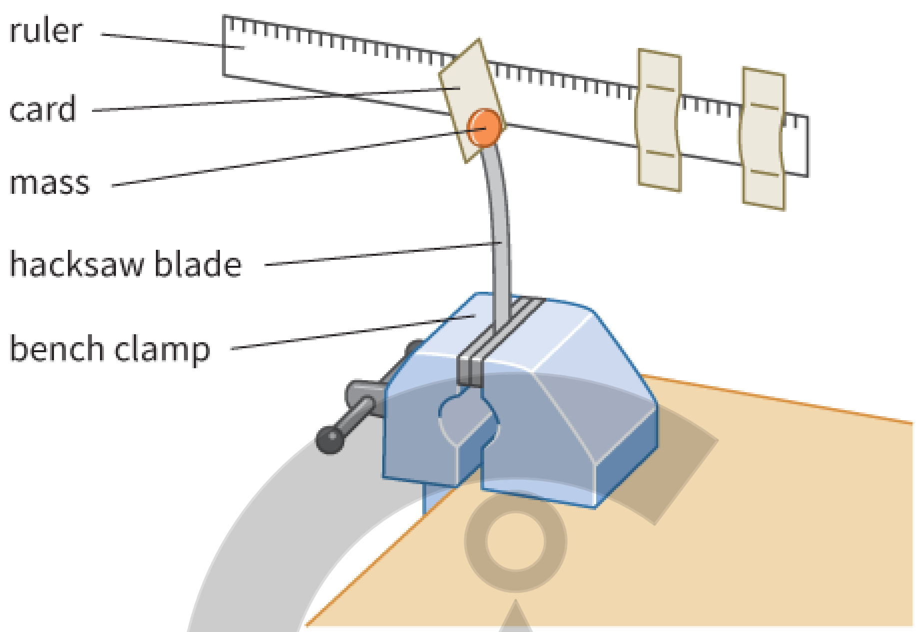


Figure 18.26: Damped oscillations with a hacksaw blade.

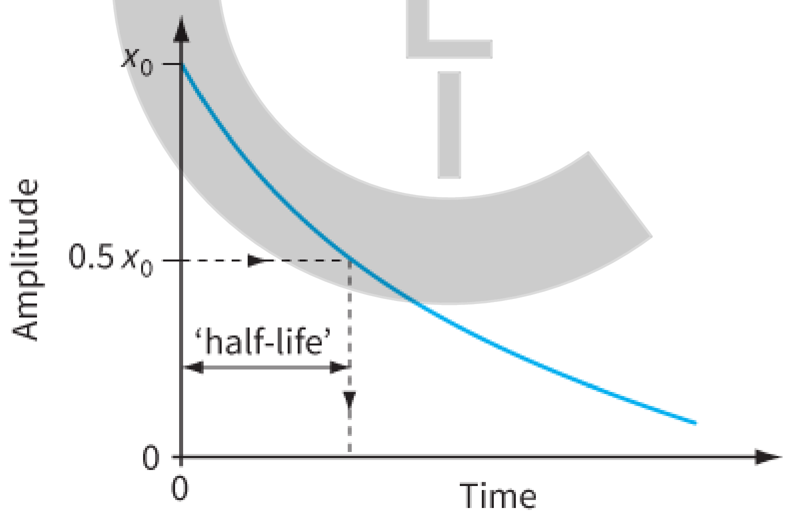


Figure 18.27: A typical graph of amplitude against time for damped oscillations.

Energy and damping

Damping can be very useful if we want to get rid of vibrations. For example, a car has springs (Figure 18.28) that make the ride much more comfortable for us when the car goes over a bump. However, we wouldn't want to spend every car journey vibrating up and down as a reminder of the last bump we went over. So the springs are damped by the shock absorbers, and we return rapidly to a smooth ride after every bump.

Damping is achieved by introducing the force of friction into a mechanical system. In an undamped oscillation, the total energy of the oscillation remains constant. There is a regular interchange between potential and kinetic energy. By introducing friction, damping has the effect of removing energy from the oscillating system, and the amplitude and maximum speed of the oscillation decrease.

Figure 18.28: The springs and shock absorbers in a car suspension system form a damped system.

Question

23 **a** Sketch graphs to show how each of the following quantities changes during the course of a single complete oscillation of an undamped pendulum: kinetic energy, potential energy, total energy.

b State how your graphs would be different for a lightly damped pendulum.

18.10 Resonance

Resonance is an important physical phenomenon that can appear in a great many different situations. A dramatic example is the Millennium Bridge in London, opened in June 2000 (Figure 18.29). With up to 2000 pedestrians walking on the bridge, it started to sway dangerously. The people also swayed in time with the bridge, and this caused the amplitude of the bridge's oscillations to increase–this is resonance. After three days, the bridge was closed. It took engineers two years to analyse the problem and then add 'dampers' to the bridge to absorb the energy of its oscillations. The bridge was then reopened and there have been no problems since.

You will have observed a much more familiar example of resonance when pushing a small child on a swing. The swing plus child has a natural frequency of oscillation. A small push in each cycle results in the amplitude increasing until the child is swinging high in the air.

Figure 18.29: The 'wobbly' Millennium Bridge in London was closed for nearly two years to correct problems caused by resonance.

PRACTICAL ACTIVITY 18.3

Observing resonance

Resonance can be observed with almost any oscillating system. The system is forced to oscillate at a particular frequency. If the forcing frequency happens to match the natural frequency of oscillation of the system, the amplitude of the resulting oscillations can build up to become very large.

Barton's pendulums

Barton's pendulums is a demonstration of this (Figure 18.30). Several pendulums of different lengths hang from a horizontal string. Each has its own natural frequency of oscillation. The 'driver' pendulum at the end is different; it has a large mass at the end, and its length is equal to that of one of the others. When the driver is set swinging, the others gradually start to move. However, only the pendulum whose length matches that of the driver pendulum builds up a large amplitude so that it is resonating.

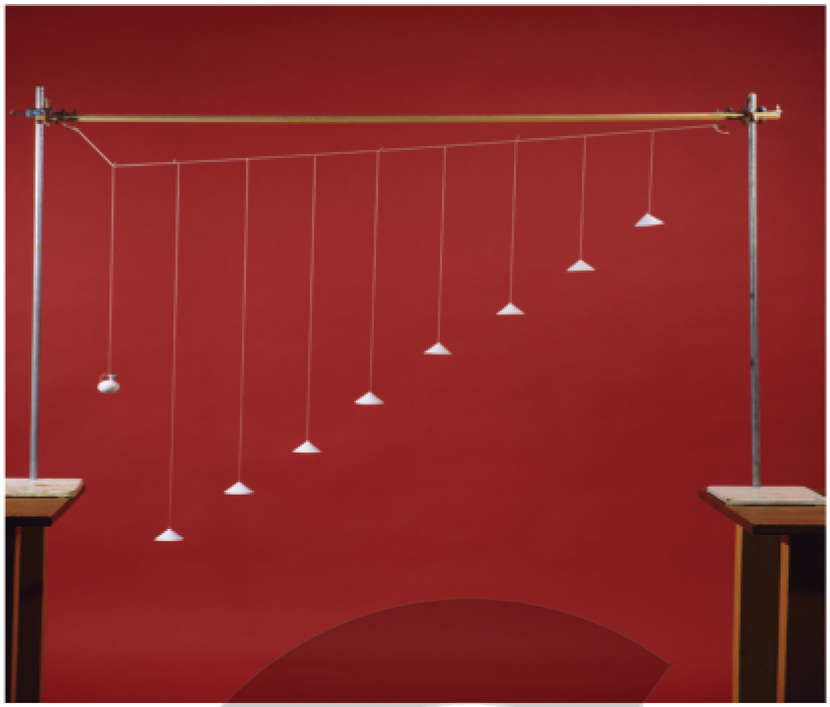


Figure 18.30: Barton's pendulums.

What is going on here? All the pendulums are coupled together by the suspension. As the driver swings, it moves the suspension, which in turn moves the other pendulums. The frequency of the matching pendulum is the same as that of the driver, and so it gains energy and its amplitude gradually builds up. The other pendulums have different natural frequencies, so the driver has little effect.

In a similar way, if you were to push the child on the swing once every three-quarters of an oscillation, you would soon find that the swing was moving backwards as you tried to push it forwards, so that your push would slow it down.

A mass–spring system

You can observe resonance for yourself with a simple mass–spring system. You need a mass on the end of a spring (Figure 18.31), chosen so that the mass oscillates up and down with a natural frequency of about 1 Hz. Now hold the top end of the spring and move your hand up and down rapidly, with an amplitude of a centimetre or two. Very little happens. Now move your hand up and down more slowly, close to 1 Hz.

You should see the mass oscillating with gradually increasing amplitude. Adjust your movements to the exact frequency of the natural vibrations of the mass and you will see the greatest effect.

Figure 18.31: Resonance with a mass on a spring.

Defining resonance

For resonance to occur, we must have a system that is capable of oscillating freely. We must also have some way in which the system is forced to oscillate. When the forcing frequency matches the natural frequency of the system, the amplitude of the oscillations grows dramatically.

If the driving frequency does not quite match the natural frequency, the amplitude of the oscillations will increase, but not to the same extent as when resonance is achieved. Figure 18.32 shows how the amplitude of oscillations depends on the driving frequency in the region close to resonance.

In resonance, energy is transferred from the driver to the resonating system more efficiently than when resonance does not occur. For example, in the case of the Millennium Bridge, energy was transferred from the pedestrians to the bridge, causing large-amplitude oscillations.

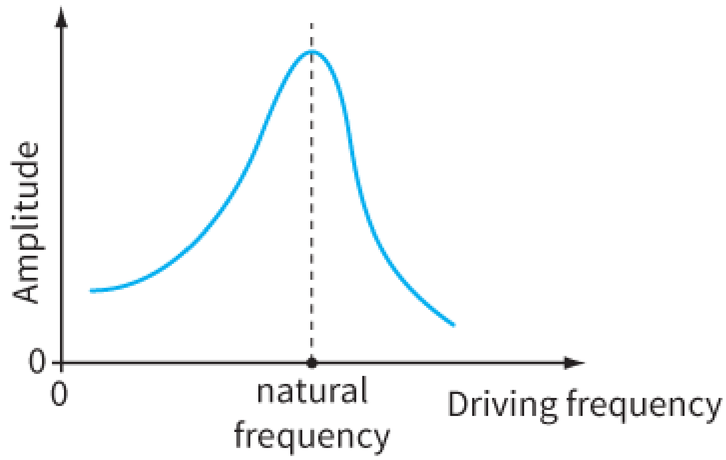


Figure 18.32: Maximum amplitude is achieved when the driving frequency matches the natural frequency of oscillation.

The following statements apply to any system in resonance:

- Its natural frequency is equal to the frequency of the driver.
- Its amplitude is maximum.
- It absorbs the greatest possible energy from the driver.

Resonance and damping

During earthquakes, buildings are forced to oscillate by the vibrations of the Earth. Resonance can occur, resulting in serious damage (Figure 18.33). In regions of the world where earthquakes happen regularly, buildings may be built on foundations that absorb the energy of the shock waves. In this way, the vibrations are ‘damped’ so that the amplitude of the oscillations cannot reach dangerous levels. This is an expensive business, and so far is restricted to the wealthier parts of the world.

Figure 18.33: Resonance during the Christchurch, New Zealand, earthquake of 22 February 2011 caused the collapse of many buildings. The earthquake, whose epicentre was in Lyttelton, just 10 kilometres south-east of Christchurch's central business district, measured 6.3. Nearly 200 lives were lost.

Damping is useful if we want to reduce the damaging effects of resonance. Figure 18.34 shows how damping alters the resonance response curve of Figure 18.32. Notice that, as the degree of damping is increased, the amplitude of the resonant vibrations decreases. The resonance peak becomes broader. There is also an effect on the frequency at which resonance occurs, which becomes lower as the damping increases.

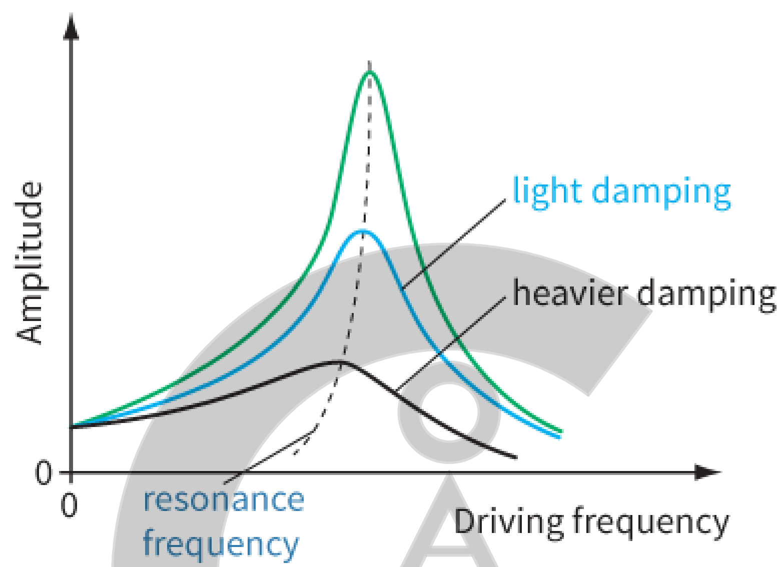


Figure 18.34: Damping reduces the amplitude of resonant vibrations.

An everyday example of damping can be seen on some doors. For example, a restaurant may have a door leading to the kitchen; this door can swing open in either direction. Such a door is designed to close by itself after someone has passed through it. Ideally, the door should swing back quickly without overshooting its closed position. To achieve this, the door hinges (or the closing mechanism) must be correctly damped. If the hinges are damped too lightly, the door will swing back and forth several times as it closes. If the damping is too heavy, it will take too long to close. With **critical damping**, the door will swing closed quickly without oscillating.

Critical damping is the minimum amount of damping required to return an oscillator to its equilibrium position without oscillating. Under-damping results in unwanted oscillations; over-damping results in a slower return to equilibrium (see Figure 18.35). A car's suspension system uses springs to smooth out bumps in the road. It is usually critically damped so that passengers do not experience nasty vibrations every time the car goes over a bump.

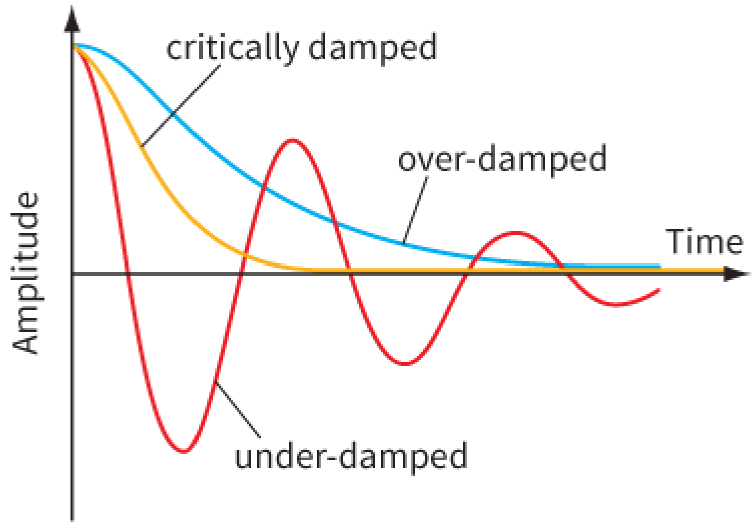


Figure 18.35: Critical damping is just enough to ensure that a damped system returns to equilibrium without oscillating.

Using resonance

As we have seen, resonance can be a problem in mechanical systems. However, it can also be useful. For example, many musical instruments rely on resonance.

Resonance is not confined to mechanical systems. It is made use of in, for example, microwave cooking. The microwaves used have a frequency that matches the natural frequency of vibration of water molecules (the microwave is the 'driver' and the molecule is the 'resonating system'). The water molecules in the food are forced to vibrate and they absorb the energy of the microwave radiation. The water gets hotter and the absorbed energy spreads through the food and cooks or heats it.

Magnetic resonance imaging (MRI) is used in medicine to produce images such as Figure 18.36, showing aspects of a patient's internal organs. Radio waves having a range of frequencies are used, and particular frequencies are absorbed by particular atomic nuclei. The frequency absorbed depends on the type of nucleus and on its surroundings. By analysing the absorption of the radio waves, a computer-generated image can be produced.

A radio or television also depends on resonance for its tuning circuitry. The aerial picks up signals of many different frequencies from many transmitters. The tuner can be adjusted to resonate at the frequency of the transmitting station you are interested in, and the circuit produces a large-amplitude signal for this frequency only.

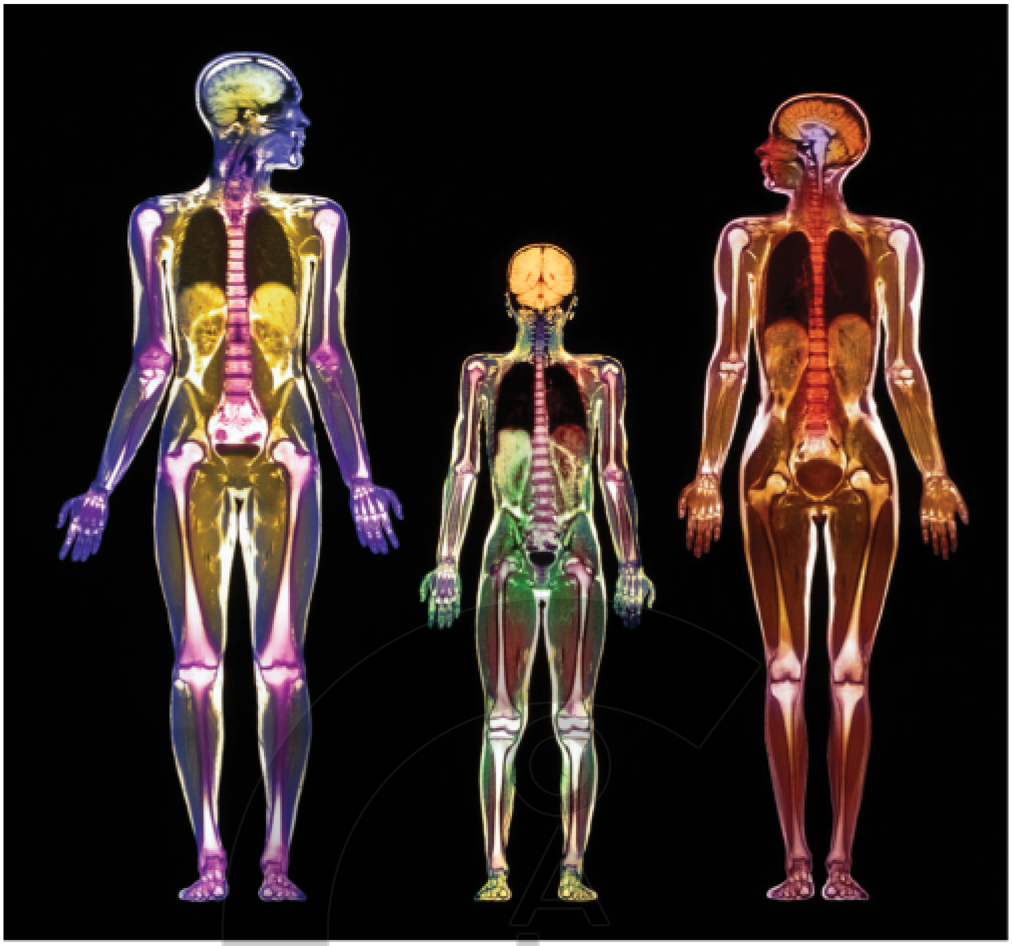


Figure 18.36: This magnetic resonance imaging (MRI) picture shows a man, a woman and a nine-year-old child. The image has been coloured to show up the bones (white), lungs (dark) and other organs.

Big ideas in physics

This study of simple harmonic motion illustrates some important aspects of physics:

- Physicists often take a complex problem (such as how the atoms in a solid vibrate) and reduce it to a simpler, more manageable problem (such as how a mass–spring system vibrates). This is simpler because we know that the spring obeys Hooke's law, so that force is proportional to displacement.
- Physicists generally feel happier if they can write mathematical equations that will give numerical answers to problems. The equation $a = -\omega^2 x$, which describes s.h.m., can be solved to give the sine and cosine equations we have considered earlier.
- Once physicists have solved one problem like this, they look around for other situations where they can use the same ideas all over again. So the mass–spring theory also works well for vibrating atoms and molecules, for objects bobbing up and down in water, and in many other situations.
- Physicists also seek to modify the theory to fit a greater range of situations. For example, what happens if the vibrating mass experiences a frictional force as it oscillates? (This is damping, as discussed earlier.) What happens if the spring doesn't obey Hooke's law? (This is a harder question to answer.)

Your A Level physics course will help you to build up your appreciation of some of these big ideas–fields (magnetic, electric, gravitational), energy and so on.

Question

24 Give an example of a situation where resonance is a problem, and a second example where resonance is useful. In each example, state what the oscillating system is and what forces it to resonate.

REFLECTION

You might have observed some of the terms and the mathematical equations in this chapter share many characteristics with those used in circular motion.

Make a list of the similarities and differences between the terms used in the two examples.

Make a list of equations used in the two examples. How are they related?

Can you use these similarities to help you understand simple harmonic motion further?

What things might you need help with to understand the chapter even better?

SUMMARY

Many systems, mechanical and otherwise, will oscillate freely when disturbed from their equilibrium position.

Some oscillators have motion described as **simple harmonic motion** (s.h.m.). For these systems, graphs of displacement, velocity and acceleration against time are sinusoidal curves–see Figure 18.37.

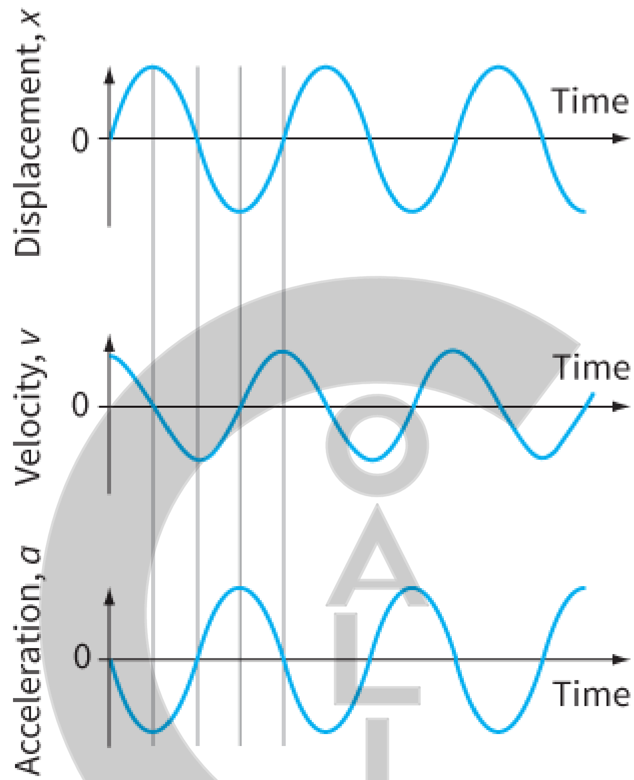


Figure 18.37: Graphs for s.h.m.

During a single cycle of s.h.m., the phase changes by 2π radians. The angular frequency ω of the motion is related to its period T and frequency f:

$$f = \frac{2\pi}{T} \text{ and } \omega = 2\pi f$$

In s.h.m., displacement x and velocity v and acceleration can be represented as functions of time t by equations of the form:

$$x = x_0 \sin \omega t \quad \text{and} \quad v = v_0 \cos \omega t \text{ and } a = -a_0 \sin \omega t$$

A body executes simple harmonic motion if its acceleration is directly proportional to its displacement from its equilibrium position, and is always directed towards the equilibrium position.

Acceleration a in s.h.m. is related to displacement x by the equation:

$$a = -\omega^2 x$$

The maximum speed v_0 in s.h.m. is given by the equation: $v_0 = \omega x_0$
The frequency or period of a simple harmonic oscillator is independent of its amplitude.
In s.h.m., there is a regular interchange between kinetic energy and potential energy.
Resistive forces remove energy from an oscillating system. This is known as damping. Damping causes the amplitude to decay with time.
Critical damping is the minimum amount of damping required to return an oscillator to its equilibrium position without oscillating.
When an oscillating system is forced to vibrate close to its natural frequency, the amplitude of vibration increases rapidly. The amplitude is maximum when the forcing frequency matches the natural frequency of the system; this is resonance.
Resonance can be a problem, but it can also be very useful.

EXAM-STYLE QUESTIONS

1 A mass, hung from a spring, oscillates with simple harmonic motion.

Which statement is correct? **[1]**

A The force on the mass is directly proportional to the angular frequency of the oscillation.

B The force on the mass is greatest when the displacement of the bob is greatest.

C The force on the mass is greatest when the speed of the bob is greatest.

D The force on the mass is inversely proportional to the time period of the oscillation.

2 The bob of a simple pendulum has a mass of 0.40 kg. The pendulum oscillates with a period of 2.0 s and an amplitude of 0.15 m.

At one point in its cycle it has a potential energy of 0.020 J.

What is the kinetic energy of the pendulum bob at this point? **[1]**

A 0.024 J

B 0.044 J

C 0.14 J

D 0.18 J

3 State and justify whether the following oscillators show simple harmonic motion:

a a basketball being bounced repeatedly on the ground. **[2]**

b a guitar string vibrating **[2]**

c a conducting sphere vibrating between two parallel, oppositely charged metal plates **[1]**

d the pendulum of a grandfather clock. **[2]**

[Total: 7]

4 The pendulum of a clock is displaced by a distance of 4.0 cm and it oscillates in s.h.m. with a period of 1.0 s.

a Write down an equation to describe the displacement x of the pendulum bob with time t. **[2]**

b Calculate:

i the maximum velocity of the pendulum bob **[2]**

ii its velocity when its displacement is 2.0 cm. **[1]**

[Total: 5]

5 A 50 g mass is attached to a securely clamped spring. The mass is pulled downwards by 16 mm and released, which causes it to oscillate with s.h.m. of time period of 0.84 s.

a Calculate the frequency of the oscillation. **[1]**

b Calculate the maximum velocity of the mass. **[1]**

c Calculate the maximum kinetic energy of the mass and state at which point in the oscillation it will have this velocity. **[2]**

d Write down the maximum gravitational potential energy of the mass (relative to its equilibrium position). You may assume that the damping is negligible. **[1]**

[Total: 5]

6 In each of the three graphs, **a**, **b** and **c** in Figure 18.38, give the phase difference between the two curves:

i as a fraction of an oscillation **[1]**

ii in degrees **[1]**

iii in radians. **[1]**

[Total: 3]

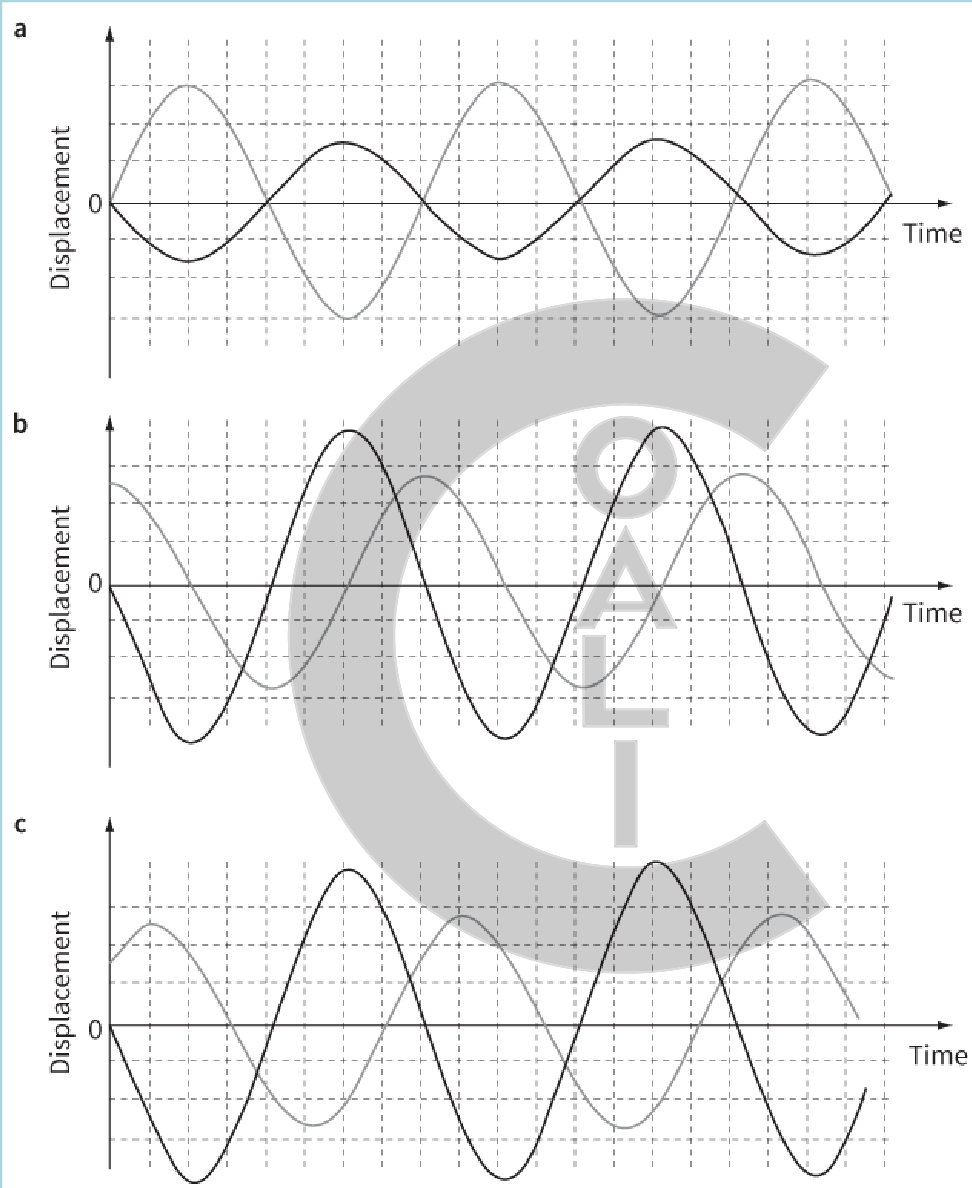


Figure 18.38

7 **a** Determine the frequency and the period of the oscillation described by this graph. **[2]**

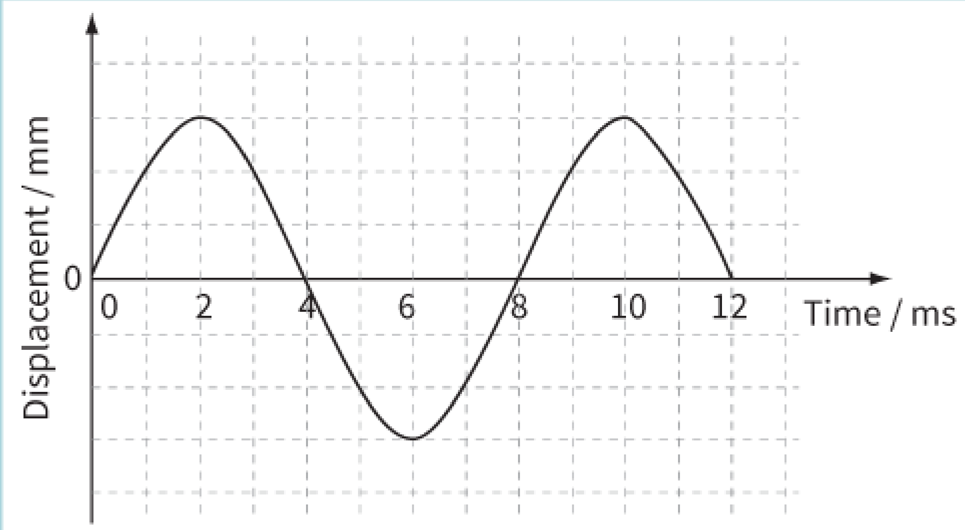


Figure 18.39

b Use a copy of the graph and on the same axes sketch:

i the velocity of the particle **[1]**

ii the acceleration of the particle. **[2]**

[Total: 5]

8 These graphs show the displacement of a body as it vibrates between two points.

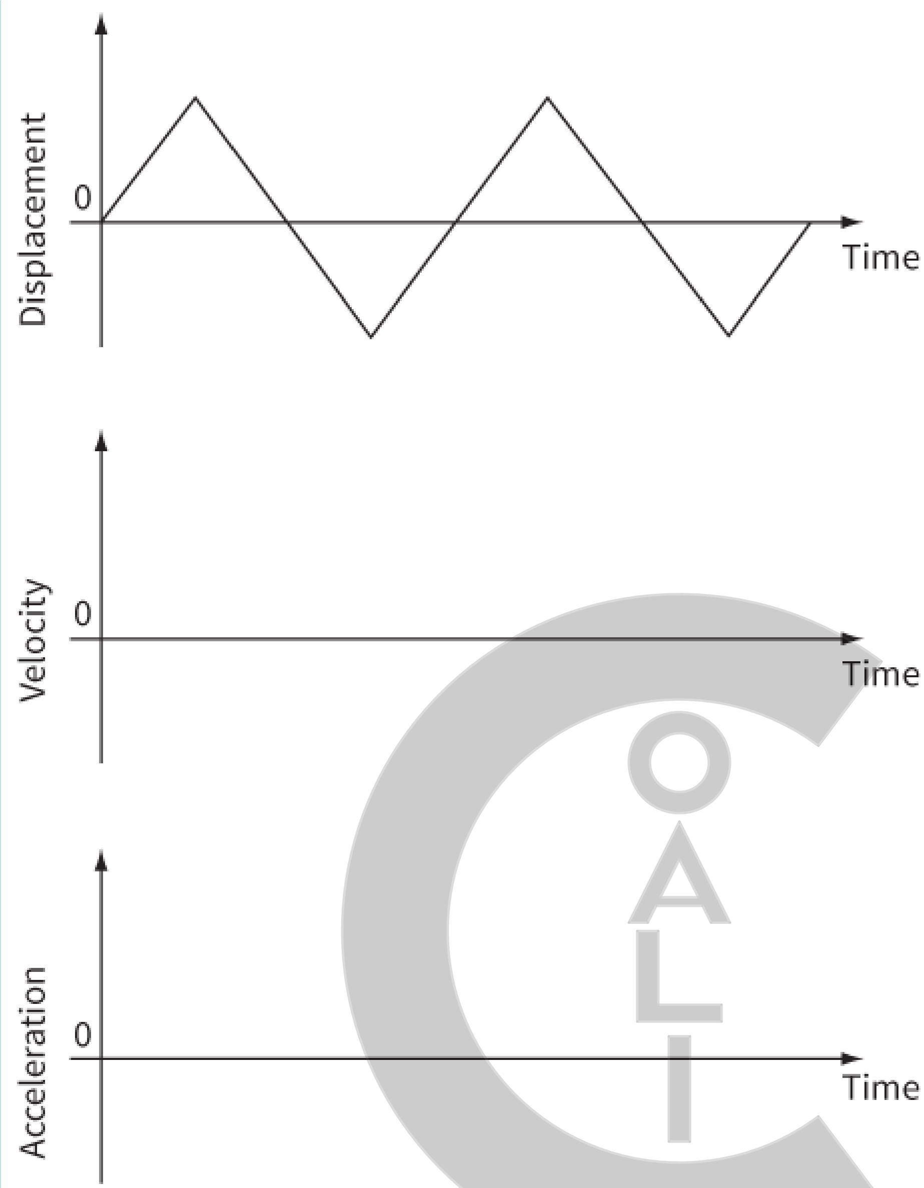


Figure 18.40

a State and explain whether the body is moving with simple harmonic motion. **[1]**

b Make a copy of the three graphs.

i On the second set of axes on your copy show the velocity of the body as it vibrates. **[1]**

ii On the third set of axes on your copy, show the acceleration of the body. **[2]**

[Total: 4]

9 This diagram shows the piston of a small car engine that oscillates in the cylinder with a motion that approximates simple harmonic motion at 4200 revs per minute (1 rev = 1 cycle). The mass of the piston is 0.24 kg.

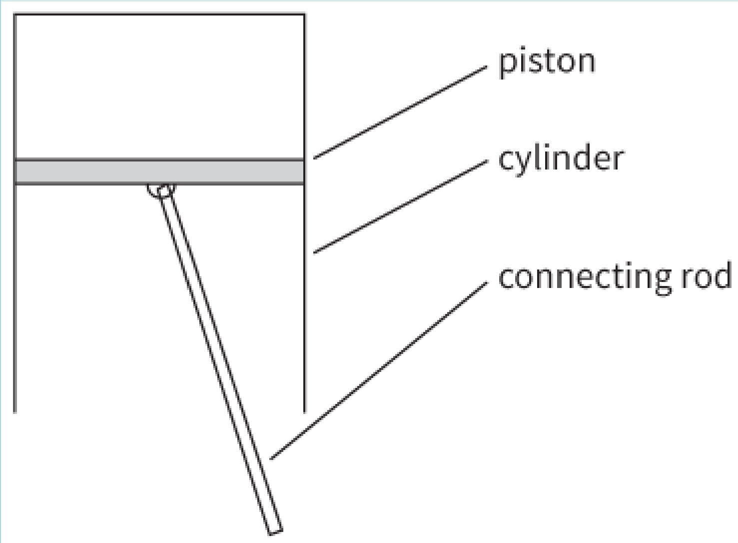


Figure 18.41

a Explain what is meant by simple harmonic motion. **[2]**

b Calculate the frequency of the oscillation. **[1]**

c The amplitude of the oscillation is 12.5 cm. Calculate:

i the maximum speed at which the piston moves **[2]**

ii the maximum acceleration of the piston **[2]**

iii the force required on the piston to produce the maximum acceleration. **[1]**

[Total: 8]

10 This diagram shows a turntable with a rod attached to it a distance 15 cm from the centre. The turntable is illuminated from the side so that a shadow is cast on a screen.

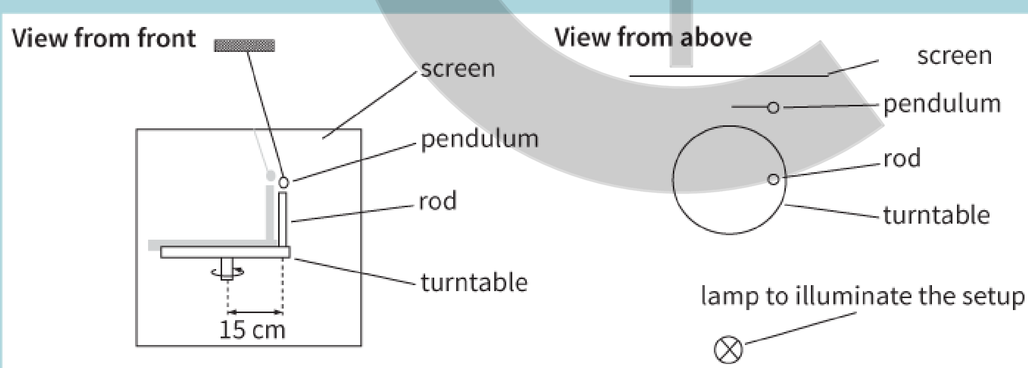


Figure 18.42

A simple pendulum is placed behind the turntable and is set oscillating so that it has an amplitude equal to the distance of the rod from the centre of the turntable.

The speed of rotation of the turntable is adjusted. When it is rotating at 1.5 revolutions per second the shadow of the pendulum and the rod are found to move back and forth across the screen exactly in phase.

a Explain what is meant by the term in phase. **[1]**

b Write down an equation to describe the displacement *x* of the pendulum from its equilibrium position and the angular frequency of the oscillation of the **[1]**

pendulum.

c The turntable rotates through 60° from the position of maximum displacement shown in the diagram.

i Calculate the displacement (from its equilibrium position) of the pendulum at this point. **[3]**

ii Calculate its speed at this point. **[2]**

iii Through what further angle must the turntable rotate before it has this speed again? **[1]**

[Total: 8]

11 When a cricket ball hits a cricket bat at high speed it can cause a standing wave to form on the bat. In one such example, the handle of the bat moved with a frequency of 60 Hz with an amplitude of 2.8 mm.

The vibrational movement of the bat handle can be modelled on simple harmonic motion.

a State the conditions for simple harmonic motion. **[2]**

b Calculate the maximum acceleration of the bat handle. **[2]**

c Given that the part of the bat handle held by the cricketer has a mass of 0.48 kg, calculate the maximum force produced on his hands. **[1]**

d The oscillations are damped and die away after about five complete cycles. Sketch a displacement–time graph to show the oscillations. **[2]**

[Total: 7]

12 Seismometers are used to detect and measure the shock waves that travel through the Earth due to earthquakes.

This diagram shows the structure of a simple seismometer. The shock wave will cause the mass to vibrate, causing a trace to be drawn on the paper scroll.

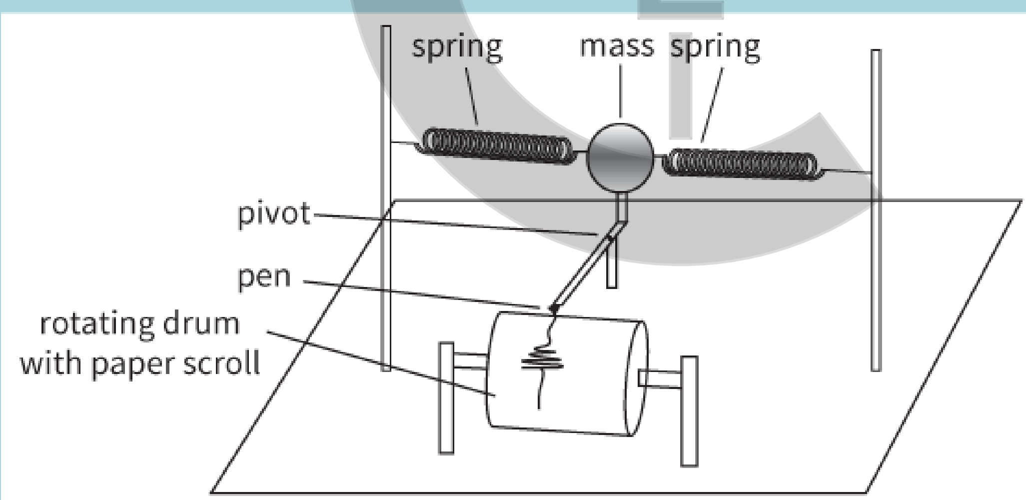


Figure 18.43

a The frequency of a typical shock wave is between 30 and 40 Hz. Explain why the natural frequency of the spring–mass system in the seismometer should be very much less than this range of frequencies. **[3]**

This graph shows the acceleration of the mass against its displacement when the seismometer is recording an earthquake.

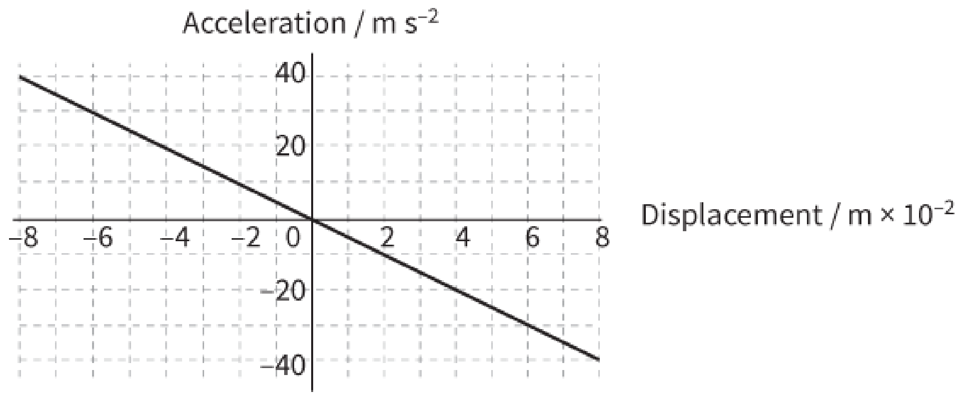


Figure 18.44

b What evidence does the graph give that the motion is simple harmonic? **[2]**

c Use information from the graph to calculate the frequency of the oscillation. **[4]**

[Total: 9]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the terms displacement, amplitude, period, frequency, angular frequency and phase difference	18.3			
express the period in terms of both frequency and angular frequency	18.3, 18.6			
understand that in simple harmonic motion there is a varying force on the oscillator, which is proportional to the displacement of the oscillator from a point and it is always directed towards that point	18.4			
recall, use and understand the importance of the equation: $a = -\omega^2 x$	18.7			
understand that the solution to the equation $a = -\omega^2 x$ is $x = x_0 \sin \omega t$	18.7			
use the equation: $v = v_0 \cos \omega t$	18.7			
use the equation: $v = \pm\omega\sqrt{(x_0^2 - x^2)}$	18.7			
understand the interchange between potential and kinetic energy in simple harmonic motion	18.8			
understand that the total energy of a simple harmonic oscillator remains constant and is determined by the amplitude of the oscillator, its mass and its frequency	18.8			
recall and use the equation $E = \frac{1}{2}m\omega^2 x_0$ for the total energy of an oscillator	18.8			
understand that a resistive force acting on an oscillator causes damping	18.9			
understand the term critical damping	18.10			
sketch displacement graphs showing the different types of damping	18.5			

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the concept of resonance	18.10			
understand that resonance occurs when the driving frequency equals the natural frequency of the oscillating system.	18.10			

〉 Chapter 19

Thermal physics

LEARNING INTENTIONS

In this chapter you will learn how to:

- relate a rise in temperature of an object to internal energy, the sum of the random distribution of kinetic and potential energies of the molecules in a system
- recall and use the first law of thermodynamics
- calculate the work done when the volume of a gas changes at constant pressure
- measure temperature using a physical property and state examples of such properties
- use the thermodynamic scale of temperature, and understand that the lowest possible temperature is zero kelvin and that this is known as absolute zero
- relate transfer of (thermal) energy as being due to a difference in temperature and understand thermal equilibrium
- define and use specific heat capacity and specific latent heat, and outline how these quantities can be measured.

BEFORE YOU START

- Write down the boiling point and melting point of water, and the names scientists use to describe *changes of state* (from solid to liquid, and from liquid to solid, and so on).
- List some difference between atoms and molecules. You can treat them both as simply 'particles' and not worry about how many atoms a molecule contains.

FROM WATER TO STEAM

When water boils, it changes state – it turns to water vapour. A liquid has become a gas. This is a familiar process, but Figure 19.1 shows a dramatic example of such a change of state. This is a geyser in New Zealand, formed when water is trapped underground where it is in contact with hot rocks. The temperature and pressure of the water build up until it suddenly erupts above the surface to form a tall plume of scalding water and water vapour. What happens underground to cause this effect? Would this be a useful effect to have happen near where you live?

Figure 19.1: At regular intervals of time, the White Lady Geyser, near Rotorua in New Zealand, throws up a plume of water (liquid) and water vapour (gas).

19.1 Changes of state

The kinetic model of matter can be used to describe the structure of solids, liquids and gases. You should recall that the kinetic model describes the behaviour of matter in terms of moving particles (atoms, molecules, and so on). Figure 19.2 should remind you of how we picture the three states of matter at the atomic scale:

- In a solid, the particles are close together, tightly bonded to their neighbours, and vibrating about fixed positions.
- In a gas, the particles have broken free from their neighbours; they are widely separated and are free to move around within their container.

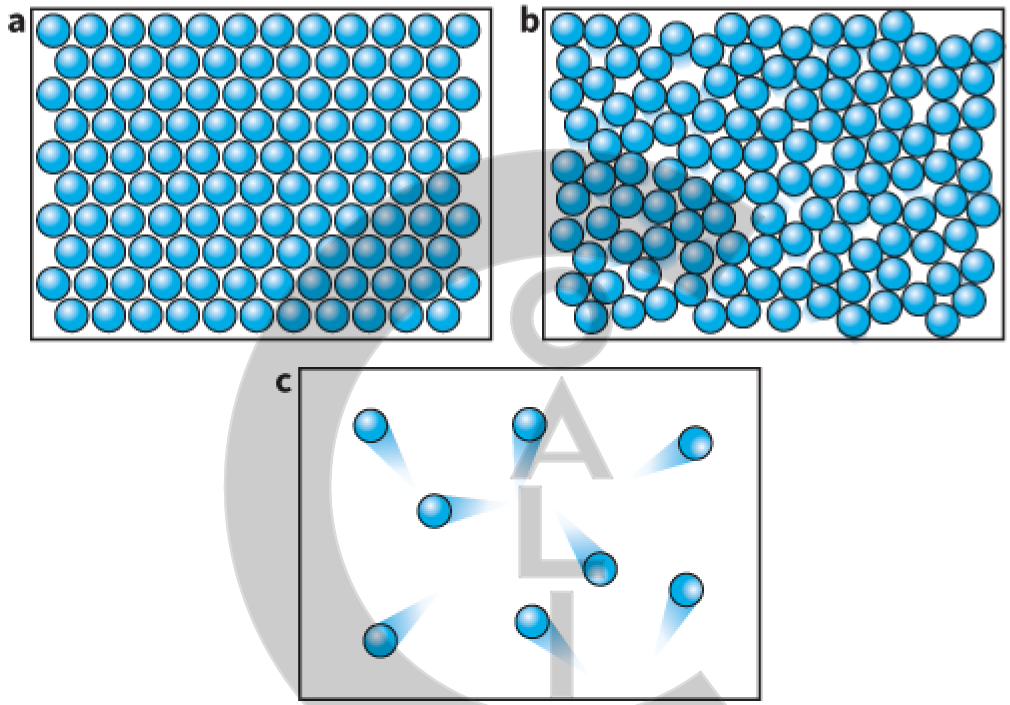


Figure 19.2: Typical arrangements of atoms in **a** a solid, **b** a liquid and **c** a gas.

Question

1 Describe a liquid in terms of the arrangement of its particles, the bonding between them and their motion.

In this chapter, we will extend these ideas to look at the energy changes involved when materials are heated and cooled.

19.2 Energy changes

Energy must be supplied to raise the temperature of a solid, to melt it, to heat the liquid and to boil it. Where does this energy go? It is worth taking a close look at a single change of state and thinking about what is happening on the atomic scale. Figure 19.3a shows a suitable arrangement. A test tube containing octadecanoic acid (a white, waxy substance at room temperature) is warmed in a water bath. At 80 °C, the substance is a clear liquid. The tube is then placed in a rack and allowed to cool. Its temperature is monitored, either with a thermometer or with a temperature probe and datalogger. Figure 19.3b shows typical results.

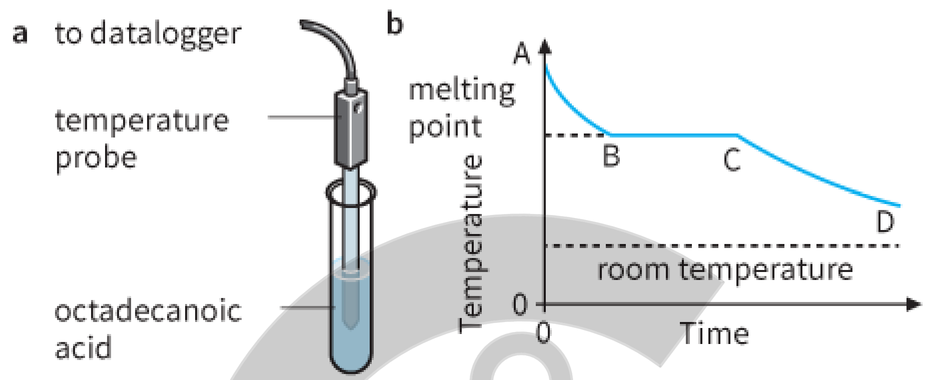


Figure 19.3: a Apparatus for obtaining a cooling curve, and **b** typical results.

The temperature drops rapidly at first, then more slowly as it approaches room temperature. The important section of the graph is the region BC. The temperature remains steady for some time. The clear liquid is gradually returning to its white, waxy solid state. It is essential to note that energy is still being lost even though the temperature is not decreasing. When no liquid remains, the temperature starts to drop again.

From the graph, we can deduce the melting point of octadecanoic acid. This is a technique used to help identify substances by finding their melting points.

Heating ice

In some ways, it is easier to think of the experiment in reverse. What happens when we heat a substance?

Imagine taking some ice from the deep freeze. Put the ice in a well-insulated container and heat it at a steady rate. Its temperature will rise; eventually, we will have a container of water vapour.

Water vapour and steam mean the same thing–an invisible gas. The ‘steam’ that you see when a kettle boils is *not* a gas; it is ‘wet steam’ – a cloud of tiny droplets of liquid water.

Figure 19.4 shows the results we might expect if we could carry out this idealised experiment. Energy is supplied to the ice at a constant rate. We will consider the different sections of this graph in some detail, in order to describe where the energy is going at each stage.

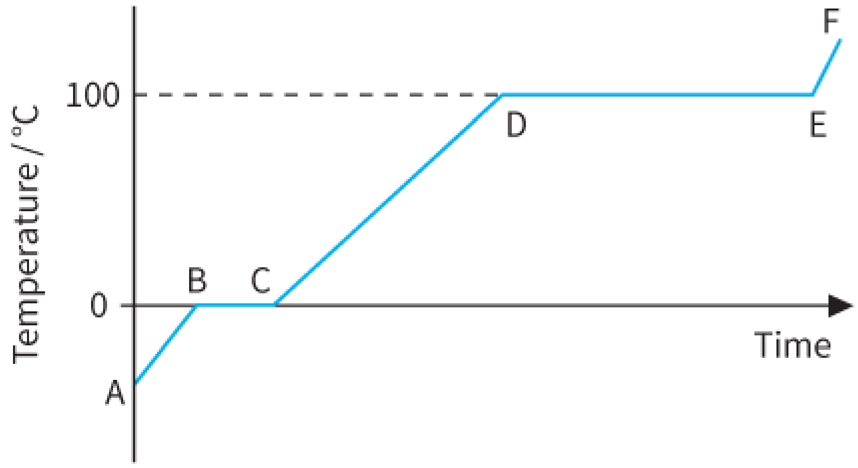


Figure 19.4: A graph of temperature against time for water, heated at a steady rate.

We need to think about the kinetic and potential energies of the molecules. If they move around more freely and faster, their kinetic energy has increased. If they break free of their neighbours and become more disordered, their electrical potential energy has increased.

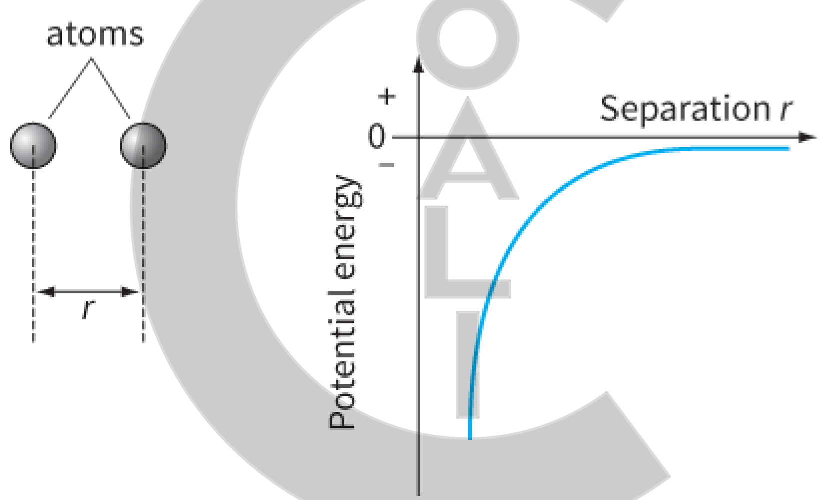


Figure 19.5: The electrical potential energy of atoms is negative and increases as they get further apart.

You know that the kinetic energy of a particle is the energy it has due to its motion. Figure 19.5 shows how the electrical potential energy of two isolated atoms depends on their separation. Work must be done (energy must be put in) to separate neighbouring atoms–think about the work you must do to snap a piece of plastic or to tear a sheet of paper. The graph shows that:

- The electrical potential energy of two atoms very close together is large and negative.
- As the separation of the atoms increases, their potential energy also increases.
- When the atoms are completely separated, their potential energy is maximum and has a value of zero.

It may seem strange that the potential energy is negative and you will see in Chapter 21 why this is so. At the moment, just notice that, as atoms or molecules become further apart, their potential energy becomes less negative and so they have more potential energy.

Now look back at the graph shown in Figure 19.4.

Section AB

The ice starts below 0 °C; its temperature rises. The molecules gain energy and vibrate more and more. Their vibrational kinetic energy is increasing. There is very little change in the mean separation between the molecules and hence very little change in their electrical potential energy.

Section BC

The ice melts at 0 °C. The molecules become more disordered. There is a modest increase in their electrical potential energy.

Section CD

The ice has become water. Its temperature rises towards 100 °C. The molecules move increasingly rapidly. Their kinetic energy is increasing. There is very little change in the mean separation between the molecules and therefore very little change in their electrical potential energy.

Section DE

The water is boiling. The molecules are becoming completely separate from one another. There is a large increase in the separation between the molecules and hence their electrical potential energy has increased greatly. Their movement becomes very disorderly.

Section EF

The steam is being heated above 100 °C. The molecules move even faster. Their kinetic energy is increasing. The molecules have maximum electrical potential energy of zero.

You should see that, when water is heated, each change of state (melting, boiling) involves the following:

- there must be an input of energy
- the temperature does not change
- the molecules are breaking free of one another
- their potential energy is increasing.

In between the changes of state:

- the input of energy raises the temperature of the substance
- the molecules move faster
- their kinetic energy is increasing.

The hardest point to appreciate is that you can put energy into the system without its temperature rising. This happens during any change of state; the energy goes to breaking the bonds between neighbouring molecules. The energy that must be supplied to cause a change of state is sometimes called ‘latent heat’. The word ‘latent’ means ‘hidden’ and refers to the fact that, when you melt something, its temperature does not rise and the energy that you have put in seems to have disappeared.

It may help to think of temperature as a measure of the average kinetic energy of the molecules. When you put a thermometer in some water to measure its temperature, the water molecules collide with the thermometer and share their kinetic energy with it. At a change of state, there is no change in kinetic energy, so there is no change in temperature.

Notice that melting the ice (section BC) takes much less energy than boiling the same amount of water (section DE). This is because, when a solid melts, the molecules are still bonded to most of their immediate neighbours. When a liquid boils, each molecule breaks free of all of its neighbours. Melting may involve the breaking of one or two bonds per molecule, whereas boiling involves breaking eight or nine.

Evaporation

A liquid does not have to boil to change into a gas. A puddle of rain-water dries up without having to be heated to 100 °C. When a liquid changes to a gas without boiling, we call this **evaporation**.

Any liquid has some vapour associated with it. If we think about the microscopic picture of this, we can see why (Figure 19.6). Within the liquid, molecules are moving about. Some move faster than others, and can break free from the bulk of the liquid. They form the vapour above the liquid. Some molecules from the vapour may come back into contact with the surface of the liquid, and return to the liquid. However, there is a net outflow of energetic molecules from the liquid, and eventually it will evaporate away completely.

You may have had your skin swabbed with alcohol or ether before an injection. You will have noticed how cold your skin becomes as the volatile liquid evaporates. Similarly, you can become very cold if you get wet and stand around in a windy place. This cooling of a liquid is a very important aspect of evaporation.

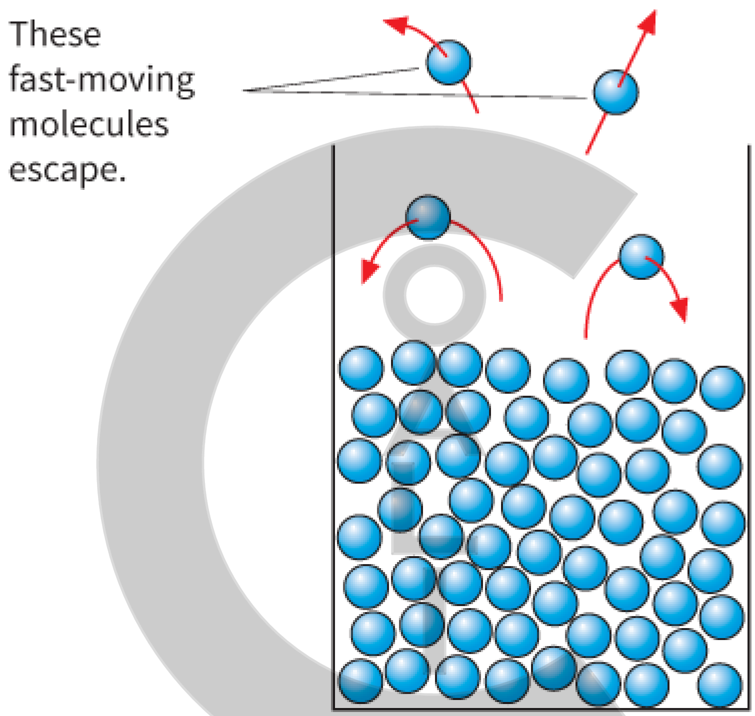


Figure 19.6: Fast-moving molecules leave the surface of a liquid – this is evaporation.

When a liquid evaporates, it is the most energetic molecules that are most likely to escape. This leaves molecules with a below-average kinetic energy. Since temperature is a measure of the average kinetic energy of the molecules, it follows that the temperature of the evaporating liquid must fall.

Question

2 Use the kinetic model of matter to explain the following:

- **a** If you leave a pan of water on the hob for a long time, it does not all boil away as soon as the temperature reaches 100 °C.
- **b** It takes less energy to melt a 1.0 kg block of ice at 0 °C than to boil away 1.0 kg of water at 100 °C.

19.3 Internal energy

All matter is made up of particles, which we will refer to here as ‘molecules’. Matter can have energy. For example, if we lift up a stone, it has gravitational potential energy. If we throw it, it has kinetic energy. Kinetic and potential energies are the two general forms of energy. We consider the stone’s potential and kinetic energies to be properties or attributes of the stone itself; we calculate their values (mgh and $\frac{1}{2}mv^2$) using the mass and speed of the stone.

Now think about another way in which we could increase the energy of the stone: we could heat it (Figure 19.7). Now where does the energy from the heater go? The stone’s gravitational potential and kinetic energies do not increase; it is not higher or faster than before. The energy seems to have disappeared into the stone.

Figure 19.7: Increasing the internal energy of a stone.

Of course, you already know the answer to this. The stone gets hotter, and that means that the molecules that make up the stone have more energy, both kinetic and electrical potential. They vibrate more and faster, and they move a little further apart. This energy of the molecules is known as the **internal energy** of the stone. The internal energy of a system (such as the heated stone) is defined as the sum of the random distribution of kinetic and potential energies of its atoms or molecules.

Molecular energy

Earlier in this chapter, where we studied the phases of matter, we saw how solids, liquids and gases could be characterised by differences in the arrangement, order and motion of their molecules. We could equally have said that, in the three phases, the molecules have different amounts of kinetic and potential energy.

Now, it is a simple problem to find the internal energy of an amount of matter. We add up the kinetic and potential energies associated with all the molecules in that matter. For example, consider the gas shown in Figure 19.8. There are ten molecules in the box, each having kinetic and potential energy. We can work out what all of these are and add them together, to get the total internal energy of the gas in the box.

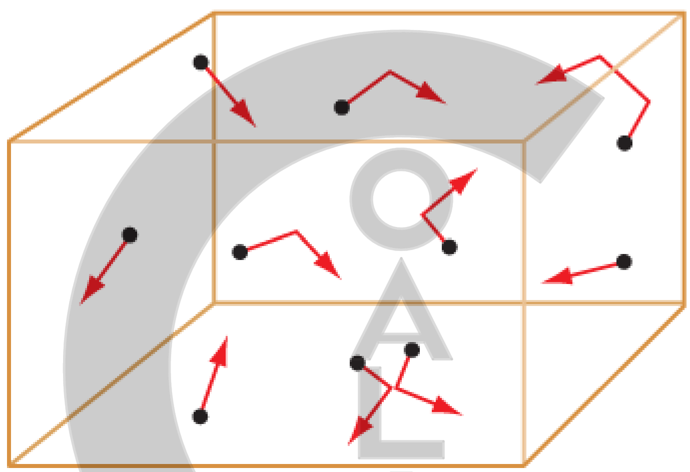


Figure 19.8: The molecules of a gas have both kinetic and potential energy.

Changing internal energy

There are two obvious ways in which we can increase the internal energy of some gas: we can heat it (Figure 19.9a), or we can do work on it by compressing it (Figure 19.9b).

Heating a gas

The walls of the container become hot and so its molecules vibrate more vigorously. The molecules of the cool gas strike the walls and bounce off faster. They have gained kinetic energy, and we say the temperature has risen.

Doing work on a gas

In this case, a wall of the container is being pushed inwards. The molecules of the cool gas strike a moving wall and bounce off faster. They have gained kinetic energy and again the temperature has risen. This explains why a gas gets hotter when it is compressed.

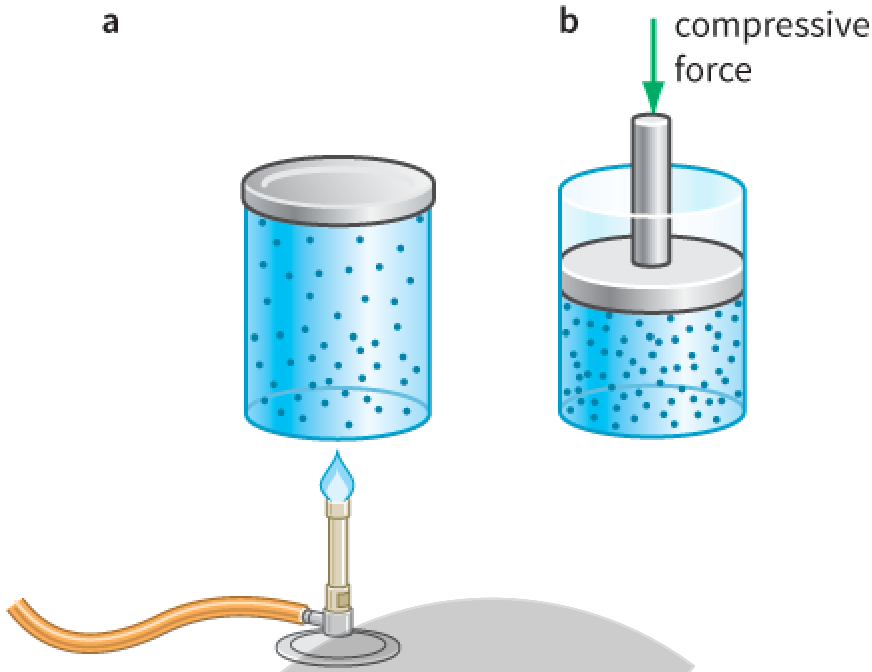


Figure 19.9: Two ways to increase the internal energy of a gas: **a** by heating it, and **b** by compressing it.

There are other ways in which the internal energy of a system can be increased: by passing an electric current through it, for example. However, doing work and heating are all we need to consider here.

The internal energy of a gas can also decrease; for example, if it loses heat to its surroundings, or if it expands so that it does work on its surroundings.

First law of thermodynamics

You will be familiar with the idea that energy is **conserved**; that is, energy cannot simply disappear, or appear from nowhere. This means that, for example, all the energy put into a gas by heating it and by doing work on it must end up in the gas; it increases the internal energy of the gas. We can write this as an equation:

increase in internal energy = energy supplied by heating + work done **on** the system

In symbols:

$$\Delta U = q + W$$

where ΔU is the increase in internal energy, q is the energy supplied **to** the system by heating and W is the work done on the system

This is known as the **first law of thermodynamics** and is a formal statement of the principle of conservation of energy. (It applies to all situations, not simply to a mass of gas.) Since you have learned previously that energy is conserved, it may seem to be a simple idea, but it took scientists a good many decades to understand the nature of energy and to appreciate that it is conserved.

You should note the sign convention that is used in the first law. A positive value of ΔU means that the internal energy **increases**, a positive value of q means that heat is added **to** the system, and a positive value of W means that work is done **on** the system. **Negative** values mean that internal energy **decreases**, heat is taken **away** from the system or work is done **by** the system.

Imagine a gas heated from the outside in a sealed container of constant volume. In this case, no work is done on the gas as the heat is added, so W is 0 and the first law equation $\Delta U = q + W$ becomes $\Delta U = q$. All the heat added becomes internal energy of the gas. If the container was able to expand a little as the heat is added then the situation needs some careful thought. Compressing a gas means work is done **on** the gas (W is positive); expanding a gas means work is done **by** the gas (W is negative as it pushes back and does work on the atmosphere). If the container expands then W is slightly negative and ΔU is slightly less than if the volume was constant.

A gas doing work

Gases exert pressure on the walls of their container. If a gas expands, the walls are pushed outwards – the gas has done work on its surroundings (W is negative, if the gas is the system). In a steam engine, expanding steam pushes a piston to turn the engine, and in a car engine, the exploding mixture of fuel and air does the same thing, so this is an important situation.

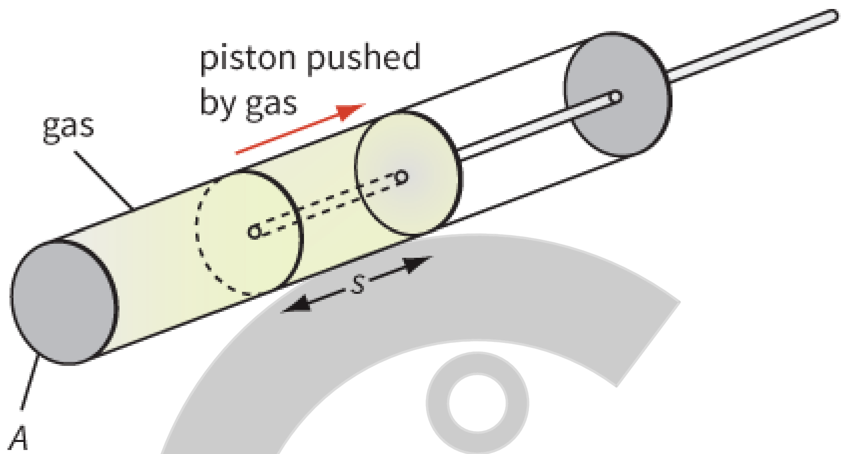


Figure 19.10: When a gas expands, it does work on its surroundings.

Figure 19.10 shows a gas at pressure p inside a cylinder of cross-sectional area A. The cylinder is closed by a moveable piston. The gas pushes the piston a distance s. If we know the force F exerted by the gas on the piston, we can deduce an expression for the amount of work done by the gas.

From the definition of pressure $\left(\text{pressure} = \frac{\text{force}}{\text{area}}\right)$, the force exerted by the gas on the piston is given by:

$$\text{force} = \text{pressure} \times \text{area}$$

$$F = p \times A$$

and the work done is force × displacement:

$$W = p \times A \times s$$

But the quantity $A \times s$ is the **increase** in volume of the gas; that is, the shaded volume in Figure 19.10. We call this ΔV, where the Δ indicates that it is a change in V. Hence, the work done by the gas in expanding is:

$$W = p\Delta V$$

KEY EQUATION

$$W = p\Delta V$$

Work done when the volume of a gas changes at constant pressure.

Notice that we are assuming that the pressure p does not change as the gas expands. This will be true if the gas is expanding against the pressure of the atmosphere, which changes only very slowly.

How does the first law of thermodynamics apply if you compress a gas? This can be done in different ways but we can consider two limiting ways.

Not allowing heat to enter or leave the system

This can be done by pushing the piston into the syringe very fast or by insulating the syringe. In this case, q is zero and $\Delta U = W$. All the work done by pushing in the piston increases the internal energy of the molecules. In this case, the kinetic energy of the molecules increases and the temperature increases, unless there is a change of state.

At constant temperature

Imagine pushing the piston very slowly into a syringe containing gas; so slowly that the temperature stays constant at room temperature. This change is known as an **isothermal change**. The kinetic energy of the molecules remains constant.

The molecules become slightly closer together and this may mean that their internal energy U becomes slightly less but the change is very small (unless the gas becomes a liquid). If U is constant, then ΔU is zero and $0 = q + W$. This means that, if you push the piston in and do positive work W, then q is negative, and heat is lost from the syringe. You can think of this as doing positive work on the system and, with no extra internal energy, the system must lose some heat to the surroundings, perhaps by conduction of heat through the walls of the syringe. Similarly, if you pull the piston out very slowly, W is negative and q is positive and heat enters the system.

Questions

3 Use the first law of thermodynamics to answer the following.

- **a** A gas is heated by supplying it with 250 kJ of thermal energy; at the same time, it is compressed so that 500 kJ of work is done on the gas. Calculate the change in the internal energy of the gas.
- **b** The same gas is heated as before with 250 kJ of energy. This time the gas is allowed to expand so that it does 200 kJ of work on its surroundings. Calculate the change in the internal energy of the gas.

4 When you blow up a balloon, the expanding balloon pushes aside the atmosphere. How much work is done against the atmosphere in blowing up a balloon from a very small volume to a volume of 2 litres (0.002 m^3)? (Atmospheric pressure = 1.0×10^5 N m^{-2}.)

19.4 The meaning of temperature

Picture a beaker of boiling water. You want to measure its temperature, so you pick up a thermometer that is lying on the bench. The thermometer reads 20 °C. You place the thermometer in the water and the reading goes up … 30 °C, 40 °C, 50 °C. This tells you that the thermometer is getting hotter; energy is being transferred from the water to the thermometer.

KEY IDEA

Thermal energy is transferred from a region of higher temperature to a region of lower temperature.

Eventually, the thermometer reading reaches 100 °C and it stops rising. Because the reading is steady, you can deduce that energy is no longer being transferred to the thermometer and so its scale tells you the temperature of the water.

This simple, everyday activity illustrates several points:

- We are used to the idea that a thermometer shows the temperature of something with which it is in contact. In fact, it tells you **its own temperature**.

  As the reading on the scale was rising, it wasn't showing the temperature of the water. It was showing that the temperature of the thermometer was rising.
- Energy is transferred from a hotter object to a cooler one. The temperature of the water was greater than the temperature of the thermometer, so energy transferred from one to the other.
- When two objects are at the same temperature, there is no transfer of energy between them. That is what happened when the thermometer reached the same temperature as the water, so it was safe to say that the reading on the thermometer was the same as the temperature of the water.

From this, you can see that temperature tells us about the direction in which energy flows. If two objects are placed in contact (so that energy can flow between them), it will flow from the hotter to the cooler. Energy flowing from a region of higher temperature to a region of lower temperature is called **thermal energy**. (Here, we are not concerned with the mechanism by which the energy is transferred. It may be by conduction, convection or radiation.)

When two objects are at the same temperature, they are in **thermal equilibrium** with each other. There will be no net transfer of thermal energy between them when they are in contact with each other – see Figure 19.11.

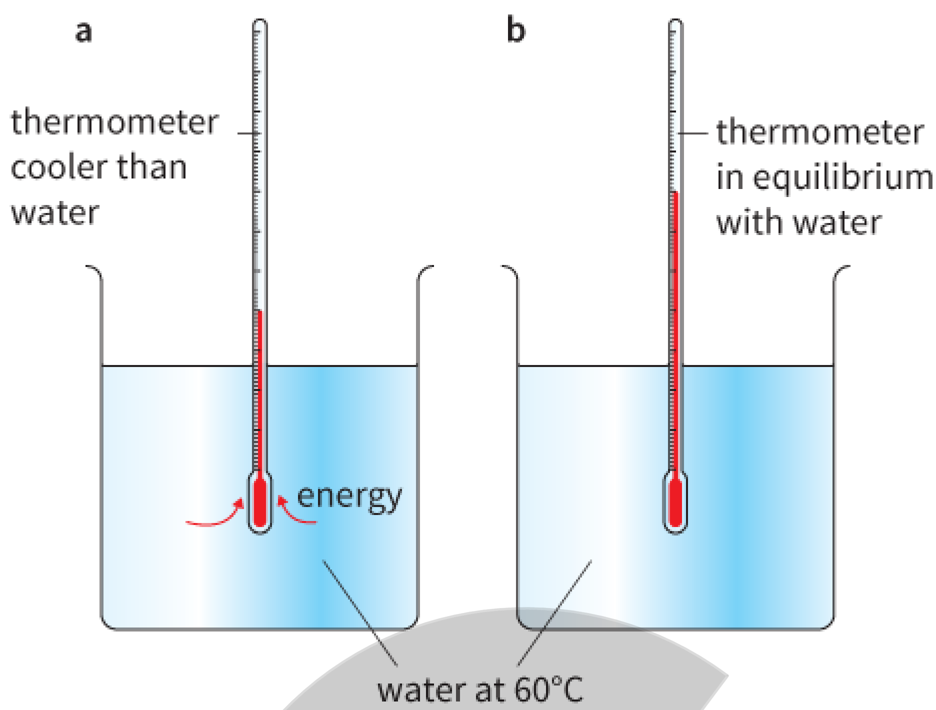


Figure 19.11: **a** Thermal energy is transferred from the hot water to the cooler thermometer because of the temperature difference between them. **b** When they are at the same temperature, there is no transfer of thermal energy and they are in thermal equilibrium.

The thermodynamic (Kelvin) scale

The Celsius scale of temperature is a familiar, everyday scale of temperature. It was originally based on the properties of water with the melting point of pure ice as 0°C and the boiling point of pure water as 100°C.

There is nothing special about these two temperatures. In fact, both the melting point and boiling point change if the pressure changes or if the water is impure. The **thermodynamic scale**, also known as the Kelvin scale, is a better scale in that one of its fixed points, **absolute zero**, is very important.

It is not possible to have a temperature lower than 0 K. Sometimes it is suggested that, at this temperature, matter has no energy left in it. This is not strictly true; it is more correct to say that, for any matter at absolute zero, it is impossible to **remove** any more energy from it. Hence, absolute zero is the temperature at which all substances have the minimum internal energy. (The kinetic energy of the atoms or molecules is zero and their electrical potential energy is minimum.)

We use different symbols to represent temperatures on these two scales: θ for the Celsius scale, and T for the thermodynamic (Kelvin) scale. To convert between the two scales, we use these relationships:

$$\theta\,(\text{in}\,^\circ\text{C}) = T\,(\text{in K}) - 273.15$$
$$T\,(\text{in K}) = \theta\,(\text{in}\,^\circ\text{C}) + 273.15$$

For most practical purposes, we round off the conversion factor to 273 as shown in the conversion chart (Figure 19.12).

KEY EQUATIONS

$$\theta\,(\text{in}\,^\circ\text{C}) = T\,(\text{in K}) - 273.15$$
$$T\,(\text{in K}) = \theta\,(\text{in}\,^\circ\text{C}) + 273.15$$

To convert temperatures between degrees Celsius and Kelvin.

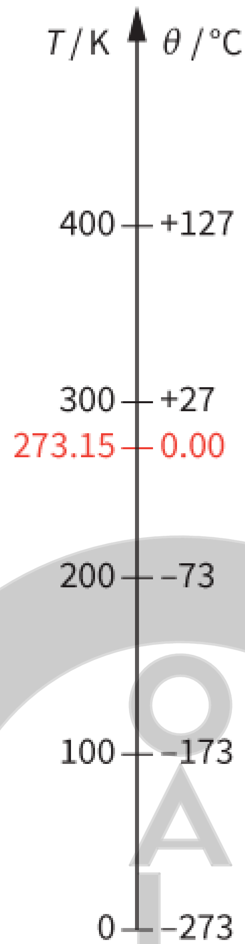


Figure 19.12: A conversion chart relating temperatures on the thermodynamic (Kelvin) and Celsius scales.

The thermodynamic scale of temperature is designed to overcome a problem with scales of temperature, such as the Celsius scale, which depends on the melting point and boiling point of pure water. To measure a temperature on this scale, you might use a liquid-in-glass thermometer. However, the expansion of a liquid may be non-linear. This means that if you compare the readings from two different types of liquid-in-glass thermometer, for example a mercury thermometer and an alcohol thermometer, you can only be sure that they will agree at the two fixed points on the Celsius scale. At other temperatures, their readings may differ.

KEY IDEA

Thermodynamic temperatures do not depend on the property of any particular substance.

The thermodynamic scale is said to be an absolute scale as it is not defined in terms of a property of any particular substance. It is based on the idea that the average kinetic energy of the particles of a substance increases with temperature. The average kinetic energy is the same for all substances at a particular thermodynamic temperature; it does not depend on the material itself. In fact, as you will see in Chapter 20, the average kinetic energy of a gas molecule is proportional to the thermodynamic temperature. So, if we can measure the average kinetic energy of the particles of a substance, we can deduce the temperature of that substance.

The thermodynamic scale has two fixed points:

- absolute zero, which is defined as 0 K

- the triple point of water; the temperature at which ice, water and water vapour can co-exist, which is defined as 273.16 K (equal to 0.01 °C).

So the gap between absolute zero and the triple point of water is divided into 273.16 equal divisions. Each division is 1 K. The scale is defined in this slightly odd way so that the scale divisions on the thermodynamic scale are equal in size to the divisions on the Celsius scale, making conversions between the two scales relatively easy.

A **change** in temperature of 1 K is thus equal to a **change** in temperature of 1 °C.

19.5 Thermometers

A thermometer is any device that can be used to measure temperature. Each type of thermometer makes use of some physical property of a material that changes with temperature. The most familiar is the length of a column of liquid in a tube, which gets longer as the temperature increases because the liquid expands. This is how a liquid-in-glass thermometer works; it depends on a change in density of a liquid. Other physical properties that can be used as the basis of thermometers include:

- the resistance of an electrical resistor or thermistor
- the e.m.f. (voltage) produced by a thermocouple
- the colour of an electrically heated wire
- the volume of a fixed mass of gas at constant pressure.

In each case, the thermometer must be calibrated at two or more known temperatures (such as the melting and boiling points of water, which correspond to 0 °C and 100 °C), and the scale between divided into equal divisions. There is no guarantee that two thermometers will agree with each other except at these fixed points.

Questions

5 **a** Convert each of the following temperatures from the Celsius scale to the thermodynamic scale: 0 °C, 20 °C, 120 °C, 500 °C, −23 °C, −200 °C.

b Convert each of the following temperatures from the thermodynamic scale to the Celsius scale: 0 K, 20 K, 100 K, 300 K, 373 K, 500 K.

6 The electrical resistance of a pure copper wire is mostly due to the vibrations of the copper atoms. Table 19.1 shows how the resistance of a length of copper wire is found to change as it is heated. Copy the table and add a column showing the temperatures in K. Draw a graph to show these data. (Start the temperature scale of your graph at 0 K.) Explain why you might expect the resistance of copper to be zero at this temperature.

Temperature / °C	Resistance / Ω
10	3120
50	3600
75	3900
100	4200
150	4800
220	5640
260	6120

Table 19.1: The variation of resistance with temperature for a length of copper wire.

In Chapter 10, we saw that electrical resistance changes with temperature. For metals, resistance increases with temperature at a fairly steady rate. However, for a thermistor, the resistance changes rapidly over a relatively narrow range of temperatures. A small change in temperature results in a large change in resistance, so a thermometer based on a thermistor will be sensitive over that range of temperatures.

A **thermocouple** is another electrical device which can be used as the sensor of a thermometer. Figure 19.13 shows the principle. Wires of two different metals, X and Y, are required. A length of metal X has a length of metal Y soldered to it at each end. This produces two **junctions**, which are the important parts of the thermocouple. If the two junctions are at different temperatures, an e.m.f. will be produced between the two free ends of the thermocouple, and can be measured using a voltmeter. The greater the difference in temperatures, the greater the voltage produced; however, this e.m.f. may not vary linearly with temperature, i.e., a graph of e.m.f. against temperature is not usually a straight line.

Electrical thermometers can measure across a great **range** of temperatures, from 0 K to hundreds or even thousands of kelvin.

Table 19.2 compares resistance and thermocouple thermometers.

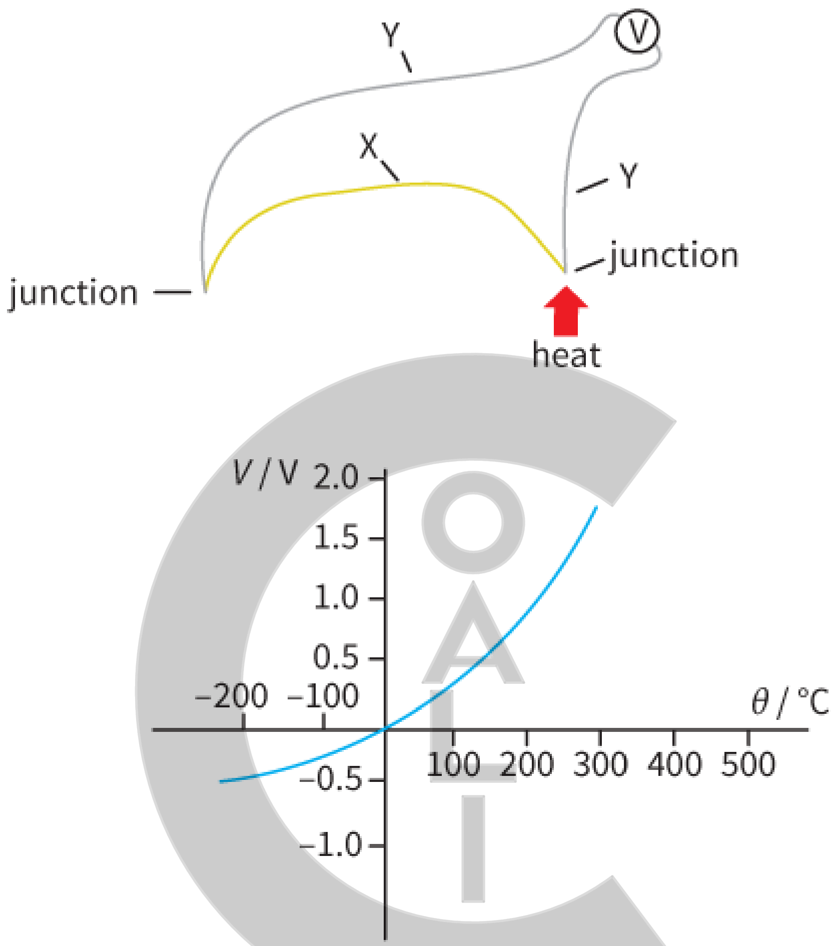


Figure 19.13: The construction of a thermocouple thermometer; the voltage produced depends on the temperature (as shown in the calibration graph) and on the metals chosen.

Feature	Resistance thermometer	Thermocouple thermometer
robustness	very robust	robust
range	thermistor: narrow range resistance wire: wide range	can be very wide
size	larger than thermocouple; has greater thermal capacity therefore slower acting	smaller than resistance thermometers; has smaller thermal capacity so quicker acting and can measure temperature at a point

Feature	Resistance thermometer	Thermocouple thermometer
sensitivity	thermistor: high sensitivity over narrow range resistance wire: less sensitive	can be sensitive if appropriate metals chosen
linearity	thermistor: fairly linear over narrow range resistance wire: good linearity	non-linear so requires calibration
remote operation	long conducting wires allow the operator to be at a distance from the thermometer	long conducting wires allow the operator to be at a distance from the thermometer

Table 19.2: Comparing resistance and thermocouple thermometers.

Question

7 Give **one** word for each of the following:

a adding a scale to a thermometer

b all the temperatures, from lowest to highest, which a thermometer can measure

c the extent to which equal rises in temperature give equal changes in the thermometer's output

d how big a change in output is produced by a given change in temperature.

19.6 Calculating energy changes

So far, we have considered the effects of heating a substance in qualitative terms, and we have given an explanation in terms of a kinetic model of matter. Now we will look at the amount of energy needed to change the temperature of something, and to produce a change of state.

Specific heat capacity

If we heat some material so that its temperature rises, the amount of energy we must supply depends on three things, the:

- mass m of the material we are heating
- temperature change Δq we wish to achieve
- material itself.

Some materials are easier to heat than others. It takes more energy to raise the temperature of 1 kg of water by 1 °C than to raise the temperature of 1 kg of alcohol by the same amount.

We can represent this in an equation. The amount of energy E that must be supplied is given by:

$$E = mc\Delta\theta$$

where c is the **specific heat capacity** of the material.

Rearranging this equation gives:

$$c = \frac{E}{m\Delta\theta}$$

The specific heat capacity of a material can be defined as a word equation as follows:

$$\text{specific heat capacity} = \frac{\text{energy supplied}}{\text{mass temperature change}}$$

Alternatively, specific heat capacity can be defined in words as follows:

The numerical value of the specific heat capacity of a substance is the energy required per unit mass of the substance to raise the temperature by 1 K (or 1 °C).

The word 'specific' here means 'per unit mass'; that is, per kg. From this form of the equation, you should be able to see that the units of c are J kg^{-1} K^{-1} (or J kg^{-1} $°C^{-1}$). Table 19.3 shows some values of specific heat capacity measured at 0 °C.

Specific heat capacity is related to the gradient of the sloping sections of the graph shown earlier in Figure 19.4. The steeper the gradient, the faster the substance heats up and hence the lower its specific heat capacity must be. Worked example 1 shows how to calculate the specific heat capacity of a substance.

WORKED EXAMPLE

1 When 26 400 J of energy is supplied to a 2.0 kg block of aluminium, its temperature rises from 20 °C to 35 °C. The block is well insulated so that there is no energy loss to the surroundings. Determine the specific heat capacity of aluminium.

Step 1 We are going to use the equation:

$$c = \frac{E}{m\Delta\theta}$$

We need to write down the quantities that we know:

E = 26 400 J m = 2.0 kg

$\Delta\theta = (35 - 20)\,°C = 15\,°C$ (or 15 K)

Step 2 Now substitute these values and solve the equation:

$$c = \frac{E}{m\Delta\theta} = \frac{26\,400}{(2.0 \times 15)} = 880\ \text{J kg}^{-1}\ \text{K}^{-1}$$

Substance	c / J kg^{-1} K^{-1}
aluminium	880
copper	380
lead	126
glass	500–680
ice	2100
water	4180
seawater	3950
ethanol	2500
mercury	140

Table 19.3: Values of specific heat capacity.

Questions

You will need to use data from Table 19.3 to answer these questions.

8 Calculate the energy that must be supplied to raise the temperature of 5.0 kg of water from 20 °C to 100 °C.

9 Which requires more energy – heating a 2.0 kg block of lead by 30 K or heating a 4.0 kg block of copper by 5.0 K?

10 A well-insulated 1.2 kg block of iron is heated using a 50 W heater for 4.0 min. The temperature of the block rises from 22 °C to 45 °C. Find the experimental value for the specific heat capacity of iron.

PRACTICAL ACTIVITY 19.1

Determining specific heat capacity *c*

How can we determine the specific heat capacity of a material? The principle is simple: supply a known amount of energy to a known mass of the material and measure the rise in its temperature. Figure 19.14 shows one practical way of doing this for a metal.

The metal is in the form of a cylindrical block of mass 1.00 kg. An electrical heater is used to supply the energy. This type of heater is used because we can easily determine the amount of energy supplied – more easily than if we heated the metal with a Bunsen flame, for example. An ammeter and voltmeter are used to make the necessary measurements.

Figure 19.14: A practical arrangement for determining the specific heat capacity of a metal.

A thermometer or temperature sensor is used to monitor the block's temperature as it is heated. The block must not be heated too quickly; we want to be sure that the energy has time to spread throughout the metal.

The block should be insulated by wrapping it in a suitable material – this is not shown in the illustration. It would be possible, in principle, to determine *c* by making just one measurement of temperature change, but it is better to record values of the temperature as it rises and plot a graph of temperature θ against time *t*. The method of calculating *c* is illustrated in Worked example 2.

Sources of error

This experiment can give reasonably good measurements of specific heat capacities. As noted earlier, it is desirable to have a relatively low rate of heating, so that energy spreads throughout the block. If the block is heated rapidly, different parts may be at different temperatures.

Thermal insulation of the material is also vital. Inevitably, some energy will escape to the surroundings. This means that more energy must be supplied to the block for each degree rise in temperature and so the experimental value for the specific heat capacity will be too high. One way around this is to cool the block below room temperature before beginning to heat it. Then, as its temperature rises past room temperature, heat losses will be zero in principle, because there is no temperature difference between the block and its surroundings.

WORKED EXAMPLE

2 An experiment to determine the specific heat capacity *c* of a 1.00 kg aluminium block is carried out; the block is heated using an electrical heater. The current in the heater is 4.17 A and the p.d. across it is 12 V. Measurements of the rising temperature of the block are represented by the graph shown in Figure 19.15. Determine a value for the specific heat capacity *c* of aluminium.

Step 1 Write down the equation that relates energy change to specific heat capacity:

$$E = mc\Delta\theta$$

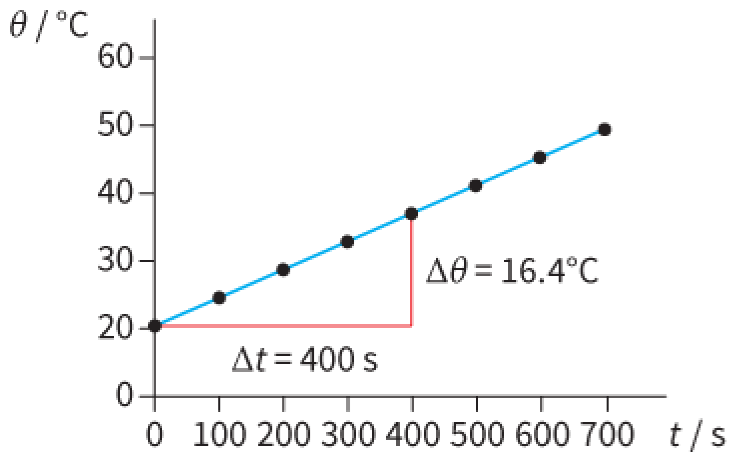


Figure 19.15: Graph of temperature against time for an aluminium block as it is heated.

Step 2 Divide both sides by a time interval Δt:

$$\frac{E}{\Delta t} = mc\left(\frac{\Delta\theta}{\Delta t}\right)$$

The quantity $\frac{E}{\Delta t}$ is the rate at which energy is supplied; that is, the power P of the heater.

The quantity $\frac{\Delta\theta}{\Delta t}$ is the rise of temperature of the block; that is, the gradient of the graph of θ against t.

Hence: $P = m \times c \times$ gradient

Step 3 Calculate the power of the heater and the gradient of the graph.

power = p.d. × current

$P = VI = 12 \times 4.17 \approx 50$ W

$$\text{gradient} = \frac{\Delta\theta}{\Delta t} = \frac{16.4}{400} = 0.041\,°\text{C}\,\text{s}^{-1}$$

Step 4 Substitute values, rearrange and solve.

$$50 = 1.00 \times c \times 0.041$$

$$c = \frac{50}{(1.00 \times 0.041)}$$

$$c = 1220\ \text{J kg}^{-1}\ \text{K}^{-1}$$

Questions

11 At higher temperatures than shown, the graph in Figure 19.15 deviates increasingly from a straight line. Suggest an explanation for this.

12 In measurements of the specific heat capacity of a metal, energy losses to the surroundings are a source of error. Is this a systematic error or a random error? Justify your answer.

13 In an experiment to measure the specific heat capacity of water, a student uses an electrical heater to heat some water. His results are shown. Calculate a value for the heat capacity of water. Comment on any likely sources of error.

mass of beaker = 150 g
mass of beaker + water = 672 g
current in the heater = 3.9 A
p.d. across the heater = 11.4 V
initial temperature = 18.5 °C
final temperature = 30.2 °C
time taken = 13.0 min

14 A block of paraffin wax was heated gently, at a steady rate. Heating was continued after the wax had completely melted. The graph of Figure 19.16 shows how the material's temperature varied during the experiment.

- **a** For each section of the graph (AB, BC and CD), describe the state of the material.
- **b** For each section, explain whether the material's internal energy is increasing, decreasing or remaining constant.
- **c** Consider the two sloping sections of the graph. State whether the material's specific heat capacity is greater when it is a solid or when it is a liquid. Justify your answer.

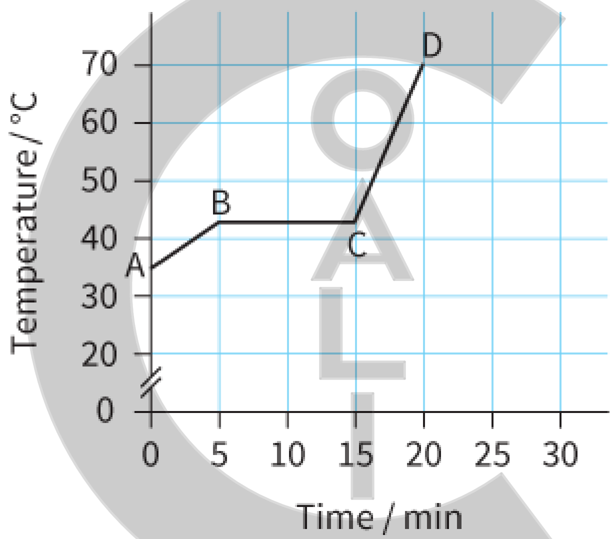


Figure 19.16: Temperature variation of a sample of wax, heated at a constant rate.

Specific latent heat

Energy must be supplied to melt or boil a substance. (In this case, there is no temperature rise to consider since the temperature stays constant during a change of state.) This energy is called latent heat.

The numerical value of the specific latent heat of a substance is the energy required per kilogram of the substance to change its state without any change in temperature.

When a substance melts, this quantity is called the **specific latent heat of fusion**; for boiling, it is the **specific latent heat of vaporisation**.

To calculate the amount of energy E required to melt or vaporise a mass m of a substance, we simply need to know its specific latent heat L:

$$E = mL$$

L is measured in J kg^{-1}. (Note that there is no 'per °C' since there is no change in temperature.) For water, the values are:

- specific latent heat of fusion of water, 330 kJ kg^{-1}
- specific latent heat of vaporisation of water, 2.26 MJ kg^{-1}

You can see that L for boiling water to form steam is roughly seven times the value for melting ice to form water. As we saw previously in the topic on heating ice, this is because, when ice melts, only one or two bonds are broken for each molecule; when water boils, several bonds are broken per molecule. Worked example 3 shows how to calculate these amounts of energy.

WORKED EXAMPLE

3 The specific latent heat of vaporisation of water is 2.26 MJ kg^{-1}. Calculate the energy needed to change 2.0 g of water into steam at 100 °C.

Step 1 We have been given the following quantities:

$m = 2.0$ g = 0.002 kg and $L = 2.26$ MJ kg^{-1}

Step 2 Substituting these values in the equation $E = mL$, we have:

energy = $0.002 \times 2.26 \times 10^6 = 4520$ J

Questions

15 The specific latent heat of fusion of water is 330 kJ kg^{-1}. Calculate the energy needed to change 2.0 g of ice into water at 0 °C. Suggest why the answer is much smaller than the amount of energy calculated in Worked example 3.

16 A sample of alcohol is heated with a 40 W heater until it boils. As it boils, the mass of the liquid decreases at a rate of 2.25 g per minute. Assuming that 80% of the energy supplied by the heater is transferred to the alcohol, estimate the specific latent heat of vaporisation of the alcohol. Give your answer in kJ kg^{-1}.

PRACTICAL ACTIVITY 19.2

Determining specific latent heat *L*

The principle of determining the specific latent heat of a material is similar to determining the specific heat capacity (but remember that there is no change in temperature).

Figure 19.17 shows how to measure the specific latent heat of vaporisation of water. A beaker containing water is heated using an electrical heater. A wattmeter (or an ammeter and a voltmeter) determines the rate at which energy is supplied to the heater. The beaker is insulated to minimise energy loss, and it stands on a balance. A thermometer is included to ensure that the temperature of the water remains at 100 °C.

The water is heated at a steady rate and its mass recorded at equal intervals of time. Its mass decreases as it boils.

A graph of mass against time should be a straight line whose gradient is the rate of mass loss. The wattmeter shows the rate at which energy is supplied to the water via the heater. We thus have:

$$\text{specific latent heat} = \left.\frac{\text{rate of supply of energy}}{\text{rate of loss of mass}}\right|$$

A similar approach can be used to determine the specific latent heat of fusion of ice. In this case, the ice is heated electrically in a funnel; water runs out of the funnel and is collected in a beaker on a balance.

As with any experiment, we should consider sources of error in measuring L and their effects on the final result. When water is heated to produce steam, some energy may escape to the surroundings so that the measured energy is greater than that supplied to the water. This systematic error gives a value of L, which is greater than the true value. When ice is melted, energy from the surroundings will conduct into the ice, so that the measured value of L will be an underestimate.

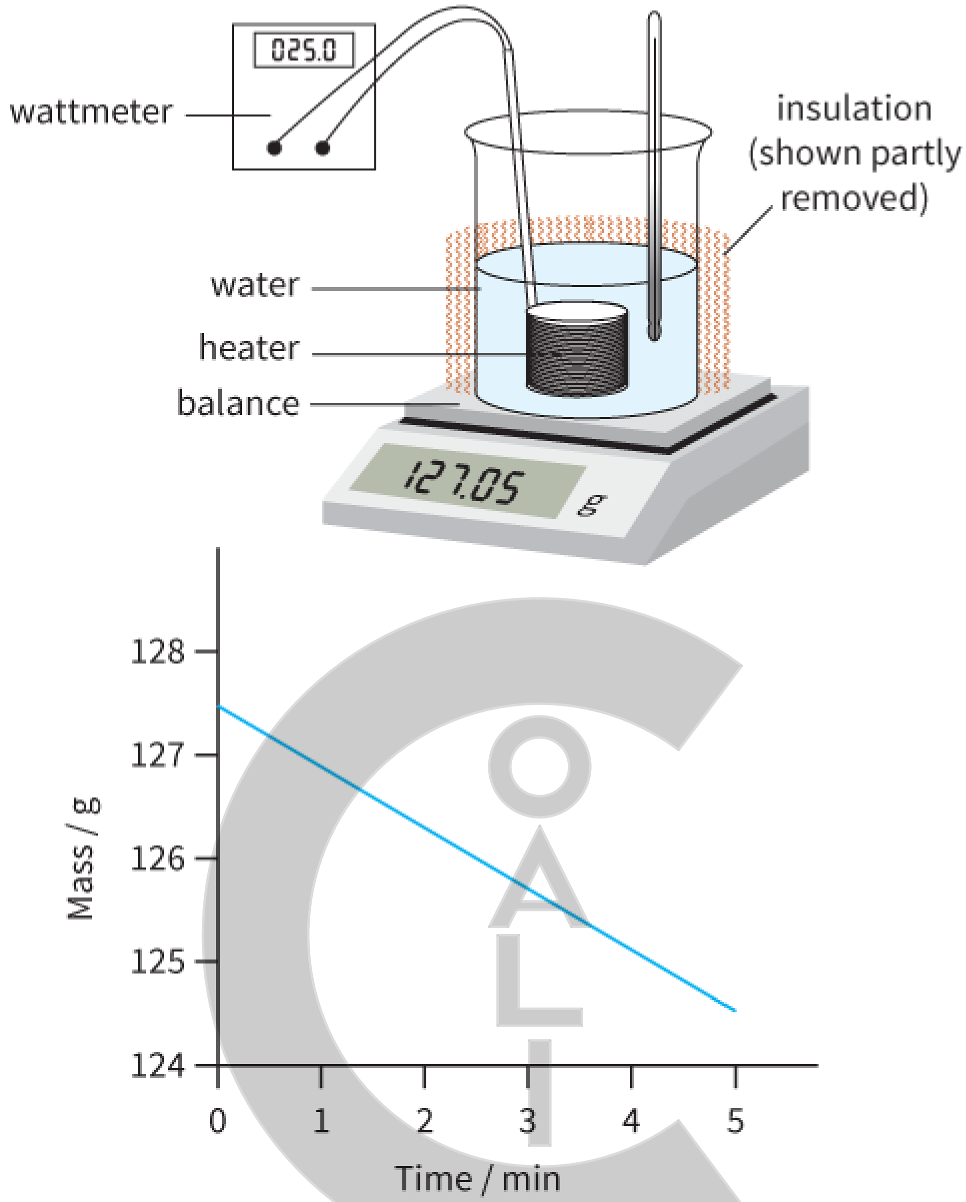


Figure 19.17: Determining the specific latent heat of vaporisation of water.

REFLECTION

List all the ideas in this chapter that are associated with an **increase** in temperature.

What strategies could you use to make sure you understand these?

SUMMARY

The kinetic model of matter allows us to explain behaviour (such as changes of state) and macroscopic properties (such as specific heat capacity and specific latent heat) in terms of the behaviour of molecules.

The internal energy of a system is the sum of the random distribution of kinetic and potential energies associated with the atoms or molecules that make up the system.

If the temperature of an object increases, there is an increase in its internal energy.

Internal energy also increases during a change of state, but there is no change in temperature.

The first law of thermodynamics expresses the conservation of energy:

increase in internal energy = energy supplied by heating + work done on the system

$W = p\Delta V$ is the work done **on** a gas when the volume of a gas changes at constant pressure. W is positive when the gas is compressed (ΔV is negative).

Temperatures on the thermodynamic (Kelvin) and Celsius scales of temperature are related by:

$$T\ (\text{K}) = \theta\ (^\circ\text{C}) + 273.15 \qquad \theta\ (^\circ\text{C}) = T\ (\text{K}) - 273.15$$

Absolute zero, 0 K, is the lowest possible temperature.

A thermometer makes use of a physical property of a material that varies with temperature.

The word equation for the specific heat capacity of a substance is:

$$\text{specific heat capacity} = \frac{\text{energy supplied}}{\text{mass} \times \text{temperature change}}$$

The specific heat capacity of a substance is the energy required per unit mass of the substance to raise the temperature by 1 K (or 1 °C).

The energy transferred in raising the temperature of a substance is given by:

$$E = mc\Delta\theta$$

The specific latent heat of a substance is the energy required per kilogram of the substance to change its state without any change in temperature:

$$E = mL$$

EXAM-STYLE QUESTIONS

1 The first law of thermodynamics can be represented by the expression: $\Delta U = q + W$.

An ideal gas is compressed at constant temperature.

Which row shows whether ΔU, q and W are negative, positive or zero during the change? **[1]**

	ΔU	q	W
A	negative	negative	positive
B	positive	positive	negative
C	zero	negative	positive
D	zero	positive	negative

2 What is the internal energy of an object? **[1]**

A the amount of heat supplied to the object

B the energy associated with the random movement of the atoms in the object

C the energy due to the attraction between the atoms in the object

D the potential and kinetic energy of the object.

3 Describe the changes to the kinetic energy, the potential energy and the total internal energy of the molecules of a block of ice as:

a it melts at 0 °C **[3]**

b the temperature of the water rises from 0 °C to room temperature. **[3]**

[Total: 6]

4 Explain, in terms of kinetic energy, why the temperature of a stone increases when it falls from a cliff and lands on the beach below. **[3]**

5 Explain why the barrel of a bicycle pump gets very hot as it is used to pump up a bicycle tyre. (Hint: the work done against friction is not large enough to explain the rise in temperature.) **[3]**

6 The so-called ‘zeroth law of thermodynamics’ states that if the temperature of body A is equal to the temperature of body B and the temperature of body B is the same as body C, then the temperature of body C equals the temperature of body A.

Explain, in terms of energy flow, why the concept of temperature would be meaningless if this law was not obeyed. **[2]**

7 **a** The first law of thermodynamics can be represented by the expression: $\Delta U = q + W$. State what is meant by all the symbols in this expression. **[3]**

b Figure 19.18 shows a fixed mass of gas that undergoes a change from A to B and then to C.

i During the change from A to B, 220 J of thermal energy (heat) is removed from the gas. Calculate the change in the internal energy of the gas. **[2]**

ii During the change from B to C, the internal energy of the gas decreases by 300 J. Using the first law of thermodynamics explain how this change can occur. **[2]**

[Total: 7]

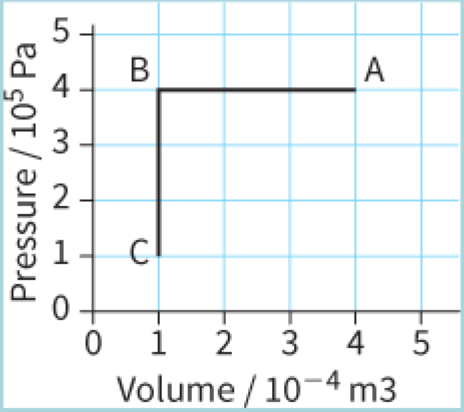


Figure 19.18

8 When a thermocouple has one junction in melting ice and the other junction in boiling water it produces an e.m.f. of 63 μV.

a What e.m.f. would be produced if the second junction was also placed in melting ice? **[1]**

b When the second junction is placed in a cup of coffee, the e.m.f. produced is 49 μV. Calculate the temperature of the coffee. **[2]**

c The second junction is now placed in a beaker of melting lead at 327 °C.

i Calculate the e.m.f. that would be produced. **[2]**

ii State the assumption that you make. **[1]**

[Total: 6]

9 The resistance of a thermistor at °C is 2000 Ω. At 100 °C the resistance falls to 200 Ω. When the thermistor is placed in water of constant temperature, its resistance is 620 Ω.

a Assuming that the resistance of the thermistor varies linearly with temperature, calculate the temperature of the water. **[2]**

b The temperature of the water on the thermodynamic scale is 280 K.

By reference to the particular features of the thermodynamic scale of temperature, **comment** on your answer to part **a**. **[3]**

[Total: 5]

10 a A 500 W kettle contains 300 g of water at 20 °C. Calculate the minimum time it would take to raise the temperature of the water to boiling point. **[5]**

b The kettle is allowed to boil for 2 minutes. Calculate the mass of water that remains in the kettle. State any assumptions that you make. **[4]**

(Specific heat capacity of water = 4.18×10^3 J kg^{-1} °C^{-1}; specific latent heat of vaporisation of water = 2.26×10^6 J kg^{-1}.)

[Total: 9]

11 a Define specific heat capacity of a substance. **[2]**

b A mass of 20 g of ice at −15 °C is taken from a freezer and placed in a beaker containing 200 g of water at 26 °C. Data for ice and water are given in Table 19.5.

	Specific heat capacity / J kg^{-1} K^{-1}	Specific latent heat of fusion / J kg^{-1}
ice	2.1×10^3	3.3×10^5
water	4.2×10^3	

Table 19.5

i Calculate the amount of thermal energy (heat) needed to convert all the ice at −15 °C to water at 0 °C. **[2]**

ii Calculate the final temperature *T* of the water in the beaker assuming that the beaker has negligible mass. **[3]**

[Total: 7]

12 a Define specific latent heat and explain the difference between latent heat of fusion and latent heat of vaporisation. **[3]**

b An electric heater generating power of 120 W is immersed in a beaker of liquid that is placed on a balance. When the liquid begins to boil it is noticed that the mass of the beaker and liquid decreases by 6.2 g every minute.

i State how this shows that the liquid is boiling at a steady rate. **[1]**

ii Calculate a value for the specific latent heat of vaporisation of the liquid. **[2]**

iii State and explain whether the value determined in **ii** is likely to be larger or smaller than the accepted value. **[2]**

[Total: 8]

13 a Explain why energy is needed for boiling even though the temperature of the liquid remains constant. **[2]**

This diagram shows an apparatus that can be used to measure the specific latent heat of vaporisation of nitrogen.

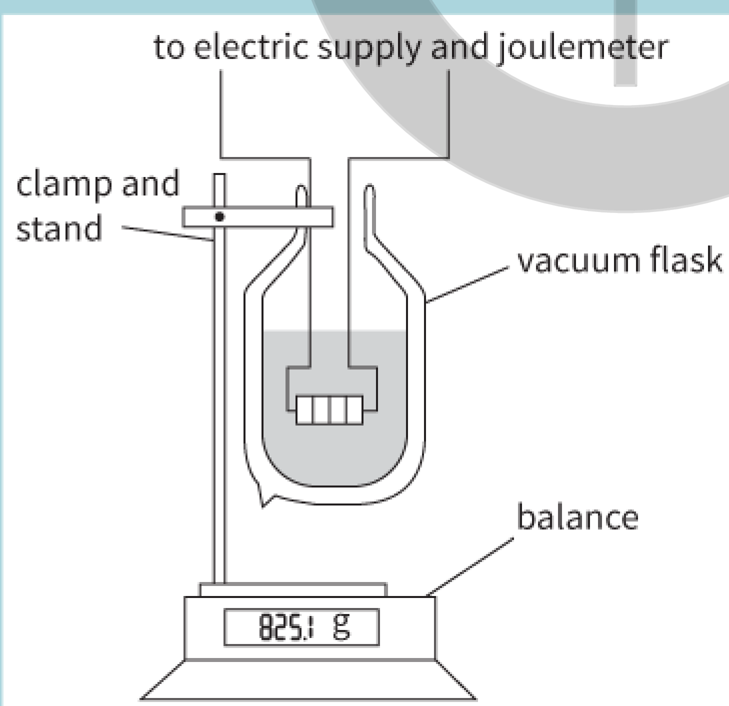


Figure 19.19

b Suggest why the nitrogen is contained in a vacuum flask. **[1]**

c The change in mass of the nitrogen is measured over a specific time interval with the heater switched off. The heater is switched on, transferring energy at 40 W, and the change of mass is found once more.

The results are shown in the table.

	Initial reading on balance / g	Final reading on balance / g	Time / minutes
heater off	834.7	825.5	4
heater on	825.5	797.1	2

Table 19.6

Calculate the specific latent heat of vaporisation of liquid nitrogen. **[4]**

[Total: 7]

14 a i Explain what is meant by internal energy. **[2]**

ii Explain what is meant by the absolute zero of temperature. **[2]**

b An electric hot water heater has a power rating of 9.0 kW. The water is heated as it passes through the heater. Water flows through the heater at a speed of 1.2 m s^{-1} through pipes that have a total cross-sectional area of 4.8×10^{-5} m^2. The temperature of the water entering the heater is 15 °C.

i Calculate the mass of water flowing through the heater each second. **[2]**

ii Calculate the temperature at which the water leaves the heater. **[3]**

iii State any assumptions you have made in your calculation. **[1]**

iv It is possible to adjust the temperature of the water from the heater. Suggest how the temperature of the water could be increased without increasing the power of the heater. **[1]**

(Density of water = 1000 kg m^{-3}; specific heat capacity of water = 4200 J kg^{-1} °C^{-1}.)

[Total: 11]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic...	Needs more work	Almost there	Ready to move on
understand internal energy, determined by the state of the system, as the sum of a random distribution of kinetic and potential energies of the molecules of a system	19.3			
relate a rise in temperature of an object to an increase in its internal energy	19.3			
recall and use $W = p\Delta V$ for the work done when the volume of a gas changes at constant pressure and understand the difference between work done by a gas and the work done on a gas	19.3			
recall and use the first law of thermodynamics: $\Delta U = q + W$	19.3			
understand that (thermal) energy is transferred from a region of higher temperature to a region of lower temperature and that regions of equal temperature are in thermal equilibrium	19.4			
understand the use of a physical property that varies with temperature to measure temperature and state examples of such properties	19.5			
understand that thermodynamic temperature does not depend on the property of any particular substance and recall and use: $T / \text{K} = \theta / {}^\circ\text{C} + 273.15$	19.4			
understand that the lowest possible temperature is zero kelvin, absolute zero	19.4			
define and use specific heat capacity	19.6			
define and use specific latent heat and distinguish between specific latent heat of fusion and specific latent heat of vaporisation.	19.6			

〉 Chapter 20

Ideal gases

LEARNING INTENTIONS

In this chapter you will learn how to:

- measure amounts of a substance in moles and find the number of particles using molar quantities
- solve problems using the equation of state $pV = nRT$ for an ideal gas
- deduce a relationship between pressure, volume and the microscopic properties of the molecules of a gas, stating the assumptions of the kinetic theory of gases
- relate the kinetic energy of the molecules of a gas to its temperature and calculate root-mean-square speeds.

BEFORE YOU START

- With a classmate, write down what you know about Brownian motion and what it shows about the molecules in a gas.
- Try to explain to a classmate, in terms of momentum change, why a ball hitting a wall exerts a force on it.
- List Newton's laws of motion.

THE IDEA OF A GAS

Figure 20.1 shows a weather balloon being launched. Balloons like this carry instruments high into the atmosphere, to measure pressure, temperature, wind speed and other variables.

The balloon is filled with helium so that its overall density is less than that of the surrounding air. The result is an upthrust on the balloon, greater than its weight, so that it rises upwards. As the balloon moves upwards, the pressure of the surrounding atmosphere decreases so that the balloon expands. The temperature drops, which tends to make the gas in the balloon shrink. In this chapter, we will look at the behaviour of gases as their pressure, temperature and volume change.

Figure 20.1: A weather balloon being launched.

20.1 Particles of a gas

We picture the particles of a gas as being fast-moving. They bounce off the walls of their container (and off each other) as they travel around at high speed (see Figure 20.2). How do we know that these particles are moving like this?

It is much harder to visualise the particles of a gas than those of a solid, because they move about in such a disordered way, and most of a gas is empty space. The movement of gas particles was investigated in the 1820s by a Scottish botanist, Robert Brown. He was using a microscope to look at pollen grains suspended in water, and saw very small particles moving around inside the water. He then saw the same motion in particles of dust in the air. It is easier in the laboratory to look at the movement of tiny particles of smoke in air. The particles are seen to be moving in a random, haphazard and jerky motion that we believe is caused by them being hit by invisible molecules of water or air around them. The pollen and dust particles are big enough to see in an ordinary microscope but air molecules are too small to see.

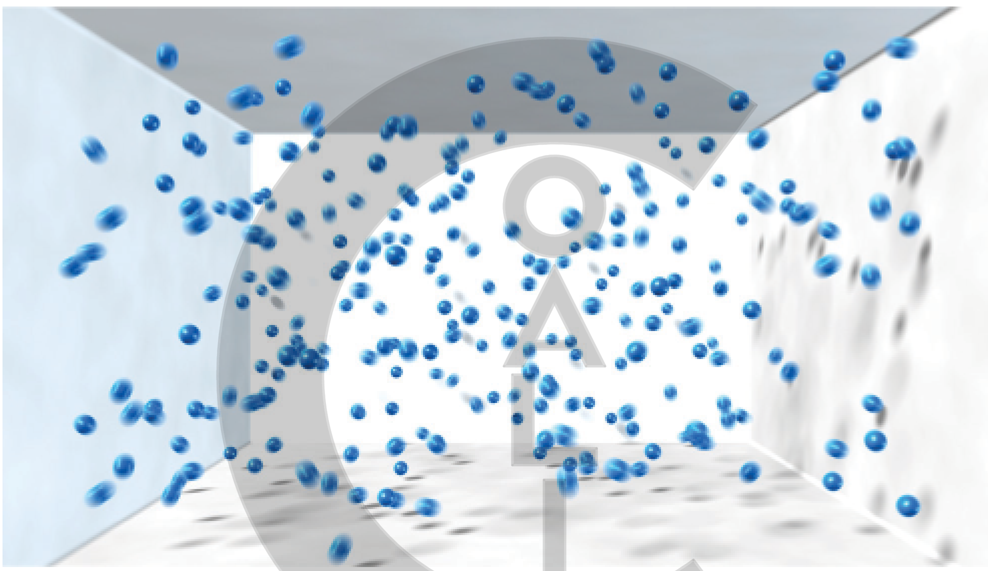


Figure 20.2: Particles of a gas – collisions with the walls of the container cause the gas' pressure on the container. (Particles do not have shadows like this. The shadows are added here to show depth.)

Fast molecules

For air at standard temperature and pressure (STP, −0 °C and 100 kPa), the average speed of the molecules is about 400 m s^{-1}. At any moment, some are moving faster than this and others more slowly. If we could follow the movement of a single air molecule, we would find that, some of the time, its speed was greater than this average; at other times, it would be less. The velocity (magnitude and direction) of an individual molecule changes every time it collides with anything else.

This value for molecular speed is reasonable. It is comparable to (but greater than) the speed of sound in air (approximately 330 m s^{-1} at STP). Very fast-moving particles can easily escape from the Earth's gravitational field. The required escape velocity is about 11 km s^{-1}. Since we still have an atmosphere, on average, the air molecules must be moving much more slowly than this value.

20.2 Explaining pressure

A gas exerts pressure on any surface with which it comes into contact. Pressure is a macroscopic property, defined as the force exerted per unit area of the surface.

The pressure of the atmosphere at sea level is approximately 100 000 Pa. The surface area of a typical person is 2.0 m^2. Hence the force exerted on a person by the atmosphere is about 200 000 N. This is equivalent to the weight of about 200 000 apples!

Fortunately, air inside the body presses outwards with an equal and opposite force, so we do not collapse under the influence of this large force. We can explain the macroscopic phenomenon of pressure by thinking about the behaviour of the microscopic particles that make up the atmosphere.

Figure 20.3 shows the movement of a single molecule of air in a box. It bounces around inside, colliding with the various surfaces of the box. At each collision, it exerts a small force on the box. The pressure on the inside of the box is a result of the forces exerted by the vast number of molecules in the box. Two factors affect the force, and hence the pressure, that the gas exerts on the box:

- the number of molecules that hit each side of the box in one second
- the force with which a molecule collides with the wall.

If a molecule of mass m hits the wall head-on with a speed v it will rebound with a speed v in the opposite direction. The change in momentum of the molecule is $2mv$. Since force is equal to rate of change of momentum, the higher the speed of the molecule the greater the force that it exerts as it collides with the wall. Hence, the pressure on the wall will increase if the molecules move faster.

If the piston in a bicycle pump is pushed inwards, but the temperature of the gas inside is kept constant, then more molecules will hit the piston in each second, but each collision will produce the same force because the temperature and therefore the average speed of the molecules is the same. The increased rate of collisions alone means that the force on the piston increases and thus the pressure rises. If the temperature of the gas in a container rises then the molecules move faster and hit the sides faster and more often; both of these factors cause the pressure to rise.

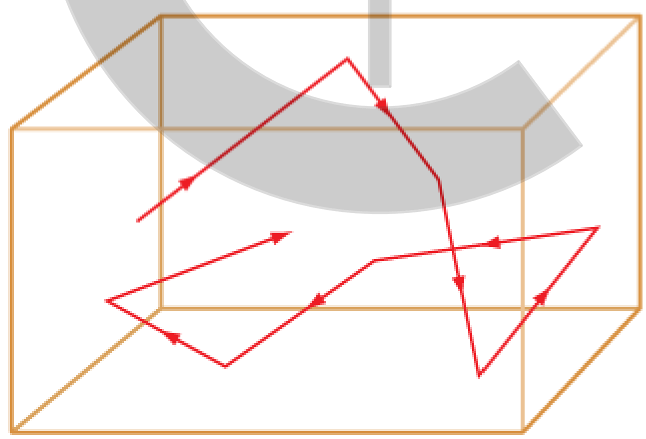


Figure 20.3: The path of a single molecule in an otherwise empty box.

Questions

1 State and explain, in terms of the kinetic model (the movement of molecules), what happens to the pressure inside a tyre when more molecules at the same temperature are pumped into the tyre.

2 Explain, using the kinetic model, why a can containing air may explode if the temperature rises.

COALI

20.3 Measuring gases

We are going to picture a container of gas, such as the box shown in Figure 20.4. There are four properties of this gas that we might measure: pressure, temperature, volume and mass. In this chapter, you will learn how these quantities are related to one another.

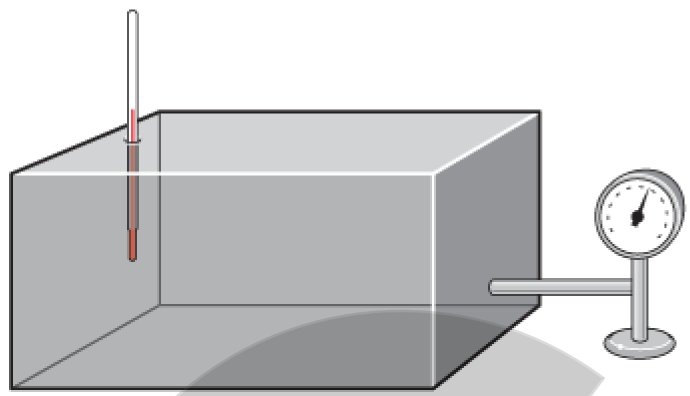


Figure 20.4: A gas has four measurable properties, which are all related to one another: pressure, temperature, volume and mass.

Pressure

This is the normal force exerted per unit area by the gas on the walls of the container. We saw in Chapter 7 that molecular collisions with the walls of the container produce a force and thus create a pressure. Pressure is measured in pascals, Pa (1 Pa = 1 N m^{-2}).

Temperature

This might be measured in °C, but in practice it is more useful to use the thermodynamic (Kelvin) scale of temperature. You should recall how these two scales are related:

$$T \text{ (K)} = \theta \text{ (°C)} + 273.15$$

Volume

This is a measure of the space occupied by the gas. Volume is measured in m^3.

Mass

This is measured in g or kg. In practice, it is more useful to consider the **amount** of gas, measured in moles. The mole is the SI unit of substance, not a unit of mass.

We have seen in Chapter 15 that each atom or molecule has a mass in unified atomic mass units (u), approximately equal to the number of nucleons (protons and neutrons) it contains.

We have also seen that 1 u = 1.66×10^{-27} kg.

Thus, each atom of carbon-12 has a mass:

$$12\text{ u} = 12 \times 1.66 \times 10^{-27}\text{ kg}$$
$$= 1.99 \times 10^{-26}\text{ kg}$$

So, 0.012 kg of carbon-12 contains $\frac{0.012}{1.99\times10^{-26}} = 6.02 \times 10^{23}$ molecules.

A mole of any substance (solid, liquid or gas) contains a standard number of particles (molecules or atoms). This number is known as the **Avogadro constant**, N_A. The value for N_A is $6.02 \times 10^{23}\text{ mol}^{-1}$. We can easily determine the number of atoms in a sample if we know how many moles it contains. For example:

2.0 mol of helium contains

$$2.0 \times 6.02 \times 10^{23} = 1.20 \times 10^{24}\text{ atoms}$$

10 mol of carbon contains

$$10 \times 6.02 \times 10^{23} = 6.02 \times 10^{24}\text{ atoms}$$

We will see later that, if we consider equal numbers of moles of two different gases under the same conditions, their physical properties are the same.

Questions

3 The mass of one atom of carbon-12 is 12 u. Determine:

- **a** the mass of one atom of carbon-12 in kg, given that 1 u = 1.66×10^{-27} kg
- **b** the number of atoms and the number of moles in 54 g of carbon
- **c** the number of atoms in 1.0 kg of carbon.

4
- **a** Calculate the mass in grams of a single atom of uranium-235 of mass 235 u.
- **b** A small pellet of uranium-235 has a mass of 20 mg. For this pellet, calculate:
  - **i** the number of uranium atoms
  - **ii** the number of moles.

5 'It can be useful to recall that 1.0 kg of ordinary matter contains in the order of 10^{26} atoms.' Making suitable estimates, test this statement.

20.4 Boyle's law

This law relates the pressure p and volume V of a gas. It was discovered in 1662 by Robert Boyle.

If a gas is compressed at constant temperature, its pressure increases and its volume decreases. A decrease in volume occupied by the gas means that there are more particles per unit volume and more collisions per second of the particles with unit area of the wall. Because the temperature is constant, the average speed of the molecules does not change. This means that each collision with the wall involves the same change in momentum, but with more collisions per second on unit area of the wall there is a greater rate of change of momentum and, therefore, a larger pressure on the wall.

Pressure and volume are inversely related.

We can write **Boyle's law** as:

The pressure exerted by a fixed mass of gas is inversely proportional to its volume, provided the temperature of the gas remains constant.

Note that this law relates two variables, pressure and volume, and it requires that the other two, mass and temperature, remain constant.

Boyle's law can be written as:

$$p \propto \frac{1}{V}$$

or simply:

$$pV = \text{Constant}$$

We can also represent Boyle's law as a graph, as shown in Figure 20.5. A graph of p against $\frac{1}{V}$ is a straight line passing through the origin, showing direct proportionality.

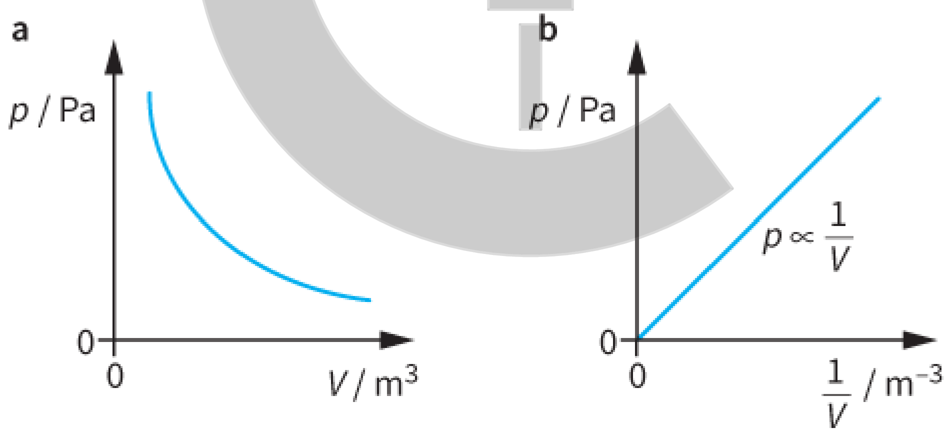


Figure 20.5: Graphical representations of the relationship between pressure and volume of a gas (Boyle's law).

For solving problems, you may find it more useful to use the equation in this form:

$$p_1V_1 = p_2V_2$$

Here, p_1 and V_1 represent the pressure and volume of the gas before a change, and p_2 and V_2 represent the pressure and volume of the gas after the change. Worked example 1 shows how to use this equation.

WORKED EXAMPLE

1 A cylinder contains 0.80 m^3 of nitrogen gas at a pressure of 1.2 atmosphere (1 atm = 1.01×10^5 Pa). A piston slowly compresses the gas to a pressure of 6.0 atm. The temperature of the gas remains constant. Calculate the final volume of the gas.

Note from the question that the temperature of the gas is constant, and that its mass is fixed (because it is contained in a cylinder). This means that we can apply Boyle's law.

Step 1 We are going to use Boyle's law in the form $p_1V_1 = p_2V_2$. Write down the quantities that you know, and that you want to find out.

$$p_1 = 1.2 \text{ atm} \quad V_1 = 0.80 \text{ m}^3$$
$$p_2 = 6.0 \text{ atm} \quad V_2 = ?$$

Note that we don't need to worry about the particular units of pressure and volume being used here, so long as they are the same on both sides of the equation. The final value of V_2 will be in dm^3 because V_1 is in m^3.

Step 2 Substitute the values in the equation, rearrange and find V_2:

$$p_1V_1 = p_2V_2$$
$$1.2 \times 0.8 = 6.0 \times V_2$$
$$V_2 = \frac{1.2 \times 0.8}{6.0}$$
$$V_2 = 0.16 \text{ m}^3$$

So the volume of the gas is reduced to 0.16 m^3.

The pressure increases by a factor of 5, so the volume decreases by a factor of 5.

Question

6 A balloon contains 0.04 m^3 of air at a pressure of 120 kPa. Calculate the pressure required to reduce its volume to 0.025 m^3 at constant temperature.

20.5 Changing temperature

Boyle's law requires that the temperature of a gas is fixed. What happens if the temperature of the gas is allowed to change? Figure 20.6 shows the results of an experiment in which a fixed mass of gas is cooled at constant pressure. The gas contracts; its volume decreases.

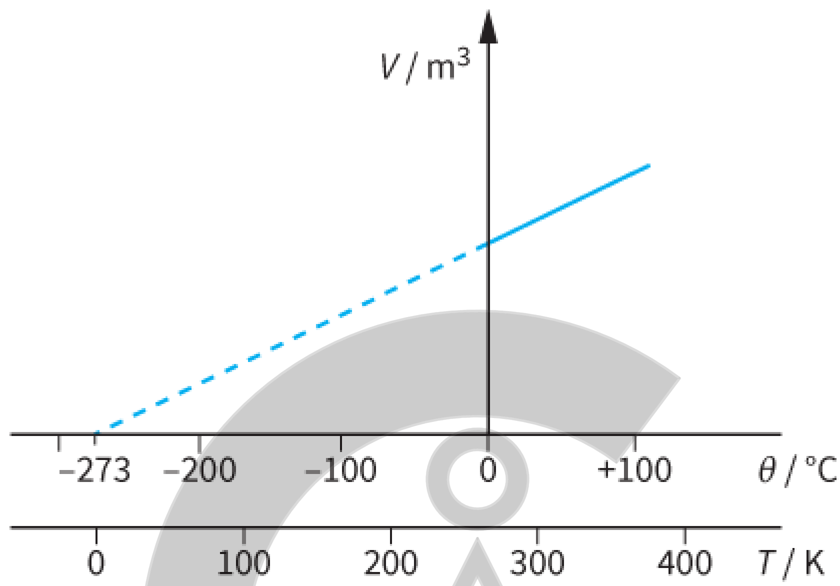


Figure 20.6: The volume of a gas decreases as its temperature decreases.

This graph does not show that the volume of a gas is proportional to its temperature on the Celsius scale. If a gas contracted to zero volume at 0 °C, the atmosphere would condense on a cold day and we would have a great deal of difficulty in breathing! However, the graph **does** show that there is a temperature at which the volume of a gas does, in principle, shrink to zero. Looking at the lower temperature scale on the graph, where temperatures are shown in kelvin (K), we can see that this temperature is 0 K, or absolute zero. (Historically, this is how the idea of absolute zero first arose.)

We can represent the relationship between volume *V* and thermodynamic temperature *T* as:

$$V \propto T$$

or simply:

$$\frac{V}{T} = \text{constant}$$

Note that this relationship only applies to a fixed mass of gas and to constant pressure.

This relationship is an expression of **Charles's law**, named after the French physicist Jacques Charles, who in 1787 experimented with different gases kept at constant pressure.

If we combine Boyle's law and Charles's law, we can arrive at a single equation for a fixed mass of gas:

$$\frac{pV}{T} = \text{constant}$$

Shortly, we will look at the constant quantity that appears in this equation, but first we will consider the extent to which this equation applies to real gases.

KEY EQUATION

$$\frac{pV}{T} = \text{constant}$$

Fixed mass of gas.

Real and ideal gases

The relationships between *p*, *V* and *T* that we have considered are based on experimental observations of gases such as air, helium, nitrogen and so on, at temperatures and pressures around room temperature and pressure. In practice, if we change to more extreme conditions, such as low temperatures or high pressures, gases start to deviate from these laws as the gas atoms exert significant electrical forces on each other. For example, Figure 20.7 shows what happens when nitrogen is cooled down towards absolute zero. At first, the graph of volume against temperature follows a good straight line. However, as it approaches the temperature at which it condenses, it deviates from ideal behaviour and at 77 K it condenses to become liquid nitrogen.

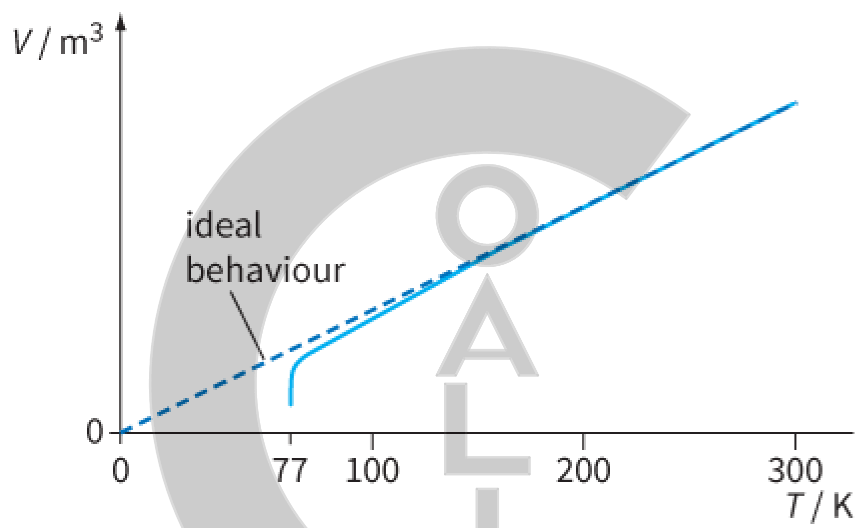


Figure 20.7: A real gas (in this case, nitrogen) deviates from the behaviour predicted by Charles's law at low temperatures.

Thus, we have to attach a condition to the relationships discussed earlier. We say that they apply to an **ideal gas**.

When we are dealing with real gases, we have to be aware that their behaviour may be significantly different from the ideal gas.

An ideal gas is thus one for which we can apply the equation:

$$\frac{pV}{T} = \text{Constant for a fixed mass of gas}$$

20.6 Ideal gas equation

So far, we have seen how p, V and T are related. It is possible to write a single equation relating these quantities that takes into account the amount of gas being considered.

We can write the equation in the following form:

$$pV = nRT$$

where n is the amount (number of moles) of an ideal gas.

Or in the form:

$$pV = NkT$$

where N is the number of molecules and k is the Boltzmann constant described later in topic 20.8.

This equation is called the **equation of state** for an ideal gas (or the **ideal gas equation)**. It relates all four of the variable quantities discussed at the beginning of this chapter. The constant of proportionality R is called the universal molar gas constant. Its experimental value is:

$$R = 8.31\ \text{J mol}^{-1}\ \text{K}^{-1}$$

Note that it doesn't matter what gas we are considering–it could be a very 'light' gas like hydrogen, or a much 'heavier' one like carbon dioxide. So long as it is behaving as an ideal gas, we can use the same equation of state with the same constant R.

KEY EQUATION

equation of state:

$$pV = nRT \text{ or } pV = NkT$$

Calculating the number *n* of moles

Instead of knowing the mass of one molecule in unified atomic mass units, sometimes we may be given the molar mass (the mass of one mole) and the mass of gas we are concerned with, to find how many moles are present. To do this, we use the relationship:

$$\text{number of moles} = \frac{\text{mass(g)}}{\text{molar mass(g mol}^{-1}\text{)}}$$

For example: How many moles are there in 1.6 kg of oxygen?

$$\text{molar mass of oxygen-16} = 32\ \text{g mol}^{-1}$$

$$\text{number of moles} = \frac{1600\ \text{g}}{32\ \text{g mol}^{-1}} = 50\ \text{mol}$$

(Note that this tells us that there are 50 moles of oxygen **molecules** in 1.6 kg of oxygen. An oxygen molecule consists of two oxygen atoms – its formula is O_2 – so 1.6 kg of oxygen contains 100 moles of oxygen **atoms**.)

Now look at Worked examples 2 and 3.

WORKED EXAMPLE

2 Calculate the volume occupied by one mole of an ideal gas at room temperature (20 °C) and pressure (1.013×10^5 Pa).

Step 1 Write down the quantities given.

$p = 1.013 \times 10^5$Pa $\quad n = 1.0$

$T = 293$ K

Hint: *Note that the temperature is converted to kelvin.*

Step 2 Substituting these values in the equation of state gives:

$$V = \frac{nRT}{P} = \frac{1 \times 8.31 \times 293}{1.103 \times 10^5} = 0.0240\ \text{m}^3 = 2.40 \times 10^{-2}\ \text{m}^3 = 24.0\ \text{dm}^3$$

Hint: *1 dm = 0.1 m; hence 1 dm^3 = 10^{-3} m^3.*

This value, the volume of one mole of gas at room temperature and pressure, is well worth remembering. It is certainly known by most chemists.

3 A car tyre contains 0.020 m^3 of air at 27 °C at a pressure of 3.0×10^5 Pa. Calculate the mass of the air in the tyre. (Molar mass of air = 28.8 g mol^{-1}.)

Step 1 Here, we need first to calculate the number of moles of air using the equation of state. We have:

$p = 3.0 \times 10^5$ Pa $\quad V = 0.02$ m^3 $\quad T = 27$ °C = 300 K

Hint: *Don't forget to convert the temperature to kelvin.*

So, from the equation of state:

$$n = \frac{pV}{RT} = \frac{30 \times 10^5 \times 0.02}{8.31 \times 300} = 2.41\ \text{mol}$$

Step 2 Now we can calculate the mass of air:

mass = number of moles × molar mass

mass = 2.41 × 28.8 = 69.4 g ≈ 69 g

Questions

For the questions that follow, you will need the following value:

$R = 8.31$ J mol^{-1} K^{-1}

7 At what temperature (in K) will 1.0 mol of a gas occupy 1.0 m^3 at a pressure of 1.0×10^4 Pa?

8 Nitrogen consists of molecules N_2. The molar mass of nitrogen is 28 g mol^{-1}. For 100 g of nitrogen, calculate:

a the number of moles

b the volume occupied at room temperature and pressure (20 °C; 1.01×10^5 Pa).

9 Calculate the volume of 5.0 mol of an ideal gas at a pressure of 1.0×10^5 Pa and a temperature of 200 °C.

10 A sample of gas contains 3.0×10^{24} molecules. Calculate the volume of the gas at a temperature of 300 K and a pressure of 120 kPa.

11 At what temperature would 1.0 kg of oxygen occupy 1.0 m^3 at a pressure of 1.0×10^5 Pa? (Molar mass of O_2 = 32 g mol^{-1}.)

12 A cylinder of hydrogen has a volume of 0.100 m^3. Its pressure is found to be 20 atmospheres at 20 °C.

a Calculate the mass of hydrogen in the cylinder.

b If it were instead filled with oxygen to the same pressure, how much oxygen would it contain?

(Molar mass of H_2 = 2.0 g mol^{-1}; molar mass of O_2 = 32 g mol^{-1};

1 atmosphere = 1.01×10^5 Pa.)

20.7 Modelling gases: the kinetic model

In this chapter, we are concentrating on the macroscopic properties of gases (pressure, volume, temperature). These can all be readily measured in the laboratory. The equation:

$$\frac{pV}{T} = \text{constant}$$

is an empirical relationship. In other words, it has been deduced from the results of experiments. It gives a good description of gases in many different situations. However, an empirical equation does not **explain** why gases behave in this way. An explanation requires us to think about the underlying nature of a gas and how this gives rise to our observations.

A gas is made of particles (atoms or molecules). Its pressure arises from collisions of the particles with the walls of the container; more frequent or harder collisions give rise to greater pressure. Its temperature indicates the average kinetic energy of its particles; the faster they move, the greater their average kinetic energy and the higher the temperature.

The **kinetic theory of gases** is a theory that links these microscopic properties of particles (atoms or molecules) to the macroscopic properties of a gas. Table 20.1 shows the assumptions on which the theory is based.

On the basis of these assumptions, it is possible to use Newtonian mechanics to show that pressure is inversely proportional to volume (Boyle's law), volume is directly proportional to thermodynamic (kelvin) temperature (Charles's law), and so on. The theory also shows that the particles of a gas have a range of speeds – some move faster than others.

Learn the four assumptions of the kinetic theory shown in Table 20.1.

Assumption	Explanation/comment
A gas contains a large number of particles (atoms or molecules) moving at random that collide elastically with the walls and with each other.	Kinetic energy cannot be lost. The internal energy of the gas is the total kinetic energy of the particles.
The forces between particles are negligible, except during collisions.	If the particles attracted each other strongly over long distances, they would all tend to clump together in the middle of the container.
The volume of the particles is negligible compared to the volume occupied by the gas.	When a liquid boils to become a gas, its particles become much farther apart.
The time of collision by a particle with the container walls is negligible compared with the time between collisions.	The molecules can be considered to be hard spheres.

Table 20.1: The basic assumptions of the kinetic theory of gases.

The kinetic theory has proved to be a very powerful model. It convinced many physicists of the existence of particles long before it was ever possible to visualise them.

Molecules in a box

We can use the kinetic model to deduce an equation that relates the macroscopic properties of a gas (pressure, volume) to the microscopic properties of its molecules (mass and speed). We start by picturing a single molecule in a cube-shaped box of side l (Figure 20.8). This molecule has mass m, and is moving with speed c parallel to one side of the box (c is not the speed of light in this case). It rattles back and forth, colliding at regular intervals with the ends of the box and thereby contributing to the pressure of the gas. We are going to work out the pressure this one molecule exerts on one end of the box and then deduce the total pressure produced by all the molecules.

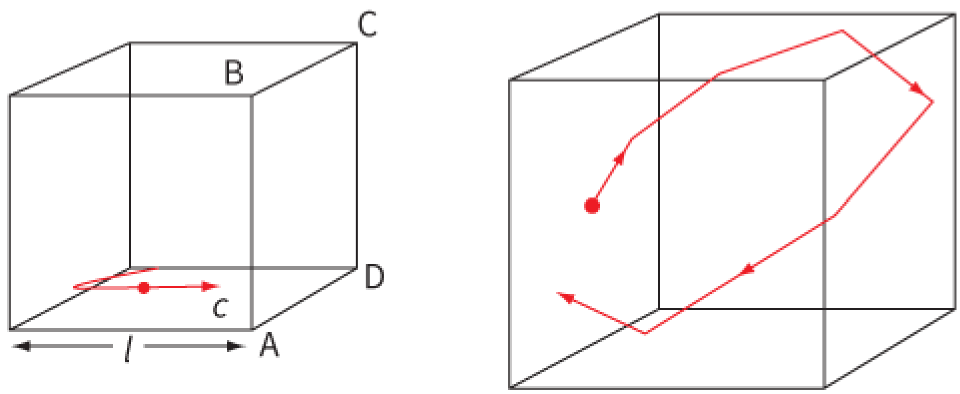


Figure 20.8: A single molecule of a gas, moving in a box.

KEY EQUATIONS

$$\text{force} = \frac{\text{change in momentum}}{\text{time taken}}$$

$$F = \frac{\Delta mv}{t}$$

$$\text{Pressure} = \frac{\text{force}}{\text{area}}$$

$$P = \frac{F}{A}$$

Note: you need to be able to derive the final equation yourself.

You need to read through the proof carefully as you will need to be able to derive the final equation yourself.

The stages involved in this calculation are:

1. Find the change in momentum as a single molecule hits a wall at 90°.
2. Calculate the number of collisions per second by the molecule on a wall.
3. Find the change in momentum per second.
4. Find the pressure on the wall.
5. Consider the effect of having three directions in which the molecule can move.

As you go through the proof, see for yourself where each stage starts and finishes.

Consider a collision in which the molecule strikes side ABCD of the cube. It rebounds elastically in the opposite direction, so that its velocity is $-c$. Its momentum changes from mc to $-mc$. The change in momentum arising from this single collision is thus:

$$\begin{aligned}\text{change in momentum} &= -mc - (+mc)\\ &= -mc - mc = -2mc\end{aligned}$$

Between consecutive collisions with side ABCD, the molecule travels a distance of $2l$ at speed c. Hence:

$$\text{time between collisions with side ABCD} = \frac{2l}{c}$$

Now we can find the force that this one molecule exerts on side ABCD, using Newton's second law of motion. This says that the force produced is equal to the rate of change of momentum:

$$\text{force} = \frac{\text{change in momentum}}{\text{time taken}} = \frac{2mc}{\left(\frac{2l}{c}\right)} = \frac{mc^2}{l}$$

(We use $+2mc$ because now we are considering the force of the molecule on side ABCD, which is in the opposite direction to the change in momentum of the molecule.)

The area of side ABCD is l^2. From the definition of pressure, we have:

$$\text{pressure } p = \frac{\text{force}}{\text{area}} = \frac{\left(\frac{mc^2}{l}\right)}{l^2} = \frac{mc^2}{l^3}$$

This is for one molecule, but there is a large number N of molecules in the box. Each has a different velocity, and each contributes to the pressure. We write the average value of c^2 as $<c^2>$, and multiply by N to find the total pressure:

$$p = \frac{Nm<c^2>}{l^3}$$

But this assumes that all the molecules are travelling in the same direction and colliding with the same pair of opposite faces of the cube. In fact, they will be moving in all three dimensions equally.

If there are components c_x, c_y and c_z of the velocity in the x-, y- and z- directions then $c^2 = c_x^2 + c_y^2 + c_z^2$. There is nothing special about any particular direction and so $<c_x^2> = <c_y^2> = <c_z^2>$ and $< c_x^2 > = \frac{1}{3} < c^2 >$

The equation for pressure worked out above involved just the component of the velocity in the x-direction and if c is the actual speed of the particle then we need to divide by 3 to find the pressure exerted.

$$p = \frac{1}{3}\left(\frac{Nm<c^2>}{l^3}\right)$$

Here, l^3 is equal to the volume V of the cube, so we can write:

$$p = \frac{1}{3}\left(\frac{Nm}{V}\right) < c^2 > \quad \text{or} \quad pV = \frac{1}{3}Nm < c^2 >$$

(Notice that, in the second form of the equation, we have the macroscopic properties of the gas – pressure and volume – on one side of the equation and the microscopic properties of the molecules on the other side.)

KEY EQUATION

Pressure of an ideal gas:

$$p = \frac{1}{3}\left(\frac{Nm}{V}\right) < c^2 > \quad \text{or} \quad pV = \frac{1}{3}Nm < c^2 >$$

Finally, the quantity Nm is the mass of all the molecules of the gas, and this is simply equal to the mass M of the gas. So $\frac{Nm}{V}$ is equal to the density ρ of the gas, and we can write:

$$p = \frac{1}{3}\rho < c^2 >$$

So the pressure of a gas depends only on its density and the mean square speed of its molecules.

A plausible equation?

It is worth thinking a little about whether the equation $p = \frac{1}{3}(\frac{Nm}{V}) < c^2 >$ seems to make sense. It should be clear to you that the pressure is proportional to the number of molecules, *N*. More molecules mean greater pressure. Also, the greater the mass of each molecule, the greater the force it will exert during a collision.

The equation also suggests that pressure *p* is proportional to the average value of the speed squared. This is because, if a molecule is moving faster, not only does it strike the container harder, but it also strikes the container more often.

The equation suggests that the pressure *p* is inversely proportional to the volume occupied by the gas. Here, we have deduced Boyle's law. If we think in terms of the kinetic model, we can see that if a mass of gas occupies a larger volume, the frequency of collision between the molecules and unit area of wall decreases. The equation shows us not just that pressure will be lower but that it is inversely proportional to volume.

These arguments should serve to convince you that the equation is plausible; this sort of argument cannot prove the equation.

Questions

13 Check that the SI base units on the left-hand side of the equation:

$$p = \frac{1}{3}(\frac{Nm}{V}) < c^2 >$$

are the same as those on the right-hand side.

14 The quantity *Nm* is the total mass of the molecules of the gas, i.e. the mass of the gas. At room temperature, the density of air is about 1.29 kg m^{-3} at a pressure of 10^5 Pa.

- **a** Use these figures to deduce the value of $<c^2>$ for air molecules at room temperature.
- **b** Find a typical value for the speed of a molecule in the air by calculating $\sqrt{< c^2 >}$. How does this compare with the speed of sound in air, approximately 330 m s^{-1}?

20.8 Temperature and molecular kinetic energy

Now we can compare the equation $p = \frac{1}{3}\left(\frac{Nm}{V}\right) < c^2 >$ with the ideal gas equation $pV = nRT$. The left-hand sides are the same, so the two right-hand sides must also be equal:

$$\frac{1}{3}Nm < c^2 >= nRT$$

We can use this equation to tell us how the absolute temperature of a gas (a macroscopic property) is related to the mass and speed of its molecules. If we focus on the quantities of interest, we can see the following relationship:

$$m < c^2 >= \frac{3nRT}{N}$$

The quantity $\frac{N}{n} = N_A$ is the Avogadro constant, the number of particles in 1 mole. So:

$$m < c^2 >= \frac{3RT}{N_A}$$

It is easier to make sense of this if we divide both sides by 2, to get the familiar expression for kinetic energy:

$$\frac{1}{2}m < c^2 >= \frac{3RT}{2N_A}$$

The quantity $\frac{R}{N_A}$ is defined as the **Boltzmann constant**, k. Its value is 1.38×10^{-23} J K^{-1}. Substituting k in place of $\frac{R}{N_A}$ gives

$$\text{kinetic energy} = \frac{1}{2}m < c^2 >= \frac{3}{2}kT$$

This is the average kinetic energy E of a molecule in the gas, and since k is a constant, the thermodynamic temperature T is proportional to the average kinetic energy of a molecule.

KEY EQUATION

Boltzmann constant:

$$k = \frac{R}{N_A}$$

The mean translational kinetic energy of an atom (or molecule) of an ideal gas is proportional to the thermodynamic temperature.

It is easier to recall this as:

mean translational kinetic energy of atom $\propto T$

KEY EQUATION

mean translational kinetic energy of atom $\propto T$

We need to consider two of the terms in this statement. First, we talk about **translational** kinetic energy. This is the energy that the molecule has because it is moving from one point in space to another; a molecule made of two or more atoms may also spin or tumble around, and is then said to have rotational kinetic energy – see Figure 20.9.

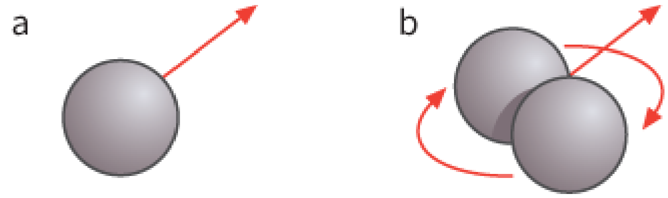


Figure 20.9: a A monatomic molecule has only translational kinetic energy. **b** A diatomic molecule can have both translational and rotational kinetic energy.

Second, we talk about **mean** (or average) kinetic energy. There are two ways to find the average kinetic energy (k.e.) of a molecule of a gas. Add up all the kinetic energies of the individual molecules of the gas and then calculate the average k.e. per molecule. Alternatively, watch an individual molecule over a period of time as it moves about, colliding with other molecules and the walls of the container and calculate its average k.e. over this time. Both should give the same answer.

The Boltzmann constant is an important constant in physics because it tells us how a property of microscopic particles (the kinetic energy of gas molecules) is related to a macroscopic property of the gas (its absolute temperature). That is why its units are joules per kelvin ($J\,K^{-1}$). Its value is very small ($1.38 \times 10^{-23}\,J\,K^{-1}$) because the increase in kinetic energy in J of a molecule is very small for each kelvin increase in temperature.

It is useful to remember the equation linking kinetic energy with temperature as 'average k.e. is three-halves *kT*'.

KEY EQUATION

$$\text{kinetic energy (of a molecule)} = \frac{3}{2}kT$$

Questions

15 The Boltzmann constant k is equal to $\frac{R}{N_A}$. From values of R and N_A, show that k has the value $1.38 \times 10^{-23}\,J\,K^{-1}$.

16 Calculate the mean translational k.e. of atoms in an ideal gas at 27 °C.

17 The atoms in a gas have a mean translational k.e. equal to 5.0×10^{-21} J. Calculate the temperature of the gas in K and in °C.

The root-mean-square speed

You may have wondered how the mean-square speed $< c^2 >$ compares with the mean speed $< c >$.

The exact relationship depends on the distribution of the speeds of the molecules. If all the molecules have the same speed, then $< c > = \sqrt{< c^2 >}$

But is this always the case?

Imagine three molecules with speeds 10, 20 and 30 $m\,s^{-1}$; (very low speeds for molecules, but easier for our calculations!).

$$\text{Their mean speed } < c > = \frac{(10+20+30)}{3} = 20\ m\,s^{-1}$$

Their square speeds are 10^2, 20^2 and 30^2.

So, their mean-square speed

$$< c^2 > = \frac{(10^2+20^2+30^2)}{3} = 467\ m^2s^{-2}$$

In this case, $\sqrt{<c^2>} = 22\ \text{m s}^{-1}$, which is **not** the same as the mean speed.

Similarly, the mean of the square of the speeds $<c^2> = 467\ \text{m}^2\ \text{s}^{-2}$ is **not** the same as the square of the mean of the speeds $(<c>)^2 = 400\ \text{m}^2\ \text{s}^{-2}$ in the example.

In general, the values for $<c>$ and $\sqrt{<c^2>}$ are similar but, because they are **not** the same, we define a special quantity called the **root-mean-square speed** $c_{\text{r.m.s.}}$.

This is the square root of the mean-square-speed; that is:

$$c_{\text{r.m.s}} = \sqrt{<c^2>}$$

In the example, for the three molecules, $c_{\text{r.m.s}} = 22\ \text{m s}^{-1}$.

KEY EQUATION

$$c_{\text{r.m.s}} = \sqrt{<c^2>}$$

Root-mean-square speed, where $c_{\text{r.m.s}}$ is the root of the mean square speed.

Questions

18 Four molecules have speeds 200, 400, 600 and 800 m s^{-1}. Calculate:

- **a** their mean speed $<c>$
- **b** the square of their mean speed $<c>^2$
- **c** their mean-square speed $<c^2>$
- **d** their root-mean-square speed $c_{\text{r.m.s.}}$.

19 Calculate the root-mean square speed of the molecules of hydrogen at 20 °C given that each molecule of hydrogen has mass 3.35×10^{-27} kg.

Mass, kinetic energy and temperature

Since mean k.e. $\propto T$, it follows that if we double the thermodynamic temperature of an ideal gas (for example, from 300 K to 600 K), we double the mean k.e. of its molecules. It doesn't follow that we have doubled their speed; because k.e. $\propto v^2$, their mean speed has increased by a factor of $\sqrt{2}$.

Air is a mixture of several gases: nitrogen, oxygen, carbon dioxide, etc. In a sample of air, the mean k.e. of the nitrogen molecules is the same as that of the oxygen molecules and that of the carbon dioxide molecules. This comes about because they are all repeatedly colliding with one another, sharing their energy. Carbon dioxide molecules have greater mass than oxygen molecules; since their mean translational k.e. is the same, it follows that the carbon dioxide molecules move more slowly than the oxygen molecules.

Questions

20 Show that, if the mean speed of the molecules in an ideal gas is doubled, the thermodynamic temperature of the gas increases by a factor of four.

21 A fixed mass of gas expands to twice its original volume at a constant temperature. How do the following change?

- **a** the pressure of the gas
- **b** the mean translational kinetic energy of its molecules.

22 Air consists of molecules of oxygen (molar mass = 32 g mol^{-1}) and nitrogen (molar mass = 28 g mol^{-1}). Calculate the mean translational k.e. of these molecules in air at 20 °C. Use your answer to calculate the root-mean-square speed of each type of molecule.

23 Show that the change in the internal energy of one mole of an ideal gas per unit change in temperature is always a constant. What is this constant?

REFLECTION

Without looking at your textbook, make a list of the kinetic theory equations and write down what each term in the equations means.

Write out a proof on your own of the main kinetic theory equation using momentum change and Newton's laws.

Write out the assumptions in your own words.

Show how kinetic theory relates temperature and molecular speed.

What things might you want more help with?

SUMMARY

For an ideal gas:

$$\frac{pV}{T} = \text{constant}$$

One mole of any substance contains N_A particles (atoms or molecules):

$$N_A = \text{Avogadro constant} = 6.02 \times 10^{23}\ \text{mol}^{-1}$$

The equation of state for an ideal gas is:

$pV = nRT$ for n moles. $pV = NRT$ for N molecules

There are four assumptions of the kinetic theory:

1. Molecules move at random, colliding elastically with the walls.
2. The volume of the molecules is small compared to the volume of the container.
3. There are no forces between atoms in the gas.
4. The time of each collision is small compared to the time between collisions.

From the kinetic model of a gas, we can deduce the relationship:

$pV = \frac{1}{3}Nm < c^2 >$ where $< c^2 >$ is the mean-square speed of the molecules.

The mean translational kinetic energy E of a particle (atom or molecule) of an ideal gas is proportional to the thermodynamic temperature T:

$$E = \frac{1}{2}m < c^2 > = \frac{3}{2}kT$$

The root-mean-sqare speed is the square root of the mean square speed of the molecules:

$$c_{r.m.s} = \sqrt{< c^2 >}$$

EXAM-STYLE QUESTIONS

1 A gas is enclosed inside a cylinder that is fitted with a freely moving piston.

The gas is initially in equilibrium with a volume V_1 and a pressure p. The gas is then cooled slowly. The piston moves into the cylinder until the volume of the gas is reduced to V_2 and the pressure remains at p.

What is the work done on the gas during this cooling? **[1]**

A $\frac{1}{2}p(V_2 - V_1)$

B $p(V_2 - V_1)$

C $\frac{1}{2}p(V_2 + V_1)$

D $p(V_2 + V_1)$

2 An ideal gas is made to expand slowly at a constant temperature.

Which statement is correct? **[1]**

A The heat energy transferred to the gas is zero.

B The internal energy of the gas increases.

C The work done by the gas is equal to the heat energy added to it.

D The work done by the gas is zero.

3 **a** State how many atoms there are in:

i a mole of helium gas (a molecule of helium has one atom) **[1]**

ii a mole of chlorine gas (a molecule of chlorine has two atoms) **[1]**

iii a kilomole of neon gas (a molecule of neon has one atom). **[1]**

b A container holds four moles of carbon dioxide of molecular formula CO_2. Calculate:

i the number of carbon dioxide molecules there are in the container **[1]**

ii the number of carbon atoms there are in the container **[1]**

iii the number of oxygen atoms there are in the container. **[1]**

[Total: 6]

4 A bar of gold-197 has a mass of 1.0 kg. Calculate:

a the mass of one gold atom in kg. **[1]**

b the number of gold atoms in the bar **[1]**

c the number of moles of gold in the bar. **[2]**

(An atom of gold contains 197 nucleons and has a mass of 197 u.)

[Total: 4]

5 A cylinder holds 140 m^3 of nitrogen at room temperature and pressure. Moving slowly, so that there is no change in temperature, a piston is pushed to reduce the volume of the nitrogen to 42 m^3.

a Calculate the pressure of the nitrogen after compression. **[2]**

b Explain the effect on the temperature and pressure of the nitrogen if the piston is pushed in very quickly. **[1]**

[Total: 3]

6 The atmospheric pressure is 100 kPa, equivalent to the pressure exerted by a **[4]**

column of water 10 m high. A bubble of oxygen of volume 0.42 cm^3 is released by a water plant at a depth of 25 m. Calculate the volume of the bubble when it reaches the surface. State any assumptions you make.

7 A cylinder contains 4.0×10^{-2} m^3 of carbon dioxide at a pressure of 4.8×10^5 Pa at room temperature.

Calculate:

a the number of moles of carbon dioxide **[2]**

b the mass of carbon dioxide. **[2]**

(Molar mass of carbon dioxide = 44 g or one molecule of carbon dioxide has mass 44 u.)

[Total: 4]

8 Calculate the volume of 1 mole of ideal gas at a pressure of 1.01×10^5 Pa and at a temperature of 0 °C. **[2]**

9 A vessel of volume 0.20 m^3 contains 3.0×10^{26} molecules of gas at a temperature of 127 °C. Calculate the pressure exerted by the gas on the vessel walls. **[3]**

10 a Calculate the root-mean-square speed of helium molecules at room temperature and pressure. (Density of helium at room temperature and pressure = 0.179 kg m^{-3}.) **[3]**

b Compare this speed with the average speed of air molecules at the same temperature and pressure. **[2]**

[Total: 5]

11 A sample of neon is contained in a cylinder at 27 °C. Its temperature is raised to 243 °C.

a Calculate the kinetic energy of the neon atoms at:

i 27 °C **[3]**

ii 243 °C. **[1]**

b Calculate the ratio of the speeds of the molecules at the two temperatures. **[2]**

[Total: 6]

12 A truck is to cross the Sahara desert. The journey begins just before dawn when the temperature is 3 °C. The volume of air held in each tyre is 1.50 m^3 and the pressure in the tyres is 3.42×10^5 Pa.

a Explain how the air molecules in the tyre exert a pressure on the tyre walls. **[3]**

b Calculate the number of moles of air in the tyre. **[3]**

c By midday the temperature has risen to 42 °C.

i Calculate the pressure in the tyre at this new temperature. You may assume that no air escapes and the volume of the tyre is unchanged. **[2]**

ii Calculate the increase in the average translational kinetic energy of an air molecule due to this temperature rise. **[2]**

[Total: 10]

13 The ideal gas equation is $pV = \frac{1}{3}Nm\langle c^2\rangle$.

a State the meaning of the symbols *N*, *m* and $\langle c^2\rangle$. **[3]**

b A cylinder of helium-4 contains gas with volume 4.1×10^4 cm^3 at a pressure of 6.0×10^5 Pa and a temperature of 22 °C. You may assume helium acts as an ideal gas and that a molecule of helium-4 contains 4 nucleons, each of mass 1.66×10^{-27} kg.

Determine:

i the amount of gas in mol **[3]**

ii the number of molecules present in the gas **[2]**

iii the root-mean-square speed of the molecules. **[3]**

[Total: 11]

14 a State what is meant by an ideal gas. **[2]**

b A cylinder contains 500 g of helium-4 at a pressure of 5.0×10^5 Pa and at a temperature of 27 °C. You may assume that the molar mass of helium-4 is 4.0 g.

Calculate:

i the number of moles of helium the cylinder holds **[1]**

ii the number of molecules of helium the cylinder holds **[1]**

iii the volume of the cylinder. **[3]**

[Total: 7]

15 a One assumption of the kinetic theory of gases is that molecules undergo perfectly elastic collisions with the walls of their container.

i Explain what is meant by a perfectly elastic collision. **[1]**

ii State three other assumptions of the kinetic theory. **[3]**

b A single molecule is contained within a cubical box of side length 0.30 m. The molecule, of mass 2.4×10^{-26} kg, moves backwards and forwards parallel to one side of the box with a speed of 400 m s^{-1}. It collides elastically with one of the faces of the box, face P.

i Calculate the change in momentum each time the molecule hits face P. **[2]**

ii Calculate the number of collisions made by the molecule in 1.0 s with face P. **[2]**

iii Calculate the mean force exerted by the molecule on face P. **[2]**

[Total: 10]

16 a A cylinder contains 1.0 mol of an ideal gas. The gas is heated while the volume of the cylinder remains constant. Calculate the energy required to raise the temperature of the gas by 1.0 °C. **[2]**

b Calculate the root-mean-square speed of a molecule of hydrogen-1 at a temperature of 100 °C.

(Mass of a hydrogen molecule = 3.34×10^{-27} kg.) **[3]**

c Calculate, for oxygen and hydrogen at the same temperature, the ratio

$$\frac{\text{root mean square speed of a hydrogen molecule}}{\text{root mean square speed of an oxygen molecule}}$$

(Mass of an oxygen molecule = 5.31×10^{-26} kg.) **[2]**

[Total: 7]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
use molar quantities and understand that one mole is an amount of substance containing N_A particles, where N_A is the Avogadro constant	20.3			
understand that an ideal gas obeys $PV \propto T$ where T is the thermodynamic temperature	20.4			
recall and use the equation of state for an ideal gas expressed as $pV = nRT$, where n = amount of substance (number of moles), and as $pV = NkT$, where N = number of molecules	20.6			
state the basic assumptions of the kinetic theory of gas	20.7			
explain how molecular movement causes the pressure exerted by a gas and derive the relationship: $pV = \frac{1}{3}Nm\langle c^2\rangle$, where $\langle c^2\rangle$ is the mean-square speed	20.7			
understand that the root-mean-square speed $c_{r.m.s.}$ is given by: $\sqrt{\langle c^2\rangle}$	20.8			
recall that the Boltzmann constant k is given by: $k = \frac{R}{N_A}$	20.8			
compare with $pV = \frac{1}{3}Nm\langle c^2\rangle$ with $pV = NkT$ to deduce that the average translational kinetic energy of a molecule is $\frac{3}{2}kT$	20.8			

〉Chapter 21

Uniform electric fields

LEARNING INTENTIONS

In this chapter you will learn how to:

- show an understanding of the concept of an electric field
- define electric field strength
- draw field lines to represent an electric field
- calculate the strength of a uniform electric field
- calculate the force on a charge in a uniform electric field
- describe how charged particles move in a uniform electric field.

BEFORE YOU START

- You will have learned about electrostatics in your previous studies and in everyday life. You will also have met the idea of magnetism.
- Make a list of the similarities between electrostatics and magnetism and also a list of the differences.
- Are the two phenomena related or not? If so, how?

ELECTRICITY IN NATURE

The lower surface of a thundercloud is usually negatively charged. When lightning strikes (Figure 21.1), an intense electric current is sent down to the ground below. You may have noticed a ‘strobe’ effect – this is because each lightning strike usually consists of four or five flashes at intervals of 50 milliseconds or so. You will already know a bit about electric (or electrostatic) fields, from your experience of static electricity in everyday life and from your studies in science. In this chapter, you will learn how we can make these ideas more formal. We will look at how electric forces are caused, and how we can represent their effects in terms of electric fields. Then we will find mathematical ways of calculating electric forces and field strengths.

There are about three million lightning strikes on the Earth each day; the energy transferred in one strike is 10 MJ. There is more than enough energy to satisfy the industrial world with all its energy needs. Why is it not harnessed? What problems can you see in harnessing it?

Figure 21.1: Lightning flashes; dramatic evidence of natural electric fields.

21.1 Attraction and repulsion

Static electricity can be useful – it is important in the process of photocopying, in dust precipitation to clean up industrial emissions and in crop-spraying, among many other applications. It can also be a nuisance. Who hasn't experienced a shock, perhaps when getting out of a car or when touching a door handle? Static electric charge has built up and gives us a shock when it discharges.

We explain these effects in terms of **electric charge**. Simple observations in the laboratory give us the following picture:

- Objects are usually electrically neutral (uncharged), but they may become electrically charged, for example, when one material is rubbed against another.
- There are two types of charge, which we call positive and negative.
- Opposite types of charge attract one another; like charges repel (Figure 21.2).
- A charged object may also be able to attract an uncharged one; this is a result of electrostatic induction.

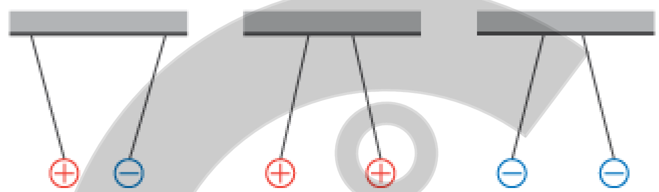


Figure 21.2: Attraction and repulsion between electric charges.

These observations are macroscopic. They are descriptions of phenomena that we can observe in the laboratory, without having to consider what is happening on the microscopic scale, at the level of particles such as atoms and electrons. However, we can give a more subtle explanation if we consider the microscopic picture of static electricity.

Using a simple model, we can consider matter to be made up of three types of particles: electrons (which have negative charge), protons (positive) and neutrons (neutral). An uncharged object has equal numbers of protons and electrons, whose charges therefore cancel out.

When one material is rubbed against another, there is friction between them, and electrons may be rubbed off one material onto the other (Figure 21.3). The material that has gained electrons is now negatively charged, and the other material is positively charged.

If a positively charged object is brought close to an uncharged one, the electrons in the second object may be attracted. We observe this as a force of attraction between the two objects. (This is known as electrostatic induction.)

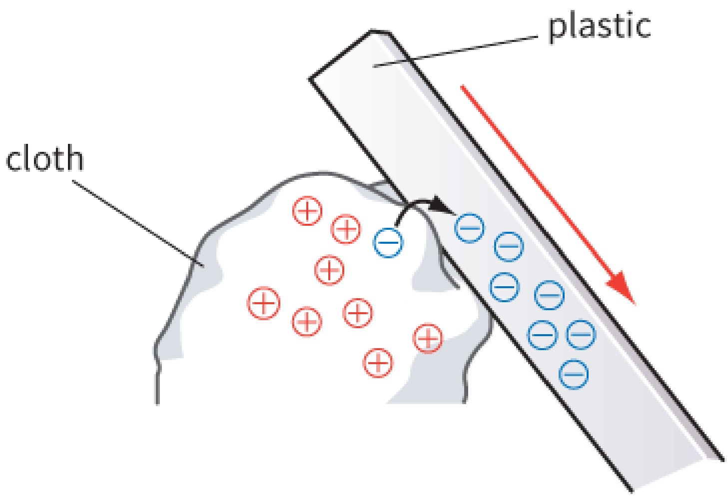


Figure 21.3: Friction can transfer electrons between materials.

It is important to appreciate that it is usually electrons that are involved in moving within a material, or from one material to another. This is because electrons, which are on the outside of atoms, are less strongly held within a material than are protons. They may be free to move about within a material (like the conduction electrons in a metal), or they may be relatively weakly bound within atoms.

PRACTICAL ACTIVITY 21.1

Investigating electric fields

If you rub a strip of plastic so that it becomes charged and then hold it close to your hair, you feel your hair being pulled upwards. The influence of the charged plastic spreads into the space around it; we say that there is an **electric field** around the charge. To produce an electric field, we need unbalanced charges (as with the charged plastic). To observe the field, we need to put something in it that will respond to the field (as your hair responds). There are two simple ways in which you can do this in the laboratory. The first uses a charged strip of gold foil, attached to an insulating handle (Figure 21.4). The second uses grains of a material such as semolina; these line up in an electric field (Figure 21.5), rather like the way in which iron filings line up in a magnetic field (Figure 21.5).

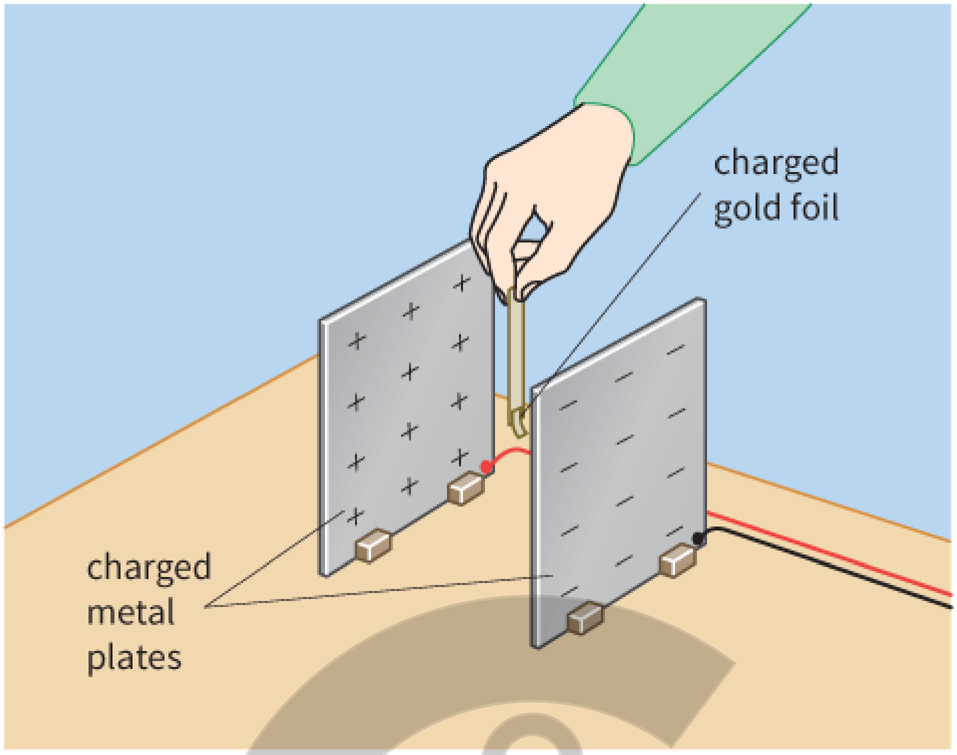


Figure 21.4: Investigating the electric field between two charged metal plates.

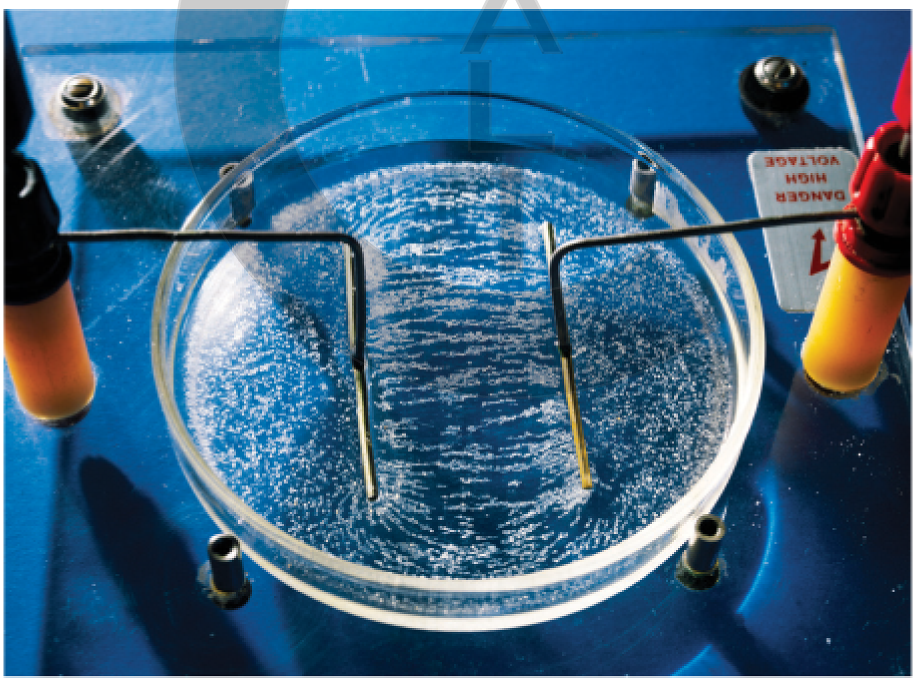


Figure 21.5: Apparatus showing a uniform electric field between two parallel charged plates.

21.2 The concept of an electric field

A charged object experiences a force in an electric field. This is what an electric field is. We say that there is an electric field anywhere where an electric charge experiences a force. An electric field is a **field of force**.

This is a rather abstract idea. You will be more familiar with the idea of a 'field of force' from your experience of magnets. There is a magnetic field around a permanent magnet; another magnet placed nearby will experience a force. You have probably plotted the field lines used to represent the field around a magnet. There is a third type of force field that we are all familiar with, because we live in it. You have already met this force in Chapter 17, the gravitational field. There are many similarities between electric fields and gravitational fields, there are also key differences.

To summarise:

- electric fields – act on objects with electric charge
- magnetic fields – act on magnetic materials, magnets and moving charges (including electric currents)
- gravitational fields – act on objects with mass.

Later, we will see that the electric force and the magnetic force are closely linked. They are generally considered as a single entity, the electromagnetic force.

Representing electric fields

We can draw electric fields in much the same way that we can draw magnetic fields or gravitational fields, by showing **field lines** (sometimes called lines of force). The three most important field shapes are shown in Figure 21.6.

As with magnetic fields, this representation tells us two things about the field: its direction (from the direction of the lines), and how strong it is (from the separation of the lines). The arrows go from positive to negative; they tell us the direction of the force on a positive charge in the field.

- A uniform field has the same strength at all points. Example: the electric field between oppositely charged parallel plates.
- A radial field spreads outwards in all directions. Example: the electric field around a point charge or a charged sphere.

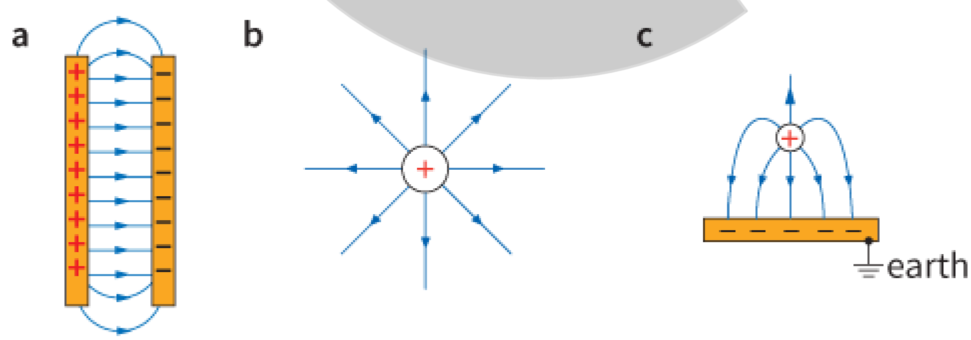


Figure 21.6: Field lines are drawn to represent an electric field. They show the direction of the force on a positive charge placed at a point in the field. **a** A uniform electric field is produced between two oppositely charged plates. **b** A radial electric field surrounds a charged sphere. **c** The electric field between a charged sphere and an earthed plate.

We can draw electric fields for other arrangements. Note the symbol for an earth, which is assumed to be uncharged (in other words, at zero volts).

Questions

1 Which of the three field diagrams in Figure 21.7 represents:

a two positive charges repelling each other?

b two negative charges?

c two opposite charges?

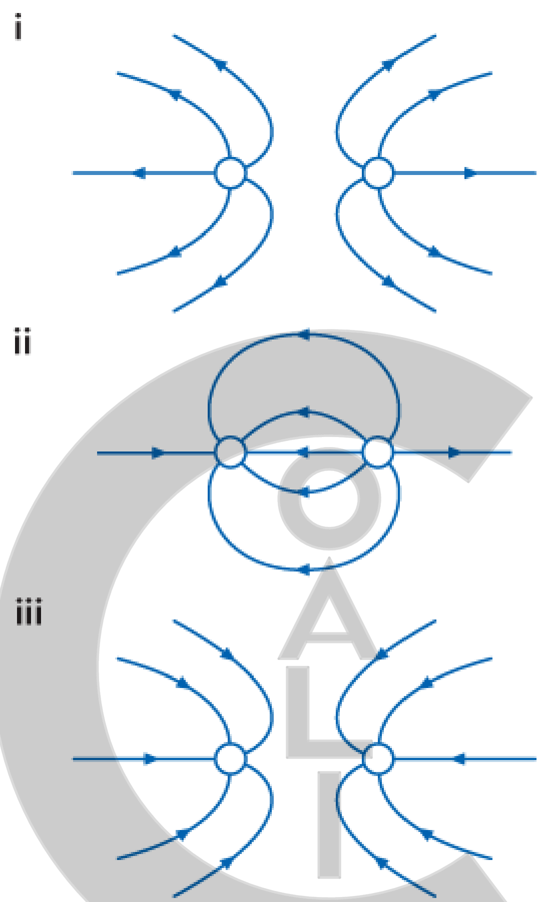


Figure 21.7: Electric fields between charges. For Question 1.

2 Many molecules are described as polar; that is, they have regions that are positively or negatively charged, though they are neutral overall. Draw a diagram to show how sausage-shaped polar molecules like those shown in Figure 21.8 might realign themselves in a solid.

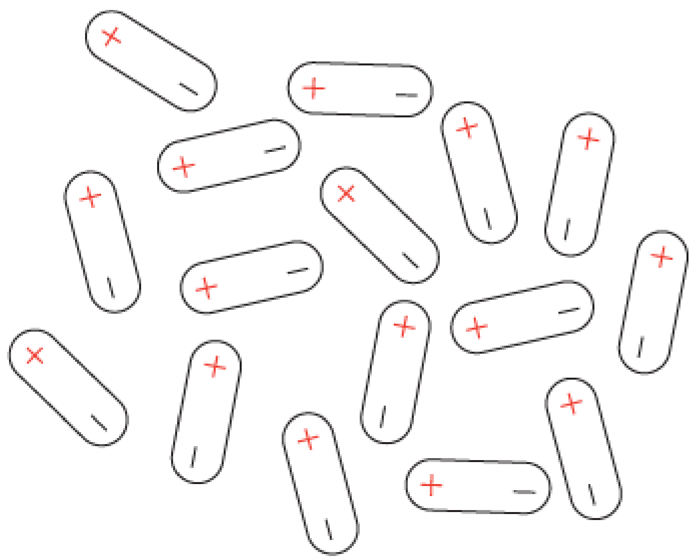


Figure 21.8: Polar molecules. For Question 2.

3 Figure 21.9 shows the electric field pattern between a thundercloud and a building. State and explain where the electric field strength is greatest.

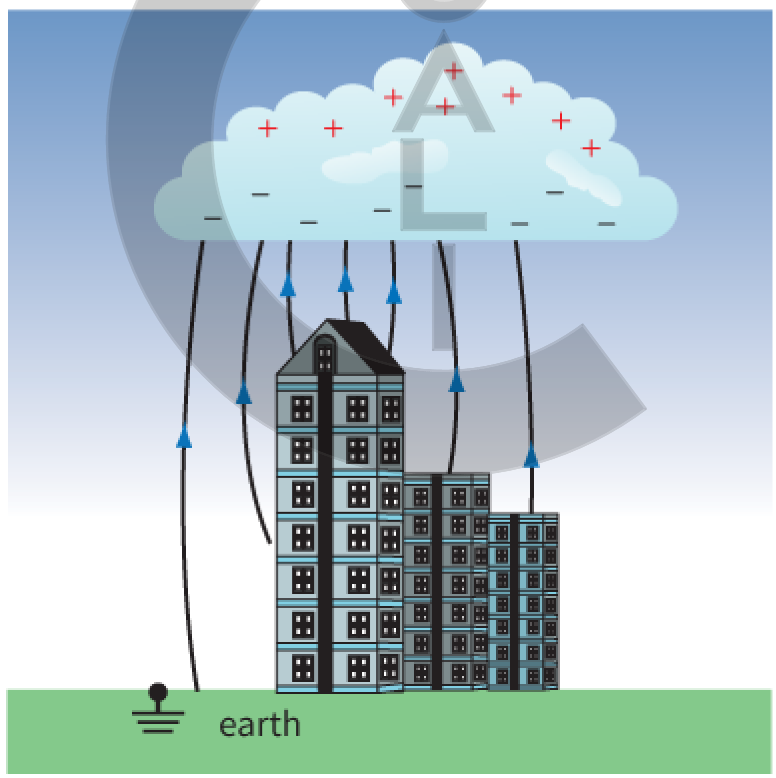


Figure 21.9: Predict where the electric field will be strongest – that's where lightning may strike.

21.3 Electric field strength

The **electric field strength** at a point is defined as the force per unit charge exerted on a stationary positive charge at that point.

To define electric field strength, we imagine putting a positive test charge $+Q$ in the field and measuring the electric force F that it feels (Figure 21.10). It is important to recognise the importance of using a positive test charge, as this determines the direction of an electric field. (If you have used a charged gold leaf to investigate a field, this illustrates the principle of testing the field with a charge.)

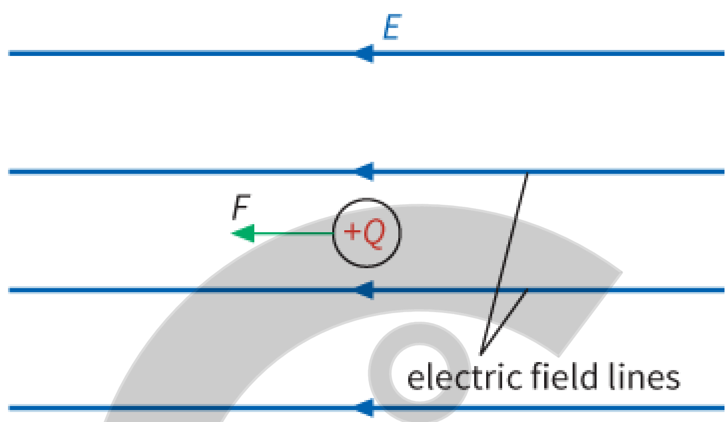


Figure 21.10: A field of strength E exerts force F on charge $+Q$.

From this definition, we can write an equation for E:

$$E = \frac{F}{Q}$$

where E is the electric field strength, F is the force on the charge and Q is the charge.

It follows that the units of electric field strength are newtons per coulomb ($N\ C^{-1}$).

KEY EQUATION

$$E = \frac{F}{Q}$$

The strength of a uniform field

You can set up a uniform field between two parallel metal plates by connecting them to the terminals of a high-voltage power supply (Figure 21.11). The strength of the field between them depends on two factors:

- the voltage V between the plates – the higher the voltage, the stronger the field: $E \propto V$
- the separation d between the plates – the greater their separation, the weaker the field: $E \propto \frac{1}{d}$

These factors can be combined to give an equation for E:

$$E = -\frac{V}{d}$$

Worked example 1 shows a derivation of this. Note that the minus sign is often omitted because we are only interested in the magnitude of the field, not its direction. In Figure 21.11, the voltage *V* increases towards the right while the force *F* acts in the opposite direction, towards the left. *E* is a vector quantity.

KEY EQUATION

$$E = \frac{\Delta V}{\Delta d}$$

Strength of a uniform field between two parallel metal plates.

If we look at this formula in a little more detail we can see that the electric field is really equal to the change in the potential (potential difference) divided by the change in distance (distance moved). This is written:

$$E = -\frac{\Delta V}{\Delta d}$$

where the symbol Δ means 'change of'.

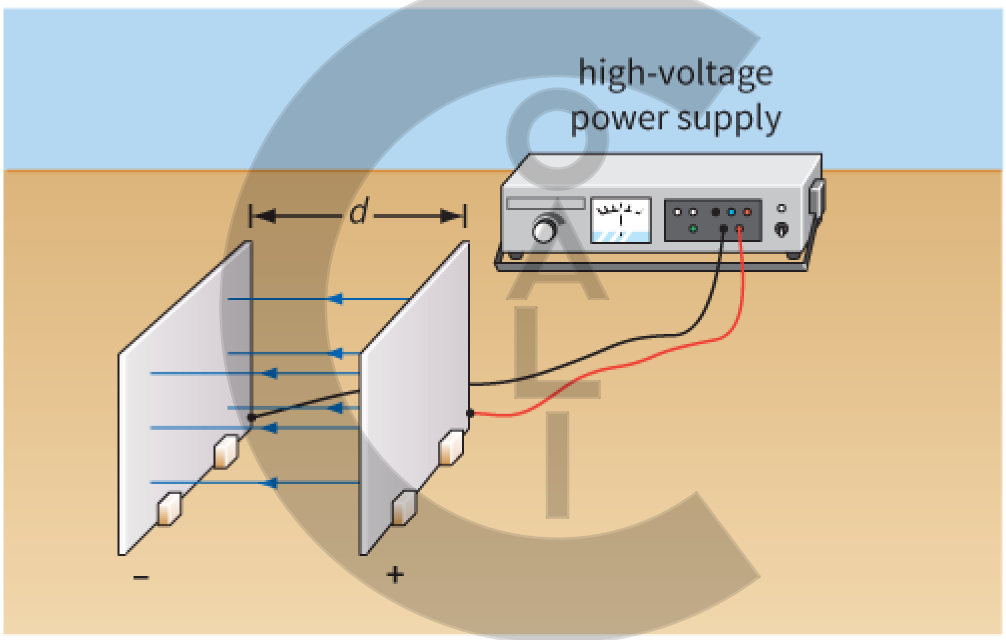


Figure 21.11: There is a uniform field between two parallel, charged plates.

From this equation, we can see that we can write the units of electric field strength as volts per metre (V m^{-1}). Note:

$$1 \text{ V m}^{-1} = 1 \text{ N C}^{-1}$$

Worked example 2 shows how to solve problems involving uniform fields.

WORKED EXAMPLES

1 Two metal plates are separated by a distance *d*. The potential difference between the plates is *V*. A positive charge *Q* is pulled at a constant speed with a constant force *F* from the negative plate all the way to the positive plate. Using the definition for electric field strength and the concept of work done, show that the magnitude of the electric field strength *E* is given by the equation:

$$E = \frac{V}{d}$$

Step 1 We have:

work done on charge = energy transformed

From their definitions, we can write:

work done = force × distance or $W = Fd$

energy transformed = VQ

Step 2 Substituting gives:

$Fd = VQ$

and rearranging gives:

$$\frac{F}{Q} = \frac{V}{d}$$

Step 3 The left-hand side of the equation is the electric field strength E. Hence:

$$E = \frac{V}{d}$$

2 Two parallel metal plates separated by 2.0 cm have a potential difference of 5.0 kV. Calculate the electric force acting on a dust particle between the plates that has a charge of 8.0×10^{-19} C.

Step 1 Write down the quantities given in the question:

$d = 2.0 \times 10^{-2}$ m

$V = 5.0 \times 10^{3}$ V

$Q = 8.0 \times 10^{-19}$ C

Hint: *When you write down the quantities it is important to include the units and to change them into base units. We have used powers of ten to do this.*

Step 2 To calculate the force F, you first need to determine the strength of the electric field:

$$\begin{aligned} E &= \frac{V}{d} \\ &= \frac{5\times10^{3}}{2.0\times10^{-2}} \\ &= 2.5\times10^{5}\,\text{V m}^{-1} \end{aligned}$$

Step 3 Now calculate the force on the dust particle:

$F = EQ$

$F = 2.5 \times 10^{5} \times 8.0 \times 10^{-19}$

$= 2.0 \times 10^{-13}$ N

Questions

4 Figure 21.12 shows an arrangement of parallel plates, each at a different voltage. The electric field lines are shown in the space between the first pair. Copy and complete the diagram to show the electric field lines in the other two spaces.

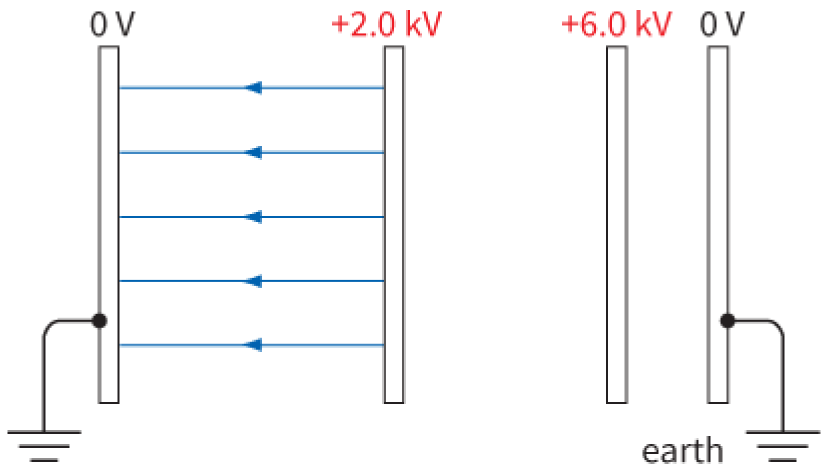


Figure 21.12: An arrangement of parallel plates. For Question 4.

5 Calculate the electric field strength at a point where a charge of 20 mC experiences a force vertically downwards of 150 N.

6 Calculate the electric field strength between two parallel charged plates, separated by 40 cm and with a potential difference between them of 1000 V.

7 An electron is situated in a uniform electric field. The electric force that acts on it is 8×10^{-16} N. What is the strength of the electric field? (Electron charge $e = 1.6 \times 10^{-19}$ C.)

8 Air is usually a good insulator. However, a spark can jump through dry air when the electric field strength is greater than about 40 000 V cm^{-1}. This is called electrical breakdown. The spark shows that electrical charge is passing through the air–there is a current. (Do not confuse this with a chemical spark such as you might see when watching fireworks; in that case, small particles of a chemical substance are burning quickly.)

a A Van de Graaff generator (Figure 21.13) is able to make sparks jump across a 4 cm gap. Estimate the voltage produced by the generator?

b The highest voltage reached by the live wire of a conventional mains supply is 325 V. In theory (but DO NOT try this), how close would you have to get to a live wire to get a shock from it?

c Estimate the voltage of a thundercloud from which lightning strikes the ground 100 m below.

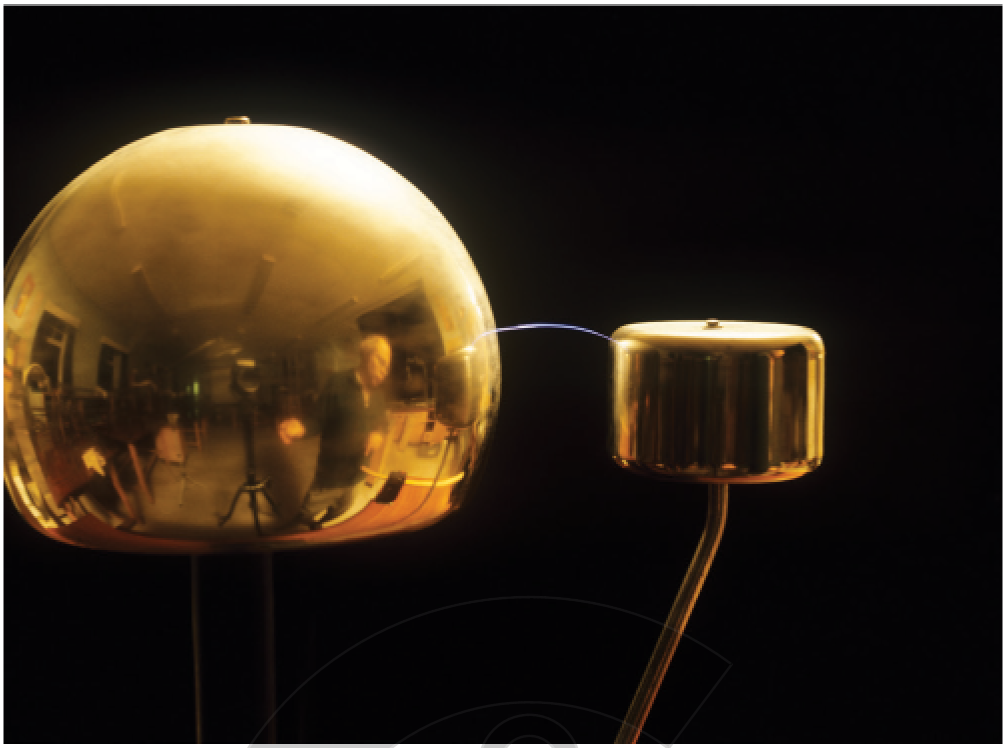


Figure 21.13: A Van de Graaff generator produces voltages sufficient to cause sparks in air.

21.4 Force on a charge

Now we can calculate the force *F* on a charge Q in the uniform field between two parallel plates. We have to combine the general equation for field strength $E = \frac{F}{Q}$ with the equation for the strength of a uniform field $E = -\frac{V}{d}$

This gives:

$$F = QE = -\frac{QV}{d}$$

For an electron with charge −*e*, this becomes:

$$F = \frac{eV}{d}$$

Figure 21.14 shows a situation where this force is important. A beam of electrons is entering the space between two charged parallel plates. How will the beam move?

We have to think about the force on a single electron. In the diagram, the upper plate is negative relative to the lower plate, and so the electron is pushed downwards. (You can think of this simply as the negatively charged electron being attracted by the positive plate, and repelled by the negative plate.)

If the electron were stationary, it would accelerate directly downwards. However, in this example, the electron is moving to the right. Its horizontal velocity will be unaffected by the force, but as it moves sideways it will also accelerate downwards. It will follow a curved path, as shown. This curve is a parabola.

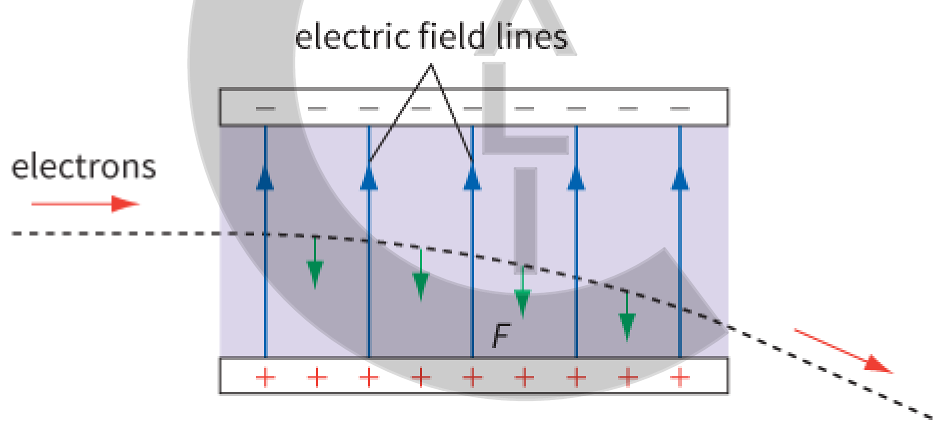


Figure 21.14: The parabolic path of a moving electron in a uniform electric field.

Note that the force on the electron is the same at all points between the plates, and it is always in the same direction (downwards, in this example).

This situation is equivalent to a ball being thrown horizontally in the Earth's uniform gravitational field (Figure 21.15). It continues to move at a steady speed horizontally, but at the same time it accelerates downwards. The result is the familiar curved trajectory shown. For the electron described, the force of gravity is tiny–negligible compared to the electric force on it.

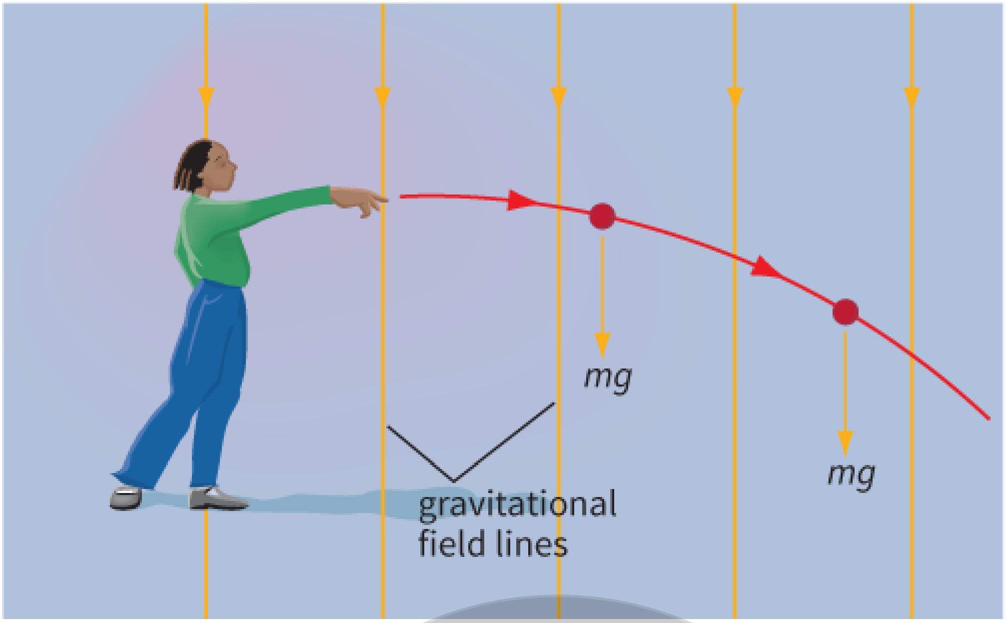


Figure 21.15: A ball, thrown in the uniform gravitational field of the Earth, follows a parabolic path.

Questions

9 In Figure 21.16, two parallel plates are shown, separated by 25 cm.

- **a** Copy the diagram and draw field lines to represent the field between the plates.
- **b** What is the potential difference between points A and B?
- **c** What is the electric field strength at C, and at D?
- **d** Calculate the electric force on a charge of +5 μC placed at C. In which direction does the force act?

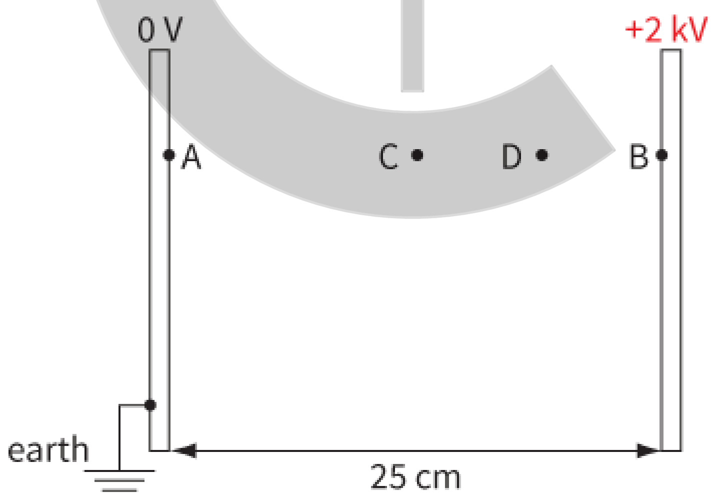


Figure 21.16: Two parallel, charged plates.

10 A particle of charge +2 μC is placed between two parallel plates, 10 cm apart and with a potential difference of 5 kV between them. Calculate the field strength between the plates, and the force exerted on the charge.

11 We are used to experiencing accelerations that are usually less than 10 m s^{-2}. For example, when we fall, our acceleration is about 9.81 m s^{-2}. When a car turns a corner sharply at speed, its acceleration is unlikely to be more than 5 m s^{-2}. However, if you were an electron, you would be used to experiencing much

greater accelerations than this. Calculate the acceleration of an electron in a television tube where the electric field strength is 50 000 V cm^{-1}. (Electron charge $-e = -1.6 \times 10^{-19}$ C; electron mass $m_e = 9.11 \times 10^{-31}$ kg.)

12 **a** Use a diagram to explain how the electric force on a charged particle could be used to separate a beam of electrons (e^-) and positrons (e^+) into two separate beams. (A positron is a positively charged particle that has the same mass as an electron but opposite charge. Positron–electron pairs are often produced in collisions in a particle accelerator.)

b Explain how this effect could be used to separate ions that have different masses and charges.

REFLECTION

When charged particles pass through a uniform electric field they are deflected.

On what factors does the deflection depend? How could this be used to compare masses of different ions? What variables must be kept the same or constant in order to give a fair comparison?

What would you do differently if you were to approach this same problem again?

SUMMARY

An electric field is a field of force, created by electric charges, and can be represented by electric field lines.
The strength of the field is the force acting per unit positive charge on a stationary positive charge at a point in the field: $$E = \frac{F}{Q}$$
In a uniform field (e.g. between two parallel charged plates), the force on a charge is the same at all points; the strength of the field is given by: $$E = -\frac{\Delta V}{\Delta x}$$
An electric charge moving initially at right-angles to a uniform electric field follows a parabolic path.

EXAM-STYLE QUESTIONS

1 A pair of charged parallel plates are arranged horizontally in a vacuum.

The upper plate carries a negative charge and the lower plate is earthed.

An electron enters the space between the plates at right angles to the electric field.

In which direction is the electric field between the plates and in which direction is the force on the electron? **[1]**

	Electric field strength	Force on the electron
A	downwards towards the lower plate	downwards towards the lower plate
B	downwards towards the lower plate	upwards towards the upper plate
C	upwards towards the upper plate	downwards towards the lower plate
D	upwards towards the upper plate	upwards towards the upper plate

Table 21.1

2 A pair of charged parallel plates are 2.0 cm apart and there is a potential difference of 5.0 kV across the plates.

A charged ion between the plates experiences a force of 1.2×10^{-13} N due to the field.

What is the charge on the ion? **[1]**

A 1.6×10^{-19} C

B 4.8×10^{-19} C

C 2.5×10^{-15} C

D 4.0×10^{-6} C

3 Figure 21.4 shows apparatus used to investigate the field between a pair of charged, parallel plates.

a Explain why the piece of gold foil deflects in the manner shown. **[1]**

b State and explain what would be observed if the gold foil momentarily touched the negatively charged plate. **[2]**

[Total: 3]

4 A charged dust particle in an electric field experiences a force of 4.4×10^{-13} N. The charge on the particle is 8.8×10^{-17} C. Calculate the electric field strength. **[2]**

5 Calculate the potential difference that must be applied across a pair of parallel plates, placed 4 cm apart, to produce an electric field of 4000 V m^{-1}. **[2]**

6 A potential difference of 2.4 kV is applied across a pair of parallel plates. The electric field strength between the plates is 3.0×10^{4} V m^{-1}.

a Calculate the separation of the plates. **[2]**

b The plates are now moved so that they are 2.0 cm apart. Calculate the electric **[2]**

field strength produced in this new position.

[Total: 4]

7 A variable power supply is connected across a pair of parallel plates. The potential difference across the plates is doubled and the distance between the plates is decreased to one-third of the original. State by what factor the electric field changes. Explain your reasoning. **[3]**

8 This diagram shows a positively charged sphere hanging by an insulating thread close to an earthed metal plate.

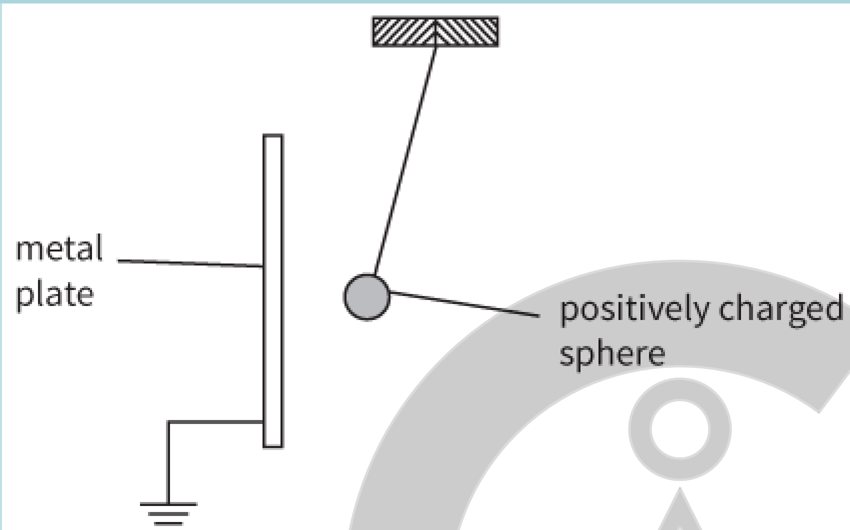


Figure 21.17

a Copy the diagram and draw five lines to show the electric field near the plate and the sphere. **[3]**

b Explain why the sphere is attracted towards the metal plate. **[2]**

c The sphere is now replaced with a similar negatively charged sphere.

i Explain what would be observed when the sphere is brought near to the earthed metal plate. **[2]**

ii Describe any changes to the electric field that would occur. **[1]**

[Total: 8]

9 This diagram shows a proton as it moves between two charged parallel plates. The charge on the proton is $+1.6 \times 10^{-19}$ C.

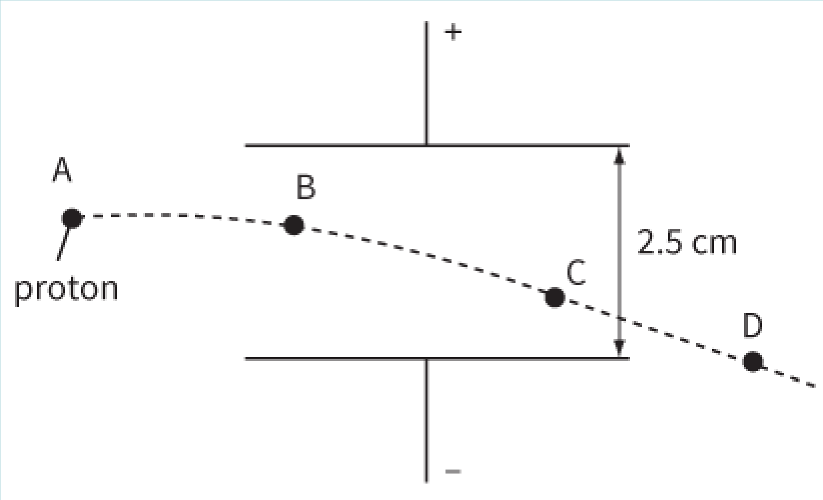


Figure 21.18

a Copy the diagram and draw the electric field between the parallel plates. **[2]**

The force on the proton when it is at position B is 6.4×10^{-14} N.

b In which direction does the force on the proton act when it is at position B? **[1]**

c What will be the magnitude of the force on the proton when it is at position C? **[1]**

d Calculate the electric field strength between the plates. **[2]**

e Calculate the potential difference between the plates. **[2]**

[Total: 8]

10 a Define what is meant by the electric field strength at a point. **[2]**

In a particle accelerator, a proton, initially at rest, is accelerated between two metal plates, as shown.

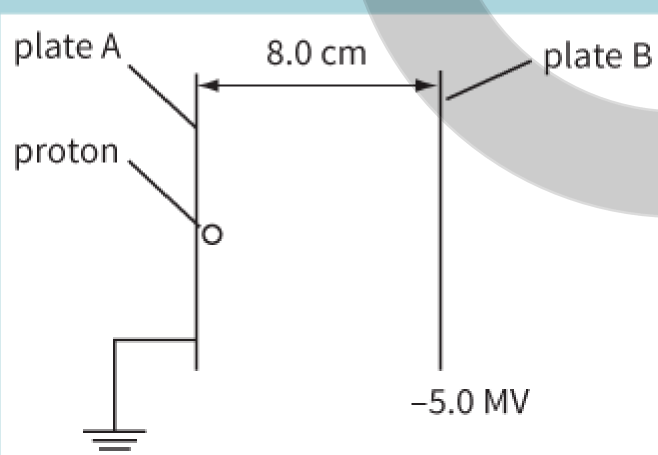


Figure 21.19

b Calculate the force on the proton due to the electric field. **[3]**

c Calculate the work done on the proton by the electric field when it moves from plate A to plate B. **[2]**

d State the energy gained by the proton. **[1]**

e Assuming that all this energy is converted to kinetic energy of the proton, calculate the speed of the proton when it reaches plate B. **[3]**

(Charge on a proton = +1.6 × 10^{-19} C; mass of a proton = 1.7 × 10^{-27} kg.)

[Total: 11]

11 a This diagram shows the structure of a spark plug in an internal combustion engine. The magnified section shows the end of the spark plug, with some of the lines of force representing the electric field.

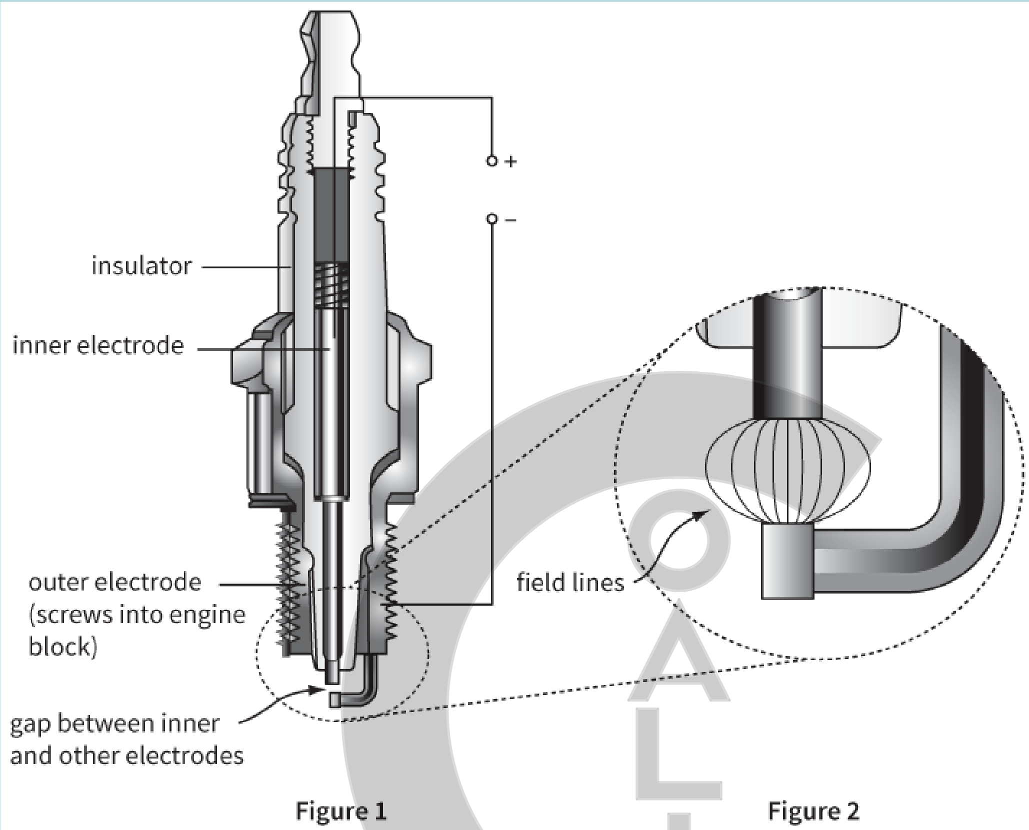


Figure 21.20

i Copy the field lines from the diagram. On your copy, draw arrows on the lines of force to show the direction of the field. **[1]**

ii What evidence does the diagram give that the field is strongest near the tip of the inner electrode? **[1]**

b The gap between the inner and outer electrodes is 1.25 mm and a field strength of 5.0 × 10^{6} N C^{-1} is required for electrical breakdown. Estimate the minimum potential difference that must be applied across the inner and outer electrodes for a spark to be produced. (You may treat the two electrodes as a pair of parallel plates.) **[2]**

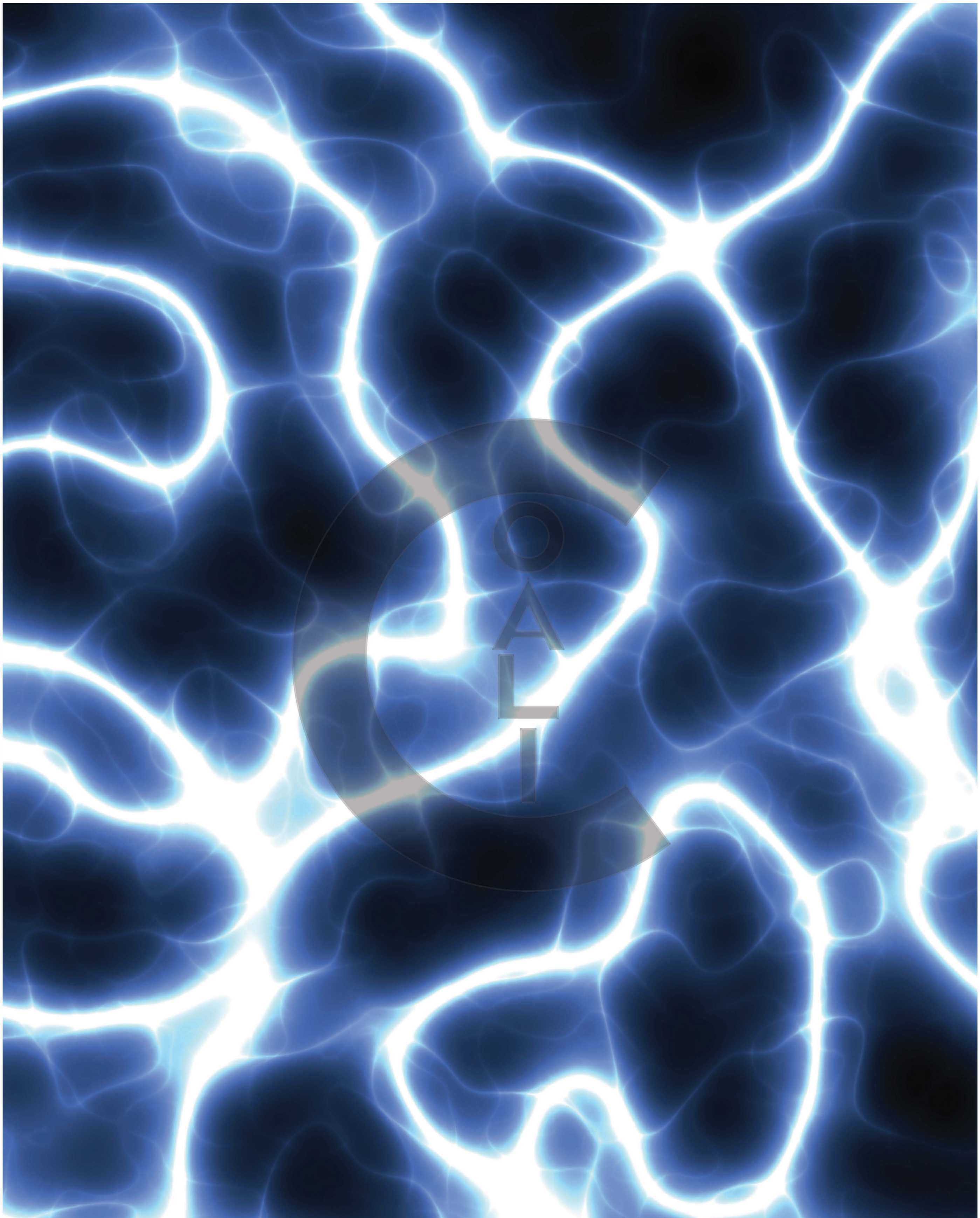
c When an electron is accelerated through a potential drop of approximately 20 V it will have sufficient energy to ionise a nitrogen atom. Show that an electron must move 4.0 μm to gain this energy. **[2]**

[Total: 6]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand that an electric field is a field of force	21.2			
define electric field as force per unit positive charge	21.3			
represent an electric field by means of field lines	21.3			
understand that the field between parallel plates is uniform	21.3			
recall and use the formula: $E = -\frac{\Delta V}{\Delta x}$	21.3			
describe the paths taken by charged particles as they pass through a uniform electric field.	21.4			



〉Chapter 22

Coulomb's law

LEARNING INTENTIONS

In this chapter you will learn how to:

- recall and use Coulomb's law
- calculate the field strength for a point charge
- recognise that for the electric field strength for a point outside a spherical conductor, the charge on the sphere may be considered to be a point charge at the centre of the sphere
- define electric potential
- calculate potential due to a point charge
- relate field strength to the potential gradient
- compare and contrast electric and gravitational fields.

BEFORE YOU START

- Cut a piece of paper into very small pieces. Rub a plastic rod (or comb) on your sleeve. Move it towards the pieces of paper. You should observe that the pieces of paper jump up and attach themselves to the comb. Hold the plastic rod still for a few minutes and you should observe something quite surprising.
- Write down what you observe and an explanation as to why this happened. Discuss your results with a fellow learner. Did they observe the same phenomenon? Did they come up with the same explanation?

LIVING IN A FIELD

The scientist in the Figure 22.1 is using a detector to measure the electric field produced by a mobile phone mast. People often worry that the electric field produced by a mobile phone transmitter may be harmful, but detailed studies have yet to show any evidence for this. If you hold a mobile phone close to your ear, the field strength will be far greater than that produced by a nearby mast.

With 5G being rolled out in various countries, what effect will this have on the local environment? Will it mean more masts and relay stations? Will copper cables be able to cope with the speed of the transmission of data needed to make 5G worthwhile? Will the investment needed to introduce 5G cause prices to rise for all customers?

Figure 22.1: Mobile phone masts produce weak electric fields – this scientist is using a small antenna to detect and measure the field of a nearby mast to ensure that it is within safe limits.

22.1 Electric fields

In Chapter 21, we presented some fundamental ideas about electric fields:

- An electric field is a field of force and can be represented by field lines.
- The electric field strength at a point is the force per unit positive charge that acts on a stationary charge:

$$\text{field strength} = \frac{\text{force}}{\text{charge}}$$

$$E = \frac{F}{Q}$$

- There is a uniform field between charged parallel plates:

$$\text{field strength} = \frac{\text{potential difference}}{\text{seperation}}$$

$$E = \frac{V}{d}$$

In this chapter, we will extend these ideas to consider how electric fields arise from electric charges. We will also compare electric fields with gravitational fields (Chapter 17).

22.2 Coulomb's law

Any electrically charged object produces an electric field in the space around it. It could be something as small as an electron or a proton, or as large as a planet or star. To say that it produces an electric field means that it will exert a force on any other charged object that is in the field. How can we determine the size of such a force?

The answer to this was first discovered by Charles Coulomb, a French physicist. He realised that it was important to think in terms of **point charges**; that is, electrical charges that are infinitesimally small so that we need not worry about their shapes. In 1785, Coulomb proposed a law that describes the force that one charged particle exerts on another. This law is remarkably similar in form to Newton's law of gravitation.

Coulomb's law states that any two point charges exert an electrical force on each other that is proportional to the product of their charges and inversely proportional to the square of the distance between them.

We consider two point charges Q_1 and Q_2 separated by a distance *r* (Figure 22.2). The force each charge exerts on the other is *F*. According to Newton's third law of motion, the point charges interact with each other and therefore exert equal but opposite forces on each other.

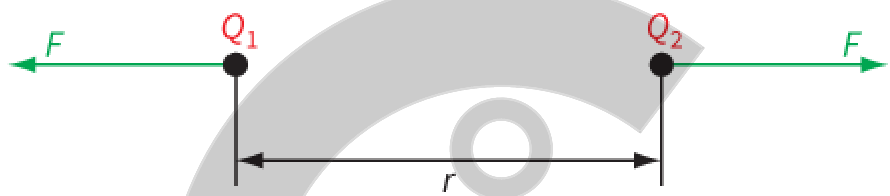


Figure 22.2: The variables involved in Coulomb's law.

According to Coulomb's law, we have:

$$\text{force} \propto \text{product of the charges}$$
$$F \propto Q_1 Q_2$$

$$\text{force} \propto \frac{1}{\text{distance}^2}$$
$$F \propto \frac{1}{r^2}$$

Therefore:

$$F \propto \frac{Q_1 Q_2}{r^2}$$

We can write this in a mathematical form:

$$F = \frac{kQ_1 Q_2}{r^2}$$

where *k* is the constant of proportionality.

This constant *k* is usually given in the form:

$$k = \frac{1}{4\pi\varepsilon_0}$$

where ε_0 is known as the **permittivity of free space** (ε is the Greek letter epsilon). The value of ε_0 is approximately 8.85×10^{-12} F m^{-1}. An equation for Coulomb's law is thus:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$$

where F is the force between two charges, Q_1 and Q_2, and r is the distance between their centres.

KEY EQUATION

Coulomb's law:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$$

Following your earlier study of Newton's law of gravitation, you should not be surprised by this relationship. The force depends on each of the properties producing it (in this case, the charges), and it is an inverse square law with distance–if the particles are twice as far apart, the electrical force is a quarter of its previous value (Figure 22.3).

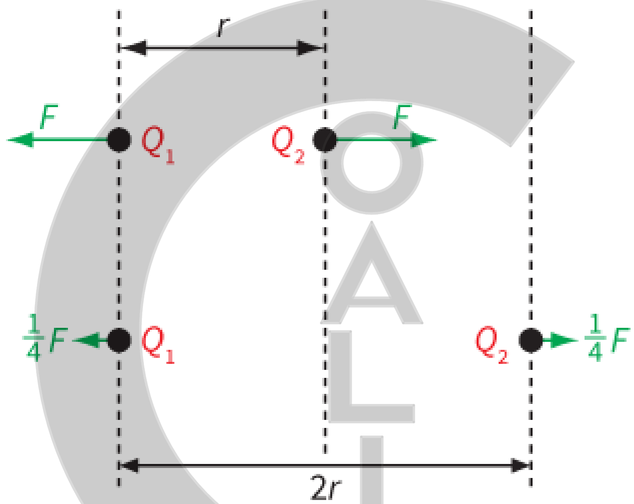


Figure 22.3: Doubling the separation results in one-quarter of the force, a direct consequence of Coulomb's law.

Note also that, if we have a positive and a negative charge, then the force F is negative. We interpret this as an attraction. Positive forces, as between two like charges, are repulsive. In gravity, we only have attraction.

So far, we have considered point charges. If we are considering uniformly charged spheres we measure the distance from the centre of one to the centre of the other – they behave as if their charge was all concentrated at the centre. Hence, we can apply the equation for Coulomb's law for both point charges (e.g. protons, electrons, etc.) and uniformly charged spheres, as long as we use the **centre-to-centre** distance between the objects.

PRACTICAL ACTIVITY 22.1

Investigating Coulomb's law

It is quite tricky to investigate the force between charged objects, because charge tends to leak away into the air or to the Earth during the course of any experiment. The amount of charge we can investigate is difficult to measure, and usually small, giving rise to tiny forces.

Figure 22.4 shows one method for investigating the inverse square law for two charged metal balls (polystyrene balls coated with conducting silver paint). As one charged ball is lowered down towards the other, their separation decreases and so the force increases, giving an increased reading on the balance.

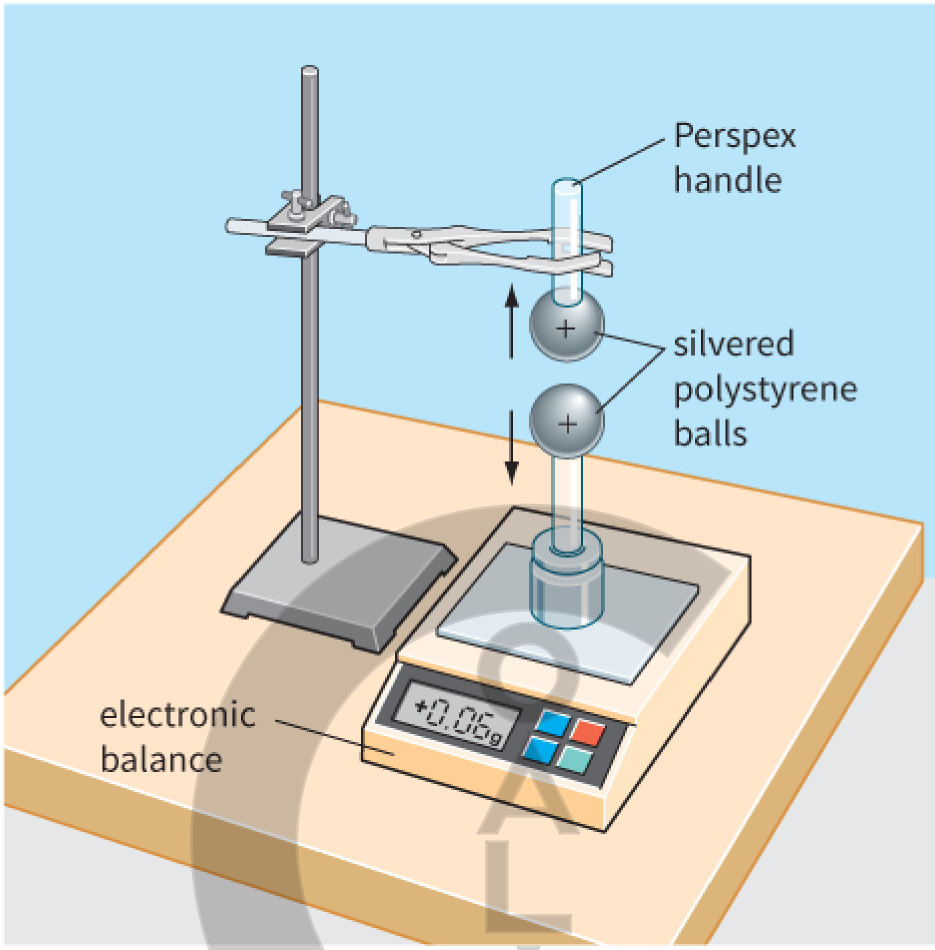


Figure 22.4: Investigating Coulomb's law.

22.3 Electric field strength for a radial field

In Chapter 21, we saw that the electric field strength at a point is defined as the force per unit charge exerted on a positive charge placed at that point, $E = \frac{F}{Q}$

So, to find the field strength near a point charge Q_1 (or outside a uniformly charged sphere), we have to imagine a small positive test charge Q_2 placed in the field, and determine the force per unit charge on it. We can then use the definition to determine the electric field strength for a point (or spherical) charge.

The force between the two point charges is given by:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$$

The electric field strength E due to the charge Q_1 at a distance of r from its centre is thus:

$$E = \frac{\text{force}}{\text{test charge}} = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2 Q_2} = \frac{Q}{4\pi\varepsilon_0 r^2}$$

where E is the electric field strength due to a point charge Q, and r is the distance from the point.

The field strength E is not a constant; it decreases as the distance r increases. The field strength obeys an inverse square law with distance–just like the gravitational field strength for a point mass. The field strength will decrease by a factor of four when the distance from the centre is doubled.

Note also that, since force is a vector quantity, it follows that electric field strength is also a vector. We need to give its direction as well as its magnitude in order to specify it completely. Worked example 1 shows how to use the equation for field strength near a charged sphere.

KEY EQUATION

Electric field strength:

$$E = \frac{Q}{4\pi\varepsilon_0 r^2}$$

WORKED EXAMPLE

1 A metal sphere of diameter 12 cm is positively charged. The electric field strength at the surface of the sphere is 4.0×10^5 V m^{-1}. Draw the electric field pattern for the sphere and determine the total surface charge.

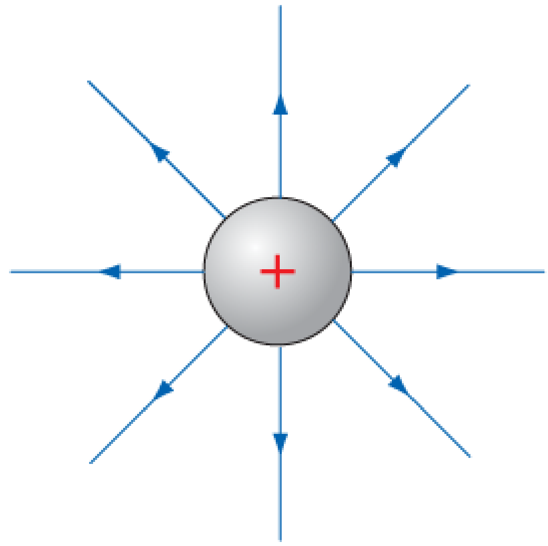


Figure 22.5: The electric field around a charged sphere.

Step 1 Draw the electric field pattern (Figure 22.5). The electric field lines must be normal to the surface and radial.

Step 2 Write down the quantities given:

electric field strength $E = 4.0 \times 10^5$ V m^{-1}

$$\text{radius } r = \frac{0.12}{2} = 0.06 \text{ m}$$

Step 3 Use the equation for the electric field strength to determine the surface charge:

$$E = \frac{Q}{4\pi\varepsilon_0 r^2}$$

$$Q = 4\pi\varepsilon_0 r^2 \times E$$

$$= 4\pi \times 8.85 \times 10^{-12} \times (0.06)^2 \times 4.0 \times 10^5$$

$$= 1.6 \times 10^{-7} \text{ C}$$

$$\equiv 0.16 \mu\text{C}$$

Questions

You will need the following data to answer the following questions. (You may take the charge of each sphere to be situated at its centre.)

$\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

1 A metal sphere of radius 20 cm carries a positive charge of +2.0 μC.

a What is the electric field strength at a distance of 25 cm from the centre of the sphere?

b An identical metal sphere carrying a negative charge of −1.0 μC is placed next to the first sphere. There is a gap of 10 cm between them. Calculate the electric force that each sphere exerts on the other.

Remember to calculate the centre-to-centre distance between the two spheres.

c Determine the electric field strength midway along a line joining the centres of the spheres.

2 A Van de Graaff generator produces sparks when the field strength at its surface is 4.0×10^4 V cm^{-1}. If the diameter of the sphere is 40 cm, what is the charge on it?

22.4 Electric potential

When we discussed gravitational potential (Chapter 17), we started from the idea of potential energy. The potential at a point is then the potential energy of unit mass at the point. We will approach the idea of electrical potential in the same way. However, you may be relieved to find that you already know something about the idea of electrical potential, because you know about voltage and potential difference. This topic shows how we formalise the idea of voltage, and why we use the expression ‘potential difference’ for some kinds of voltage.

Electric potential energy

When an electric charge moves through an electric field, its potential energy changes. Consider this example: if you want to move one positive charge closer to another positive charge, you have to push it (Figure 22.6). This is simply because there is a force of repulsion between the charges. You have to do work in order to move one charge closer to the other.

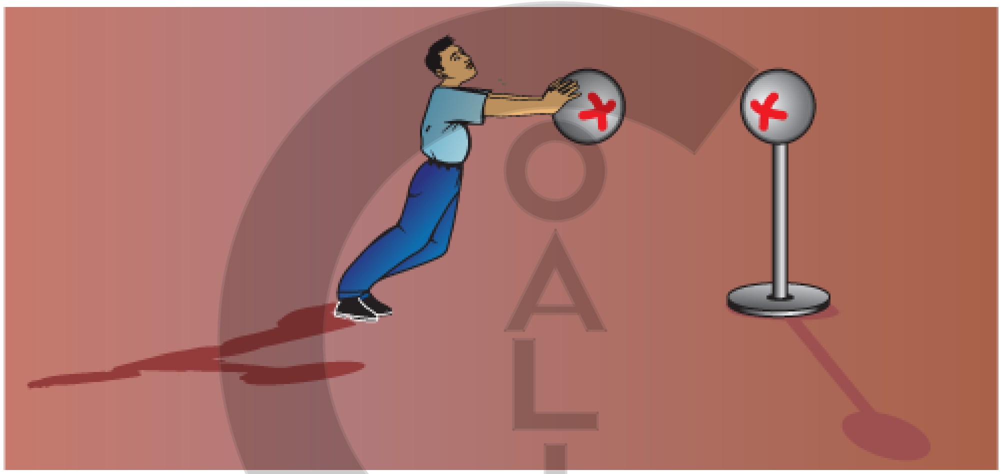


Figure 22.6: Work must be done to push one positive charge towards another.

In the process of doing work, energy is transferred from you to the charge that you are pushing. Its potential energy increases. If you let go of the charge, it will move away from the repelling charge. This is analogous to lifting up a mass; it gains gravitational potential energy as you lift it, and it falls if you let go.

Energy changes in a uniform field

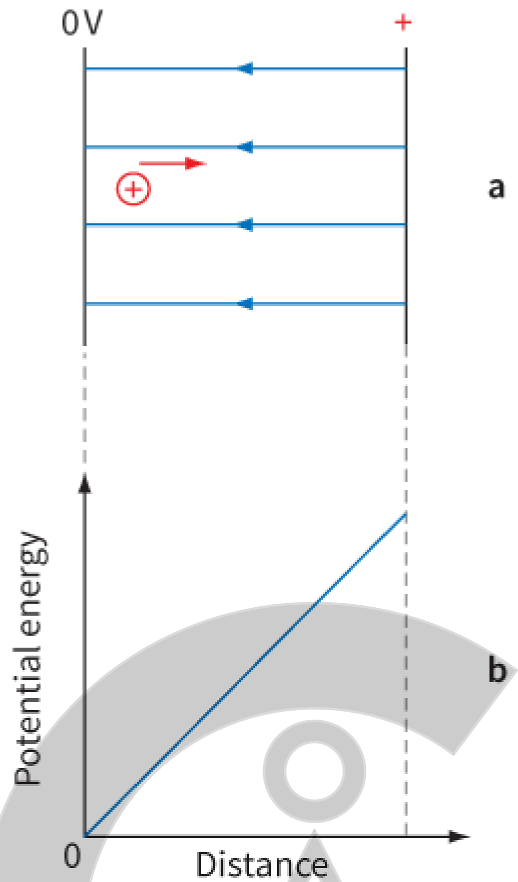


Figure 22.7: Electrostatic potential energy changes in a uniform field.

We can also think about moving a positive charge in a uniform electric field between two charged parallel plates. If we move the charge towards the positive plate, we have to do work. The potential energy of the charge is therefore increasing. If we move it towards the negative plate, its potential energy is decreasing (Figure 22.7a).

Since the force is the same at all points in a uniform electric field, it follows that the energy of the charge increases steadily as we push it from the negative plate to the positive plate. The graph of potential energy against distance is a straight line, as shown in Figure 22.7b.

We can calculate the change in potential energy of a charge Q as it is moved from the negative plate to the positive plate very simply. Potential difference is defined as the energy change (joules) per unit charge (coulombs) between two points (recall from Chapter 8 that one volt is one joule per coulomb). Hence, for charge Q, the work done in moving it from the negative plate to the positive plate is:

$$W = QV$$

We can rearrange this equation as:

$$V = \frac{W}{Q}$$

This is really how voltage V is defined. It is the energy per unit positive charge at a point in an electric field. By analogy with gravitational potential, we call this the electric potential at a point. Now you should be able to see that what we regard as the familiar idea of voltage should more correctly be referred to as electric potential. The difference in potential between two points is the potential difference (p.d.) between them.

Just as with gravitational fields, we must define the zero of potential (this is the point where we consider a charge to have zero potential energy). Usually, in a laboratory situation, we define the Earth as being at a potential of zero volts. If we draw two parallel charged plates arranged horizontally, with the lower one earthed

(Figure 22.8), you can see immediately how similar this is to our idea of gravitational fields. The diagram also shows how we can include equipotential lines in a representation of an electric field.

We can extend the idea of electric potential to measurements in electric fields. In Figure 22.9, the power supply provides a potential difference of 10 V. The value of the potential at various points is shown. You can see that the middle resistor has a potential difference across it of (8 − 2) V = 6 V.

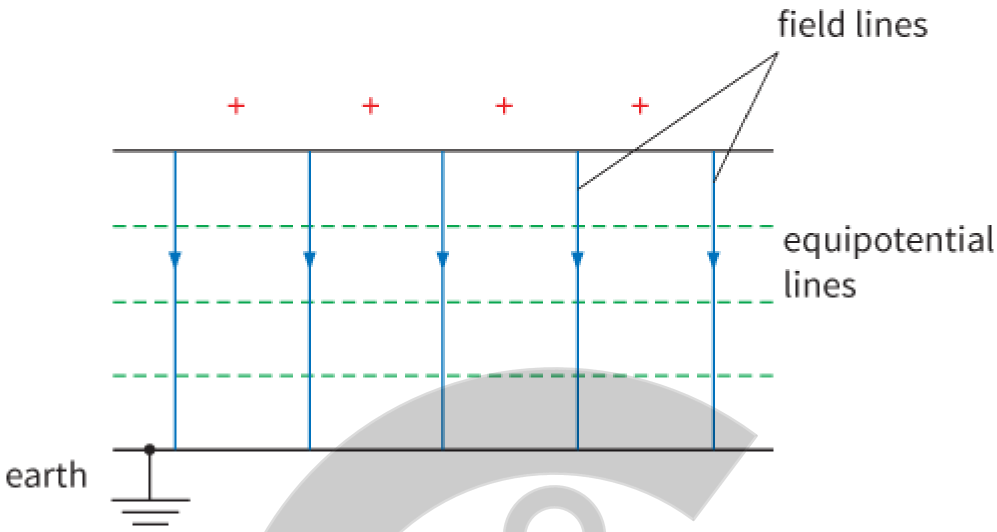


Figure 22.8: Equipotential lines in a uniform electric field.

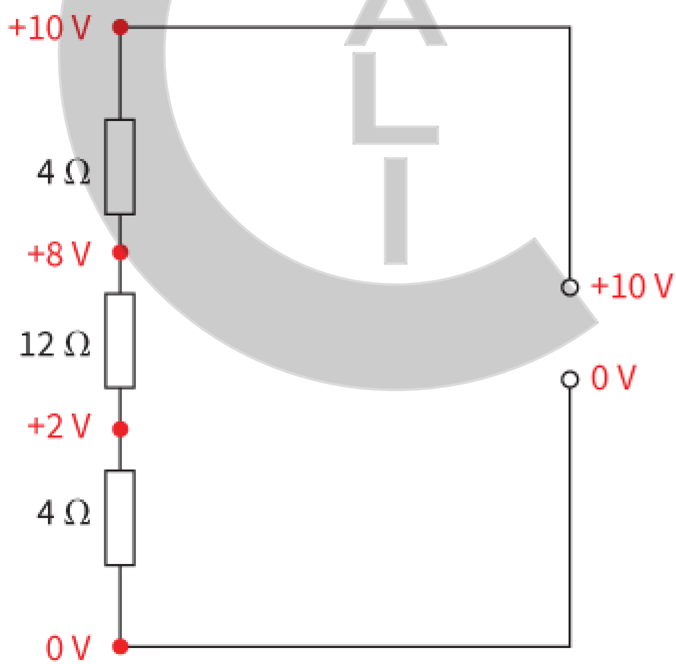


Figure 22.9: Changes in potential (shown in red) around an electric circuit.

Energy in a radial field

Imagine again pushing a small positive test charge towards a large positive charge. At first, the repulsive force is weak, and you have only to do a small amount of work. As you get closer, however, the force increases (Coulomb's law), and you have to work harder and harder.

The potential energy of the test charge increases as you push it. It increases more and more rapidly the closer you get to the repelling charge. This is shown by the graph in Figure 22.10. We can write an equation for the potential *V* at a distance *r* from a charge *Q*:

$$V = \frac{Q}{4\pi\varepsilon_0 r}$$

where *V* is the potential near a point charge *Q*, ε_0 is the permittivity of free space and *r* is the distance from the point.

(This comes from the calculus process of integration, applied to the Coulomb's law equation.)

You should be able to see how this relationship is similar to the equivalent formula for gravitational potential in a radial field:

$$\phi = -\frac{GM}{r}$$

Note that we do not need the minus sign in the electric equation as it is included in the charge. A negative charge gives an attractive (negative) field whereas a positive charge gives a repulsive (positive) field.

We can show these same ideas by drawing field lines and equipotential lines. The equipotentials get closer together as we get closer to the charge (Figure 22.11).

KEY EQUATION

$$V = \frac{Q}{4\pi\varepsilon_0 r}$$

Electric potential in a radial field due to a point charge.

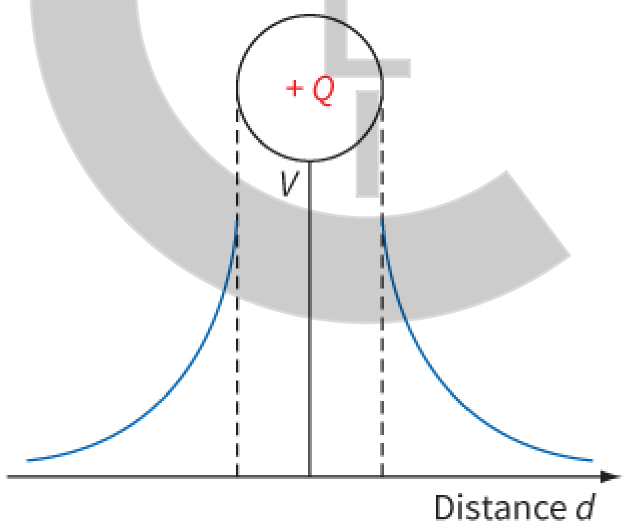


Figure 22.10: The potential changes according to an inverse law near a charged sphere.

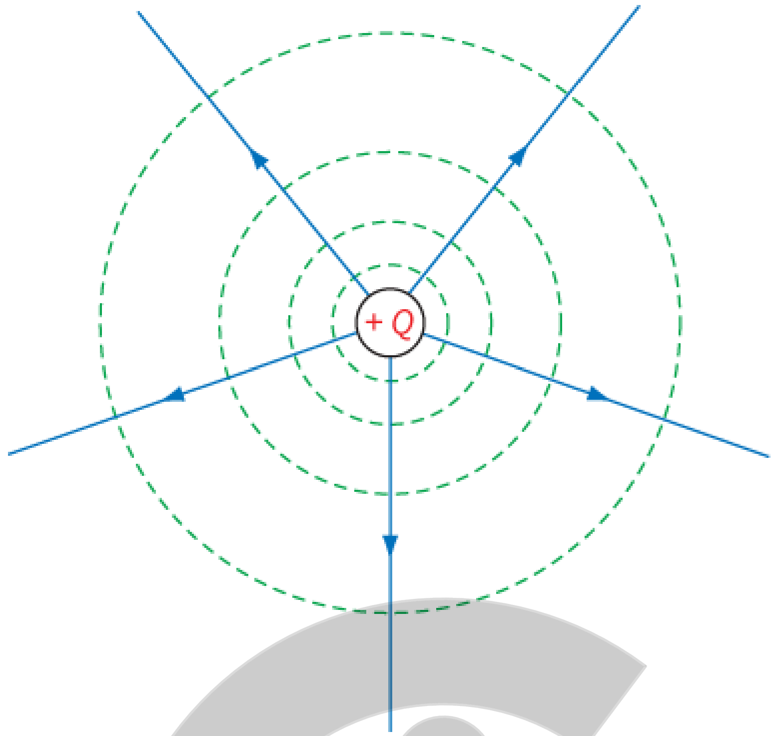


Figure 22.11: The electric field around a positive charge. The dashed equipotential lines are like the contour lines on a map; they are spaced at equal intervals of potential.

To arrive at this result, we must again define our zero of potential. Again, we say that a charge has zero potential energy when it is at infinity (some place where it is beyond the influence of any other charges). If we move towards a positive charge, the potential is positive. If we move towards a negative charge, the potential is negative.

This allows us to give a definition of electric potential: The **electric potential** at a point is equal to the work done per unit charge in bringing unit positive charge from infinity to that point.

Electric potential is a scalar quantity. To calculate the potential at a point caused by more than one charge, find each potential separately and add them. Remember that positive charges cause positive potentials and negative charges cause negative potentials.

Electrical potential energy

We have already defined electric potential energy between two points A and B as the work done in moving positive charge from point A to point B. This means that the potential energy change in moving point charge Q_1 from infinity towards a point charge Q_2 is equal to the potential at that point due to Q_2 multiplied by Q_1. In symbol form:

$$W = VQ_2$$

The potential V near the charge Q_2 is:

$$V = \frac{Q_2}{4\pi\varepsilon_0 r}$$

Thus the potential energy of the pair of point charges W (shown as E_p in the equation) is:

$$E_p = \frac{Qq}{4\pi\varepsilon_0 r}$$

KEY EQUATION

$$E_p = \frac{Qq}{4\pi\varepsilon_0 r}$$

Potential energy of a pair of point charges.

WORKED EXAMPLE

2 An α-particle approaches a gold nucleus and momentarily comes to rest at a distance of 4.5×10^{-14} m from the gold nucleus. Calculate the electric potential energy of the particles at that instant.
(Charge on the α-particle = 2*e*; charge on the nucleus = 79*e*.)

Step 1 Convert the charges to coulombs.

α-particle charge = $2 \times 1.6 \times 10^{-19}$ C

charge on the gold nucleus = $79 \times 1.6 \times 10^{-19}$ C

Step 2

$$W = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r}$$

$$= \frac{2 \times 1.6 \times 10^{-19} \times 79 \times 1.6 \times 10^{-19}}{4\pi \times 8.85 \times 10^{-12} \times 4.5 \times 10^{-14}}$$

$$= 8.1 \times 10^{-13}\ \text{J}$$

Electric potential difference near a charged sphere

We have already seen that the electric potential Δ*V* at a distance *r* from a point charge *Q* is given by the equation:

$$\Delta V = \frac{Q}{4\pi\varepsilon_0 r}$$

The potential difference between two points, one at a distance r_1 and the second at a distance r_2 from a charge *Q* is:

$$\Delta V = \frac{Q}{4\pi\varepsilon_0 r_1} - \frac{Q}{4\pi\varepsilon_0 r_2}$$

$$= \frac{Q}{4\pi\varepsilon_0}\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$$

This reflects the similar formula for the gravitational potential energy between two points near a point mass.

KEY EQUATION

$$\Delta V = \frac{Q}{4\pi\varepsilon_0}\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$$

Potential difference between two points from a charge.

Field strength and potential gradient

We can picture electric potential in the same way that we thought about gravitational potential. A negative charge attracts a positive test charge, so we can regard it as a potential ‘well’. A positive charge is the opposite–

a ‘hill’ (Figure 22.12). The strength of the field is shown by the slope of the hill or well:

field strength = −potential gradient

The minus sign is needed because, if we are going up a potential hill, the force on us is pushing us back down the slope, in the opposite direction.

KEY IDEA

electric field strength = −potential gradient

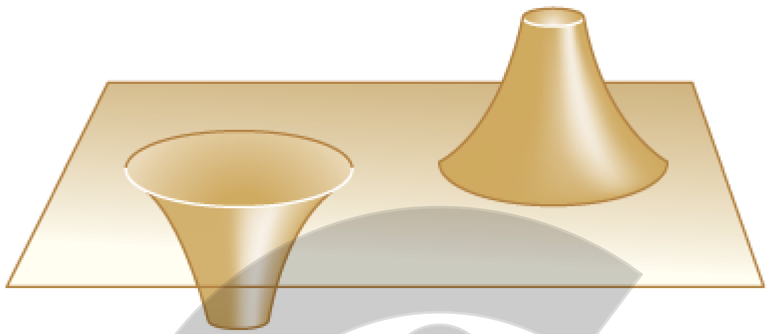


Figure 22.12: A ‘potential well’ near a negative charge, and a ‘potential hill’ near a positive charge.

This relationship applies to all electric fields. For the special case of a uniform field, the potential gradient *E* is constant. Its value is given by:

$$E = -\frac{\Delta V}{\Delta d}$$

where *V* is the potential difference between two points separated by a distance *d*.

(This is the same as the relationship $E = \frac{V}{d}$ quoted in Chapter 21.)

Worked example 3 shows how to determine the field strength from a potential–distance graph.

WORKED EXAMPLE

3 The graph (Figure 22.13) shows how the electric potential varies near a charged object. Calculate the electric field strength at a point 5 cm from the centre of the object.

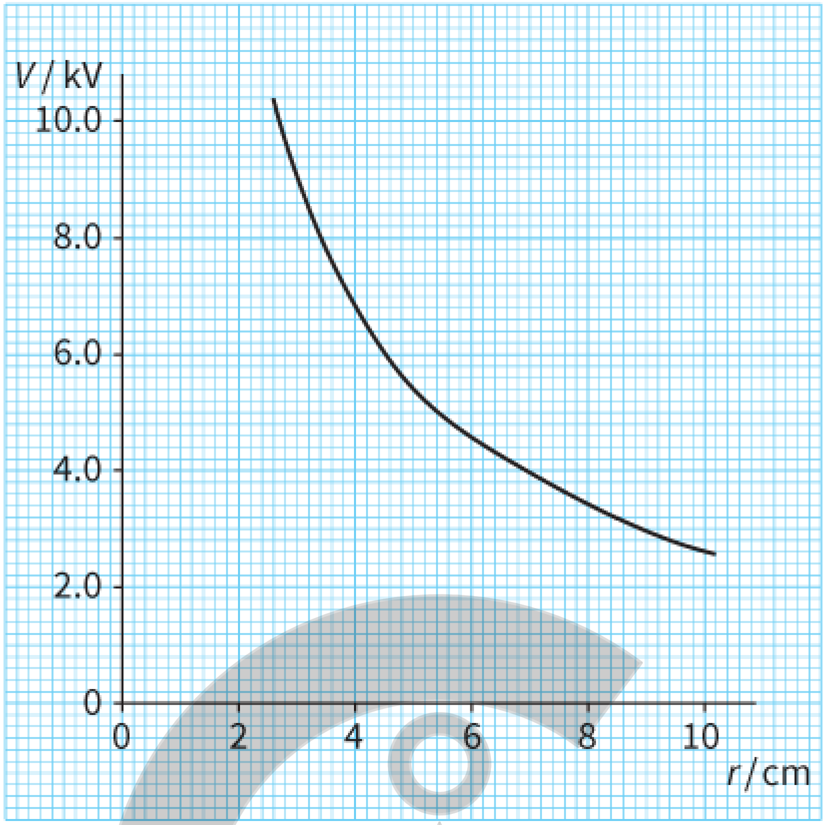


Figure 22.13: Variation of the potential *V* near a positively charged object.

Step 1 Draw the tangent to the graph at the point 5.0 cm. This is shown in Figure 22.14.

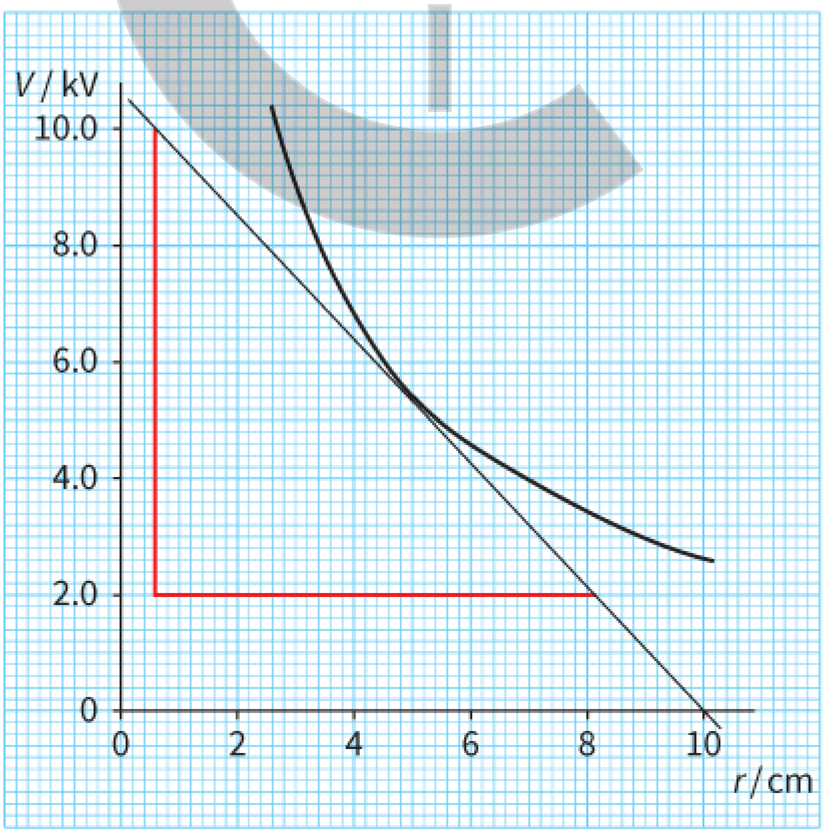


Figure 22.14: Drawing the tangent to the *V*–*r* graph to find the electric field strength *E*.

Step 2 Calculate the gradient of the tangent:

$$\begin{aligned}\text{gradient} &= \frac{\Delta V}{\Delta r}\\ &= \frac{10.0-2.0}{0.6-8.2}\\ &= -1.05\,\text{kV cm}^{-1}\\ &\equiv -1.05\times10^{5}\,\text{V m}^{-1}\\ &\approx -1.0\times10^{5}\,\text{V m}^{-1}\end{aligned}$$

The electric field strength is therefore $+1.0 \times 10^5$ V m^{-1} or $+1.0 \times 10^5$ N C^{-1}.

Remember E = −potential gradient.

Questions

3 **a** What is the electrical potential energy of a charge of +1 C placed at each of the points A, B, C, D between the charged, parallel plates shown in Figure 22.15?

b What would be the potential energy of a +2 C charge at each of these points? (C is halfway between A and B, D is halfway between C and B.)

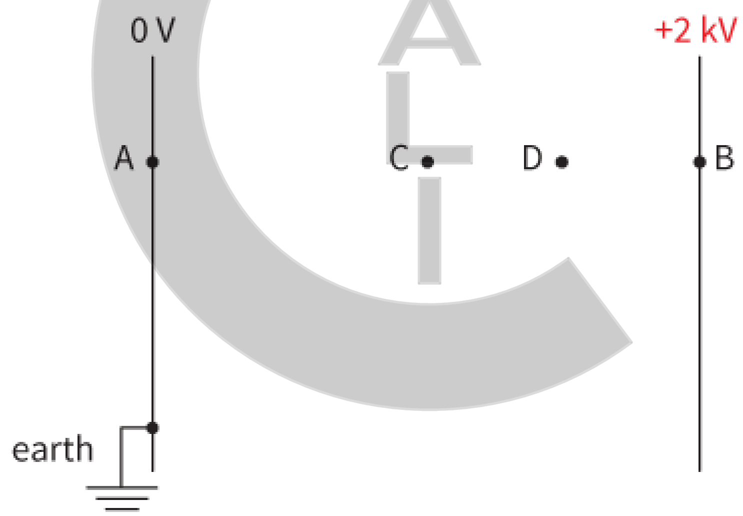


Figure 22.15: A uniform electric field. For Question 3.

4 A Van de Graaff generator has a spherical dome of radius 10 cm. It is charged up to a potential of 100 000 V (100 kV). How much charge is stored on the dome? What is the potential at a distance of 10 cm from the dome?

5 **a** How much work is done in moving a +1 C charge along the following paths shown in Figure 22.16: from E to H; from E to F; from F to G; from H to E?

b How do your answers differ for a:

i −1 C charge?

ii +2 C charge?

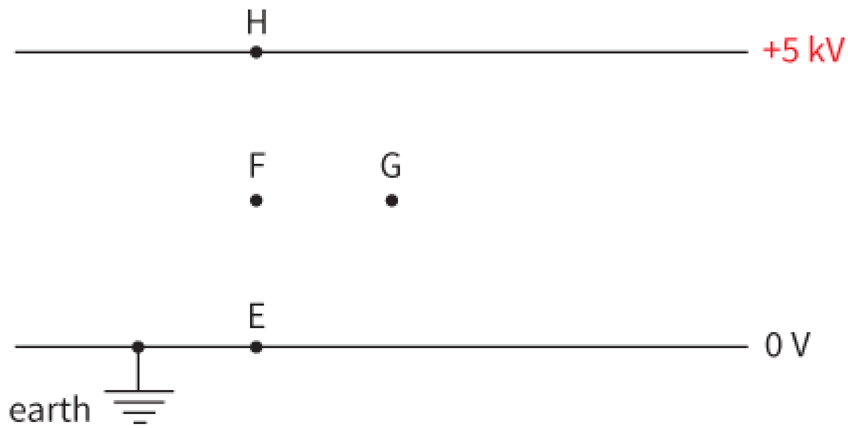


Figure 22.16: A uniform electric field. For Question 5.

22.5 Gravitational and electric fields

There are obvious similarities between the ideas we have used in this chapter to describe electric fields and those we used in Chapter 17 for gravitational fields. This can be helpful, or it can be confusing! The summary given in Table 22.1 is intended to help you to sort them out.

An important difference is this: electric charges can be positive or negative, so they can attract or repel. There are no negative masses, so there is only attraction in a gravitational field.

Gravitational fields	Electric fields
Origin arise from masses	**Origin** arise from electric charges
Vector forces only gravitational attraction, no repulsion	**Vector forces** both electrical attraction and repulsion are possible (because of positive and negative charges)
All gravitational fields field strength $g = \frac{F}{m}$ field strength is force per unit mass	**All electric fields** field strength $E = \frac{F}{Q}$ field strength is force per unit positive charge
Units F in N, g in N kg^{-1} or m s^{-2}	**Units** F in N, E in N C^{-1} or V m^{-1}
Uniform gravitational fields parallel gravitational field lines g = Constant	**Uniform electric fields** parallel electric field lines $E = \frac{V}{d}$ = constant

<table>
<tr><th>Gravitational fields</th><th>Electric fields</th></tr>
<tr><td>

Spherical gravitational fields

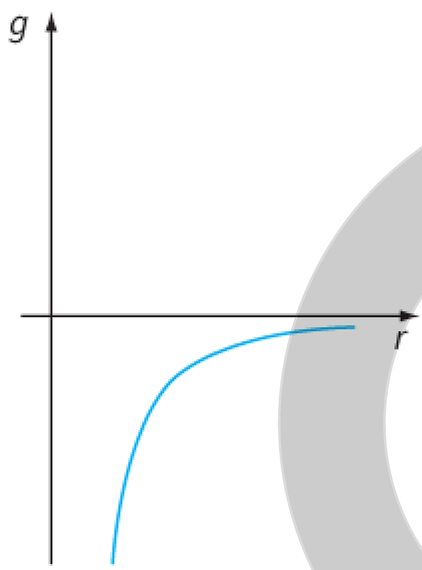
radial field lines

force given by Newton's law: $F = \frac{GMm}{r^2}$

field strength is therefore: $g = \frac{GM}{r^2}$

(Gravitational forces are always attractive, so we show *g* on a graph against *r* as negative.)

force and field strength obey an inverse square law with distance



</td><td>

Spherical electric fields

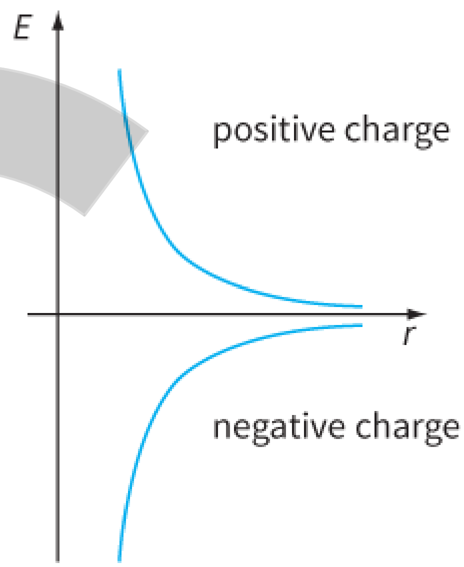
radial field lines

force given by Coulomb's law: $F = \frac{Q_1Q_2}{4\pi\varepsilon_0 r^2}$

field strength is therefore: $E = \frac{Q}{4\pi\varepsilon_0 r^2}$

(A negative charge gives an attractive field; a positive charge gives a repulsive field.)

force and field strength obey an inverse square law with distance



</td></tr>
</table>

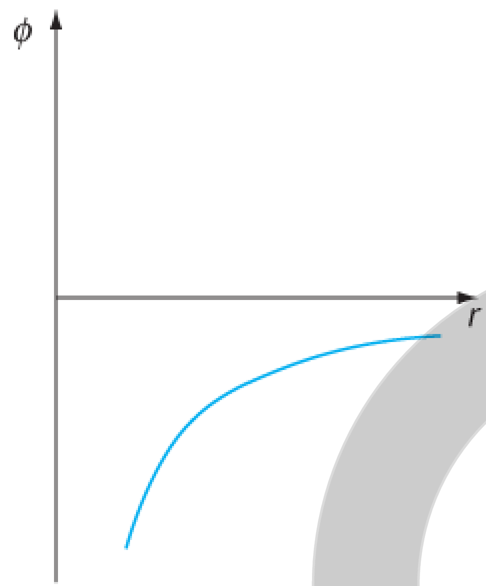
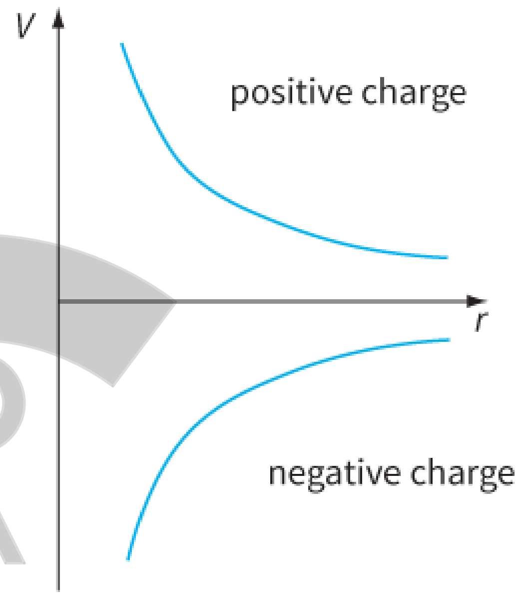
Gravitational fields	Electric fields
Gravitational potential given by: $\phi = \frac{GM}{r}$ potential obeys an inverse relationship with distance and is zero at infinity potential is a scalar quantity and is always negative 	**Electric potential** given by: $V = \frac{Q}{4\pi\varepsilon_0 r}$ potential obeys an inverse relationship with distance and is zero at infinity potential is a scalar quantity 

Table 22.1: Gravitational and electric fields compared.

Question

You will need the following data to answer the question.

proton mass = 1.67×10^{-27} kg

proton charge = $+1.60 \times 10^{-19}$ C

$\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

$G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}

6 Two protons in the nucleus of an atom are separated by a distance of 10^{-15} m. Calculate the electrostatic force of repulsion between them, and the force of gravitational attraction between them. (Assume the protons behave as point charges and point masses.) Is the attractive gravitational force enough to balance the repulsive electrical force? What does this suggest to you about the forces between protons within a nucleus?

REFLECTION

In Question 6, we showed that in the atomic nuclei the electric force is much larger than the gravitational force. Is this also true in the formation of atoms? Yet, in the formation of stars and planetary systems, the gravitational force rules.

Discus and explain why there is this difference.

What did this discussion and explanation reveal about you as a learner?

COALI

SUMMARY

Coulomb's law states that two point charges exert an electrical force on each other that is proportional to the product of their charges and inversely proportional to the square of the distance between them.
The equation for Coulomb's law is: $F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$
A point charge *Q* gives rise to a radial field. The electric field strength is given by the equation: $$E = \frac{Q}{4\pi\varepsilon_0 r^2}$$
The electric potential at a point is defined as the work done per unit positive charge in bringing charge from infinity to the point.
For a point charge, the electric potential is given by: $V = \frac{Q}{4\pi\varepsilon_0 r}$
The electric potential energy of two point charges is: $W = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r}$

EXAM-STYLE QUESTIONS

1 How does the potential V change with the distance r from a point charge? **[1]**

A $V \propto r$

B $V \propto r^2$

C $V \propto r^{-1}$

D $V \propto r^{-2}$

2 The electric field strength 20 cm from an isolated point charge is 1.9×10^4 N C^{-1}.

What is the electric field strength 30 cm from the charge? **[1]**

A 8.4×10^3 N C^{-1}

B 1.3×10^4 N C^{-1}

C 2.9×10^4 N C^{-1}

D 4.3×10^4 N C^{-1}

3 On a copy of this diagram, draw the electric fields between the charged objects. **[5]**

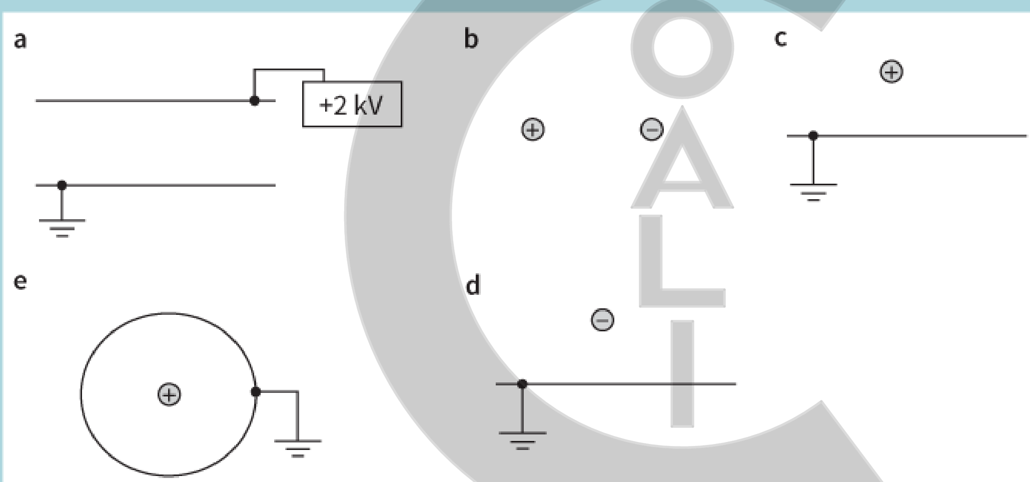


Figure 22.17

4 Two parallel plates are 4 cm apart and have a potential difference of 2.5 kV between them.

a Calculate the electric field strength between the plates. **[2]**

b A small piece of dust carrying a charge of +2.4 nC moves into the space between the plates.

i Calculate the force on the dust particle. **[2]**

ii The mass of the dust particle is 4.2 mg. Calculate the acceleration of the particle towards the negative plate. **[2]**

[Total: 6]

5 A small sphere carries a charge of 2.4×10^{-9} C. Calculate the electric field strength at a distance of:

a 2 cm from the centre of the sphere **[2]**

b 4 cm from the centre of the sphere. **[2]**

[Total: 4]

6 A conducting sphere of diameter 6.0 cm is mounted on an insulating base. The sphere is connected to a power supply that has an output voltage of 20 kV.

a Calculate the charge on the sphere. **[3]**

b Calculate the electric field strength at the surface of the sphere. **[2]**

[Total: 5]

7 The nucleus of a hydrogen atom carries a charge of $+1.60 \times 10^{-19}$ C.

Its electron is at a distance of 1.05×10^{-10} m from the nucleus.

Calculate the ionisation potential of hydrogen. **[3]**

(Hint: This is equal to the work per unit charge needed to remove the electron to infinity.)

8 a Define electric field strength. **[2]**

b Two charged conducting spheres, each of radius 1.0 cm, are placed with their centres 10 cm apart, as shown.

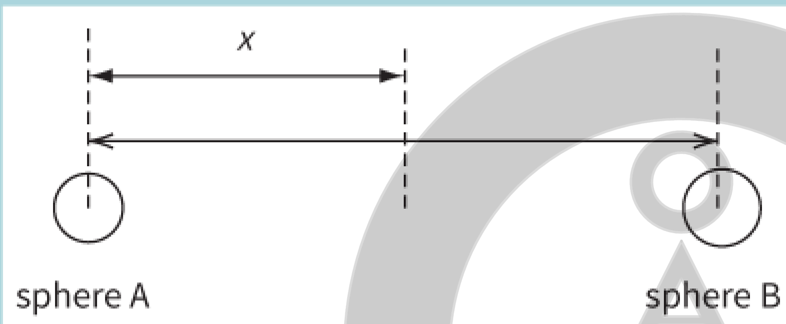


Figure 22.18

Sphere A carries a charge of $+2.0 \times 10^{-9}$ C.

The graph shows how the electric field strength between the two spheres varies with distance *x*.

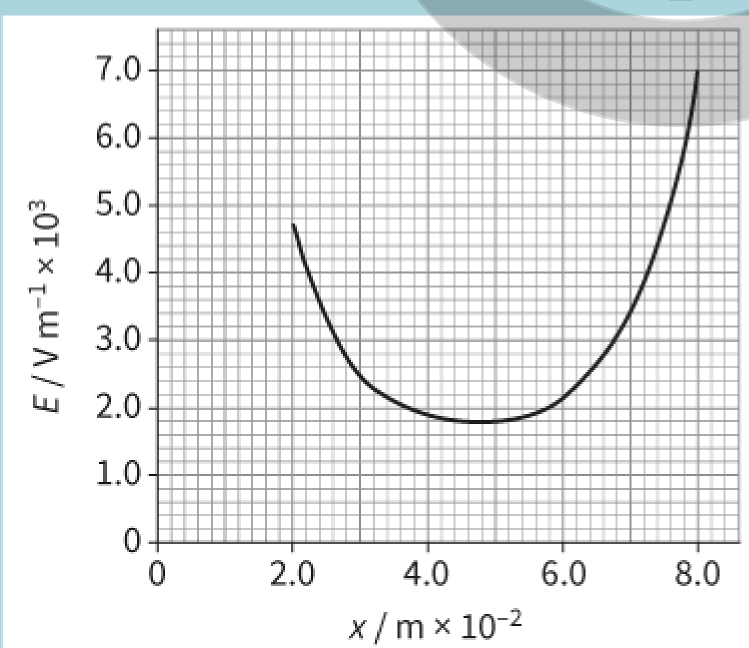


Figure 22.19

i Determine the field strength 5.0 cm from the centre of sphere A **[2]**

ii Use your result to **i** to calculate the charge on sphere B. **[3]**

c i Sphere B is now removed. Calculate the potential at the surface of sphere A. **[2]**

ii Suggest and explain how the potential at the surface of sphere A would compare before and after sphere B was removed. **[2]**

[Total: 11]

9 An α-particle emitted in the radioactive decay of radium has a kinetic energy of 8.0×10^{-13} J.

a **i** Calculate the potential difference that an α-particle, initially at rest, would have to be accelerated through to gain this energy. **[2]**

ii Calculate the speed of the α-particle at this kinetic energy. **[3]**

b This diagram shows the path of an α-particle of this energy as it approaches a gold nucleus head-on.

Figure 22.20

i State the speed of the α-particle at its point of closest approach to the gold nucleus. **[1]**

ii Write down the kinetic energy of the α-particle at this point. **[1]**

iii Write down the potential energy of the α-particle at this point. **[1]**

c Use your answer to part **b iii** to show that the α-particle will reach a distance of 4.5×10^{-14} m from the centre of the gold nucleus. **[2]**

d Suggest and explain what this information tells us about the gold nucleus. **[2]**

(Mass of an α-particle = 6.65×10^{-27} kg; charge on an α-particle = $+2e$; charge on a gold nucleus = $+79e$.)

[Total: 12]

10 **a** Define electric potential at a point. **[2]**

b This graph shows the electrical potential near an antiproton.

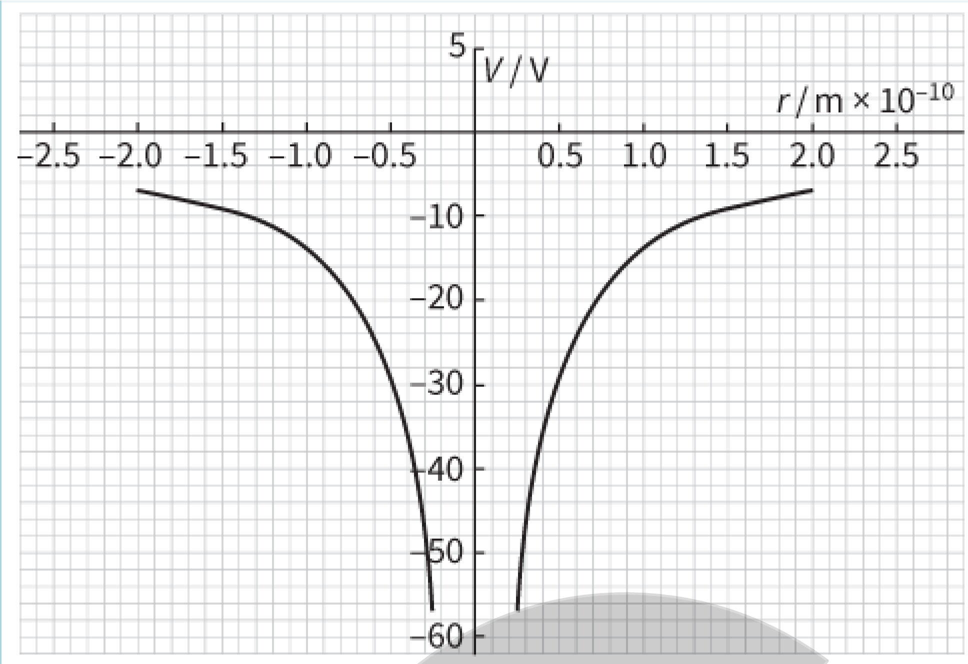


Figure 22.21

i Determine the potential at a distance 0.53×10^{-10} m from the antiproton. **[2]**

ii Determine the potential energy a positron would have at this distance. **[2]**

c Use the graph to determine the electric field at this distance from the antiproton. **[2]**

[Total: 8]

11 This diagram shows a conducting sphere of radius 0.80 cm carrying a charge of $+6.0 \times 10^{-8}$ C resting on a balance.

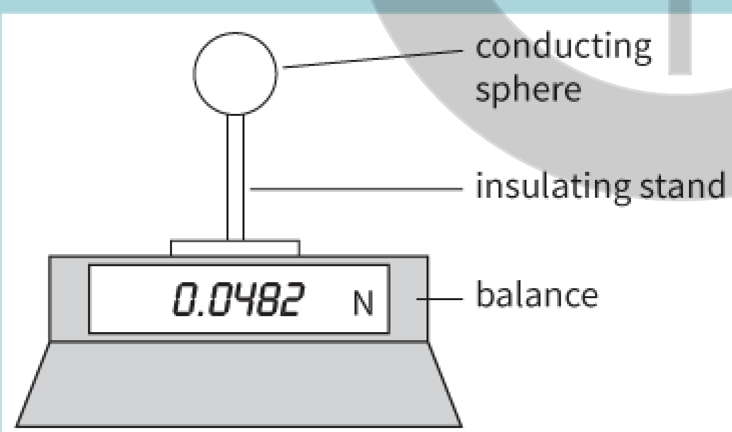


Figure 22.22

a Calculate the electric field at the surface of the sphere. **[2]**

b An identical sphere carrying a charge of -4.5×10^{-8} C is held so that its centre is 5.0 cm vertically above the centre of the first sphere.

i Calculate the electric force between the two spheres. **[2]**

ii Calculate the new reading on the balance. **[1]**

c The second sphere is moved vertically downwards through 1.5 cm. Calculate **[3]**

the work done against the electric field in moving the sphere.

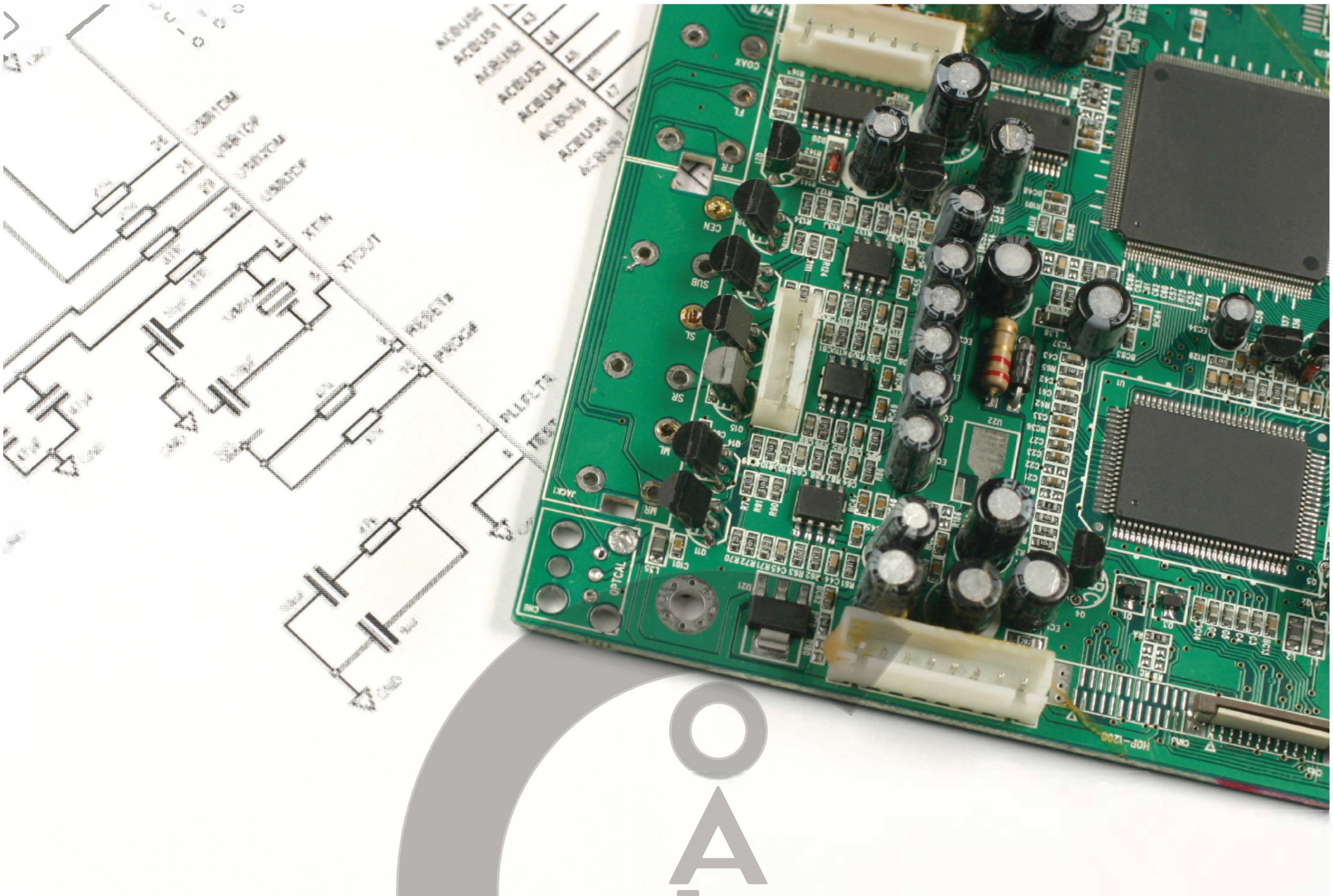
[Total: 8]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand the nature of the electric field	22.1			
represent and interpret an electric field using field lines	22.4			
recall and use Coulomb's law: $F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$	22.2			
understand that electric field *g* is defined as the electric force per unit coulomb	22.5			
derive from Coulomb's law of gravitation: $E = \frac{Q}{4\pi\varepsilon_0 r^2}$	22.3			
recall and use the equation: $E = \frac{Q}{4\pi\varepsilon_0 r^2}$	22.3			
recall and use: $E = \frac{\Delta V}{\Delta d}$	22.4			
define electric potential at a point, *V*, as the work done in bringing unit charge from infinity to that point	22.4			
recognise that the electric potential at infinity is zero	22.4			
recognise that the electric potential increases as you move closer to a positively charged object	22.4			
recognise that the electric potential decreases as you move closer to a negatively charged object	22.4			
recall and use the formula: gravitational potential $V = \frac{Q}{4\pi\varepsilon_0 r}$	22.4			
use the formula: $\Delta V = \frac{Q}{4\pi\varepsilon_0}\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$	22.4			
understand that the electric potential energy of two point masses is equal to: $W = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r}$	22.4			

COALI



〉Chapter 23

Capacitance

LEARNING INTENTIONS

In this chapter you will learn how to:

- define capacitance and state its unit, the farad
- solve problems involving charge, voltage and capacitance
- deduce the electric potential energy stored in a capacitor from a potential–charge graph
- deduce and use formulae for the energy stored by a capacitor
- derive and use formulae for capacitances in series and parallel
- recognise and use graphs showing variation of potential difference, current and charge as a capacitor discharges.
- recall and use the time constant for a capacitor–resistor circuit
- use the equation for the discharge of a capacitor through a resistor.

BEFORE YOU START

- In order to avoid an electric shock, electrical engineers regularly connect various points to Earth, even though the equipment is disconnected from the mains supply.
- What does this suggest to you is happening? How can you get a shock when the equipment is not connected to the mains? Discuss with a partner and be prepared to share your thoughts with the rest of the class.

CAPACITORS

Most electronic devices, such as radios, computers and MP3 players, make use of components called capacitors. These are usually quite small, but Figure 23.1 shows a giant capacitor, specially constructed to store electrical energy at the Fermilab particle accelerator in the United States.

Fermilab is a particle physics and accelerator laboratory. Particle accelerators, as the name suggests, accelerate particles, such as protons, up to incredibly high energies. The ‘tevatron’ at Fermilab can accelerate protons up to energies of approximately 2 TeV (10^{12} eV). High-energy, but short-lasting voltage pulses (100 000 V lasting 10^{-5} s) are required to accelerate the particles. Such pulses would disrupt the public electricity supply. To ensure the public power supply is evenly loaded and is not disrupted by peak pulses, large capacitors (temporary energy storage devices) are continuously charged and discharged 50 times per second.

Wind turbines and solar cells only generate energy in suitable weather conditions. Could huge capacitors be used to store electrical energy generated when the weather conditions are suitable for use at times when it is not? How else could the energy be stored?

Figure 23.1: One of the world's largest capacitors, built to store energy at the Fermilab particle accelerator.

23.1 Capacitors in use

Capacitors are used to store energy in electrical and electronic circuits. This means that they have many valuable applications. For example, capacitors are used in computers; they store energy in normal use and then they gradually release this energy if there is a power failure, so that the computer will operate long enough to save valuable data. Figure 23.2 shows a variety of shapes and sizes of capacitors.

Every capacitor has two leads, each connected to a metal plate. To store energy, these two plates must be given equal and opposite electric charges. Between the plates is an insulating material called the dielectric. Figure 23.3 shows a simplified version of the construction of a capacitor; in practice, many have a spiral form.

Figure 23.2: A variety of capacitors.

To move charge onto the plates of a capacitor, it must be connected to a voltage supply. The negative terminal of the supply pushes electrons onto one plate, making it negatively charged. Electrons are repelled from the other plate, making it positively charged. Figure 23.4 shows that there is a flow of electrons all the way round the circuit.

The two ammeters will give identical readings. The current stops when the potential difference (p.d.) across the capacitor is equal to the electromotive force (e.m.f.) of the supply. We then say that the capacitor is 'fully charged'.

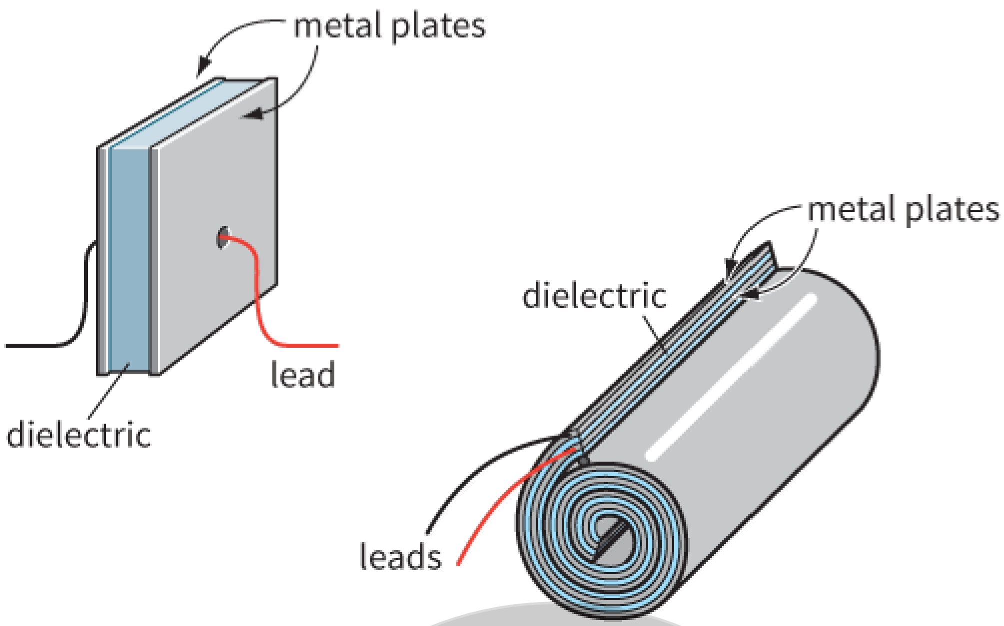


Figure 23.3: The construction of two types of capacitor.

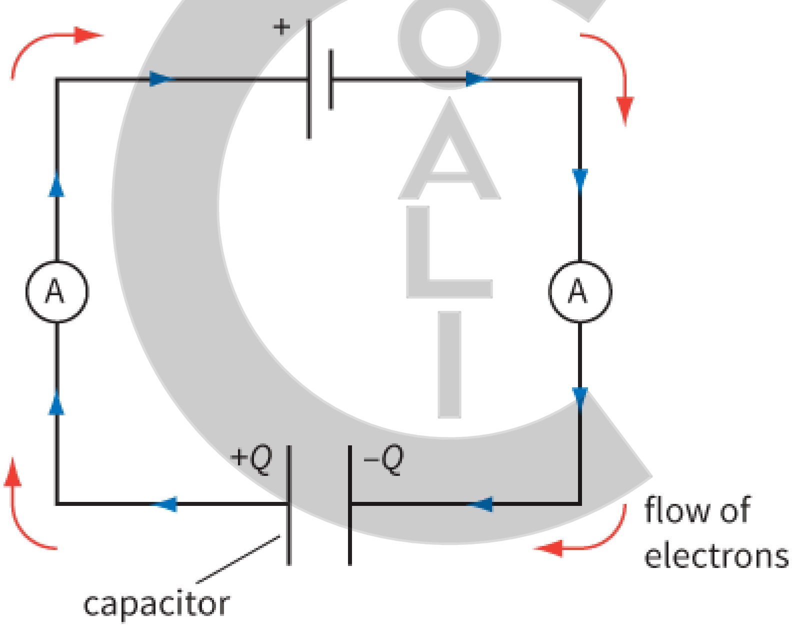


Figure 23.4: The flow of charge when a capacitor is charged up.

Note: The convention is that current is the flow of positive charge. Here, it is free electrons that flow. Electrons are negatively charged; conventional current flows in the opposite direction to the electrons (Figure 23.5).

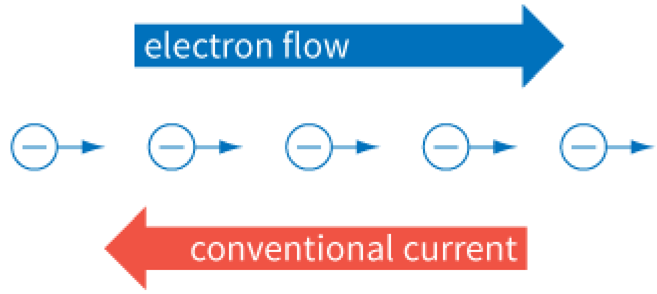


Figure 23.5: A flow of electrons to the right constitutes a conventional current to the left.

Charge on the plates

Think about a capacitor with uncharged plates. Each plate has equal amounts of positive and negative charge. Connecting the capacitor to a supply pulls charge $+Q$ from one plate and transfers it to the other, leaving behind charge $-Q$. The supply does work in separating the charges. Since the two plates now store equal and opposite charges, the total charge on the capacitor is zero. When we talk about the 'charge stored' by a capacitor, we mean the quantity Q, the magnitude of the charge stored on each plate.

To make the capacitor plates store more charge, we would have to use a supply of higher e.m.f. If we connect the leads of the charged capacitor together, electrons flow back around the circuit and the capacitor is discharged.

You can observe a capacitor discharging as follows. Connect the two leads of a capacitor to the terminals of a battery. Disconnect, and then reconnect the leads to a light-emitting diode (LED). It is best to have a protective resistor in series with the LED. The LED will glow briefly as the capacitor discharges.

In any circuit, the charge that flows past a point in a given time is equal to the area under a current–time graph (just as distance is equal to the area under a speed–time graph). So the magnitude of the charge on the plates in a capacitor is given by the area under the current–time graph recorded while the capacitor is being charged up.

The meaning of capacitance

If you look at some capacitors, you will see that they are marked with the value of their **capacitance**. The greater the capacitance, the greater is the charge on the capacitor plates for a given potential difference across it.

The capacitance C of a capacitor is defined by:

$$\text{capacitance} = \frac{\text{charge}}{\text{potential difference}}$$

$$C = \frac{Q}{V}$$

where Q is the magnitude of the charge on each of the capacitor's plates and V is the potential difference across the capacitor.

KEY EQUATION

$$\text{capacitance} = \frac{\text{charge}}{\text{potential difference}}$$

$$C = \frac{Q}{V}$$

The charge on the capacitor may be calculated using the equation:

$$Q = VC$$

This equation shows that the charge depends on two things: the capacitance C and the voltage V (double the voltage means double the charge). Note that it isn't only capacitors that have capacitance. Any object can become charged by connecting it to a voltage. The object's capacitance is then the ratio of the charge to the voltage.

Units of capacitance

The unit of capacitance is the **farad**, F. From the equation that defines capacitance, you can see that this must be the same as the units of charge (coulombs, C) divided by voltage (V):

$$1\ \text{F} = 1\ \text{C V}^{-1}$$

(It is unfortunate that the letter 'C' is used for both capacitance and coulomb. There is room for confusion here!)

In practice, a farad is a large unit. Few capacitors have a capacitance of 1 F. Capacitors usually have their values marked in picofarads (pF), nanofarads (nF) or microfarads (μF):

$1\ \text{pF} = 10^{-12}\ \text{F}$ $1\ \text{nF} = 10^{-9}\ \text{F}$ $1\ \mu\text{F} = 10^{-6}\ \text{F}$

Other markings on capacitors

Many capacitors are marked with their highest safe working voltage. If you exceed this value, charge may leak across between the plates, and the dielectric will cease to be an insulator. Some capacitors (electrolytic ones) must be connected correctly in a circuit. They have an indication to show which end must be connected to the positive of the supply. Failure to connect correctly will damage the capacitor, and can be extremely dangerous.

Questions

1. Calculate the charge on a 220 μF capacitor charged up to 15 V. Give your answer in microcoulombs (μC) and in coulombs (C).
2. A charge of 1.0×10^{-3} C is measured on a capacitor with a potential difference across it of 500 V. Calculate the capacitance in farads (F), microfarads (μF) and picofarads (pF).
3. Calculate the average current required to charge a 50 μF capacitor to a p.d. of 10 V in a time interval of 0.01 s.
4. A student connects an uncharged capacitor of capacitance C in series with a resistor, a cell and a switch. The student closes the switch and records the current I at intervals of 10 s. The results are shown in Table 23.1. The potential difference across the capacitor after 60 s is 8.5 V. Plot a current–time graph, and use it to estimate the value of C.

t / s	0	10	20	30	40	50	60
I / μA	200	142	102	75	51	37	27

Table 23.1 Data for Question 4.

23.2 Energy stored in a capacitor

When you charge a capacitor, you use a power supply to push electrons onto one plate and off the other. The power supply does work on the electrons, so their potential energy increases. You recover this energy when you discharge the capacitor.

If you charge a large capacitor (1000 μF or more) to a potential difference of 6.0 V, disconnect it from the supply and then connect it across a 6.0 V lamp, you can see the lamp glow as energy is released from the capacitor. The lamp will flash briefly. Clearly, such a capacitor does not store much energy when it is charged.

In order to charge a capacitor, work must be done to push electrons onto one plate and off the other (Figure 23.6). At first, there is only a small amount of negative charge on the left-hand plate. Adding more electrons is relatively easy, because there is not much repulsion. As the charge on the plate increases, the repulsion between the electrons on the plate and the new electrons increases, and a greater amount of work must be done to increase the charge on the plate.

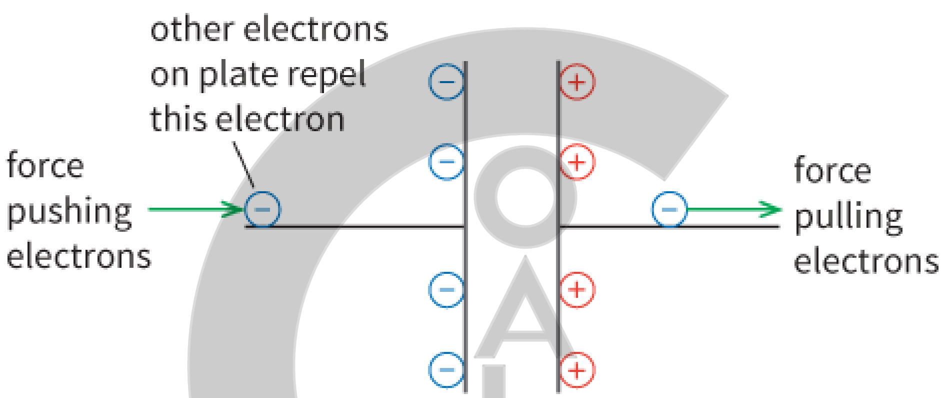


Figure 23.6: When a capacitor is charged, work must be done to push additional electrons against the repulsion of the electrons that are already present.

This can be seen qualitatively in Figure 23.7a. This graph shows how the p.d. V increases as the amount of charge Q increases. It is a straight line because Q and V are related by:

$$V = \frac{Q}{C}$$

We can use Figure 23.7a to calculate the work done in charging up the capacitor.

First, consider the work done W in moving charge Q through a constant p.d. V. This is given by:

$$W = QV$$

(You studied this equation in Chapter 9.) From the graph of Q against V (Figure 23.7b), we can see that the quantity $Q \times V$ is given by the area under the graph.

The area under a graph of p.d. against charge is equal to work done.

If we apply the same idea to the capacitor graph (Figure 23.7a), then the area under the graph is the shaded triangle, with an area of base × height. Hence, the work done in charging a capacitor to a particular p.d. is given by:

$$W = \frac{1}{2}QV$$

Substituting $Q = CV$ into this equation gives two further equations:

$$W = \frac{1}{2}CV^2$$

$$W = \frac{1}{2}\frac{Q^2}{C}$$

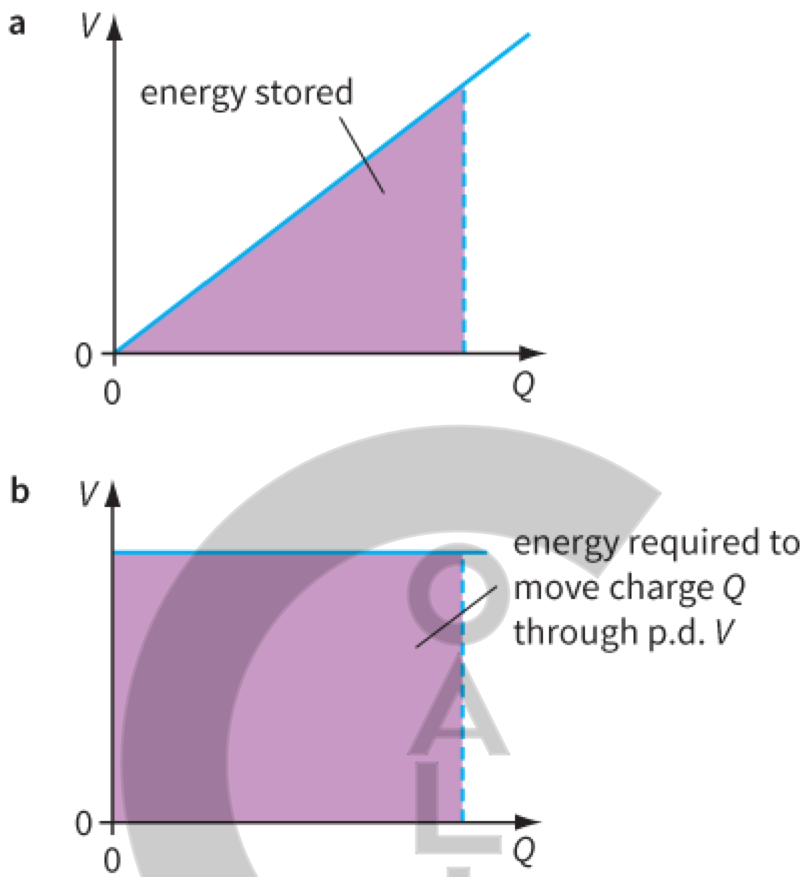


Figure 23.7: The area under a graph of voltage against charge gives a quantity of energy. The area in **a** shows the energy stored in a capacitor; the area in **b** shows the energy required to drive a charge through a resistor.

where *W* energy stored, *Q* is the charge on the capacitor, *C* is the capacitance and *V* is the potential difference across the capacitor.

These three equations show the work done in charging up the capacitor. This is equal to the energy stored by the capacitor, since this is the amount of energy released when the capacitor is discharged.

We can also see from the second formula $(W = \frac{1}{2}CV^2)$ that the energy *W* that a capacitor stores depends on its capacitance *C* and the potential difference *V* to which it is charged.

The energy *W* stored is proportional to the square of the potential difference *V* ($W \propto V^2$). It follows that doubling the charging voltage means that four times as much energy is stored.

KEY EQUATIONS

Work done by charging a capacitor:

$$W = \frac{1}{2}QV$$

$$W = \frac{1}{2}CV^2$$

$$W = \frac{1}{2}\frac{Q^2}{C}$$

WORKED EXAMPLE

1 A 2000 μF capacitor is charged to a p.d. of 10 V. Calculate the energy stored by the capacitor.

Step 1 Write down the quantities we know:

$$C = 2000\mu F$$
$$V = 10V$$

Step 2 Write down the equation for energy stored and substitute values:

$$W = \frac{1}{2}CV^2$$
$$= \frac{1}{2} \times 2000 \times 10^{-6} \times 10^2$$
$$= 0.10\ J$$

This is a small amount of energy – compare it with the energy stored by a rechargeable battery, typically of the order of 10 000 J. A charged capacitor will not keep an MP3 player running for any length of time.

Questions

5 State the quantity represented by the gradient of the straight line shown in Figure 23.7a.

6 The graph of Figure 23.8 shows how *V* depends on *Q* for a particular capacitor.

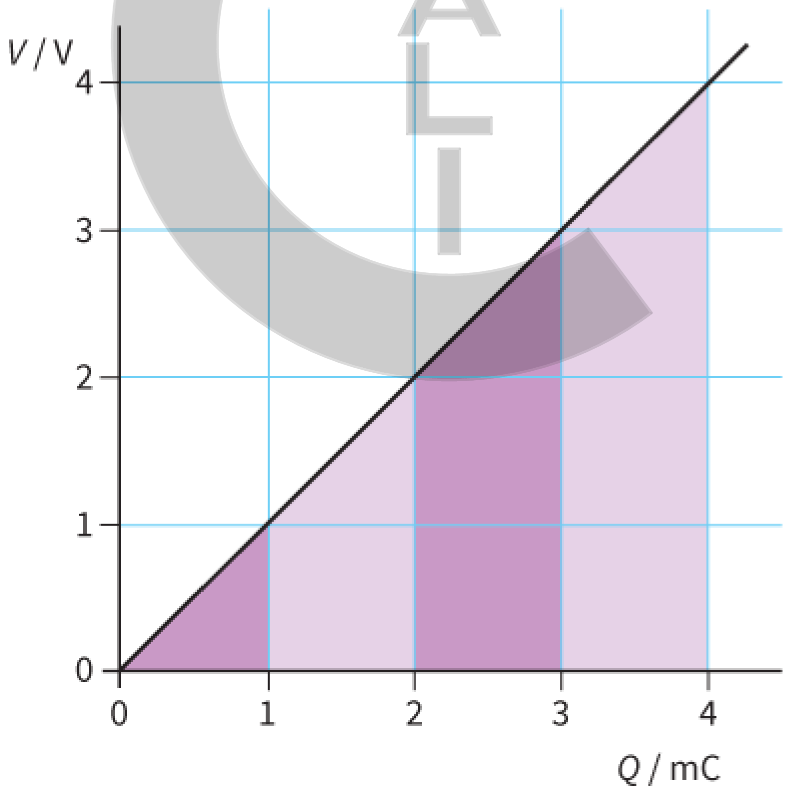


Figure 23.8: The energy stored by a capacitor is equal to the area under the graph of voltage against charge.

The area under the graph has been divided into strips to make it easy to calculate the energy stored. The first strip (which is simply a triangle) shows the energy stored when the capacitor is charged up to 1.0 V. The energy stored is:

$$\begin{aligned} \tfrac{1}{2}QV &= \tfrac{1}{2} \times 1.0\ \text{mC} \times 1.0\ \text{V} \\ &= 0.5\ \text{mJ} \end{aligned}$$

a Calculate the capacitance C of the capacitor.

b Copy Table 23.2 and complete it by calculating the areas of successive strips, to show how W depends on V.

c Plot a graph of W against V. Describe the shape of this graph.

Q / mC	V / V	Area of strip ΔW / mJ	Sum of areas W / mJ
1.0	1.0	0.5	0.5
2.0	2.0	1.5	2.0
3.0			
4.0			

Table 23.2 Data for Question 6.

PRACTICAL ACTIVITY 23.1

Investigating energy stored in a capacitor

If you have a sensitive joulemeter (capable of measuring millijoules, mJ), you can investigate the equation for energy stored. A suitable circuit is shown in Figure 23.9.

The capacitor is charged up when the switch connects it to the power supply. When the switch is altered, the capacitor discharges through the joulemeter. (It is important to wait for the capacitor to discharge completely.) The joulemeter will measure the amount of energy released by the capacitor.

By using capacitors with different values of C, and by changing the charging voltage V, you can investigate how the energy W stored depends on C and V.

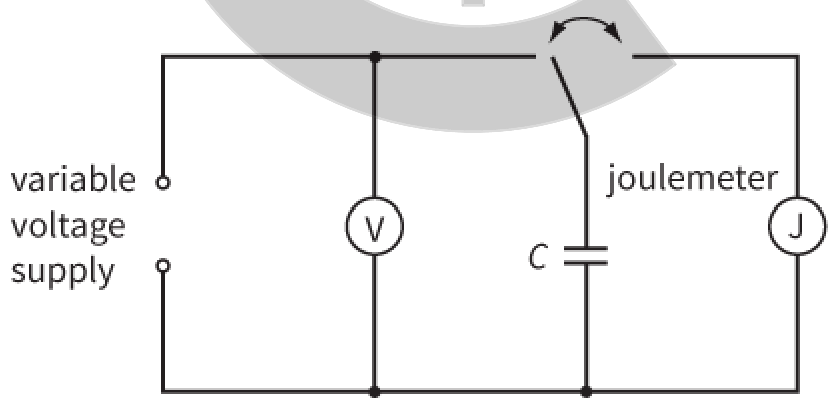


Figure 23.9: With the switch to the left, the capacitor C charges up; to the right, it discharges through the joulemeter.

Questions

7 Calculate the energy stored in the following capacitors:

a a 5000 μF capacitor charged to 5.0 V

b a 5000 pF capacitor charged to 5.0 V

c a 200 μF capacitor charged to 230 V.

8 Which involves more charge, a 100 μF capacitor charged to 200 V or a 200 μF capacitor charged to 100 V? Which stores more energy?

9 A 10 000 μF capacitor is charged to 12 V, and then connected across a lamp rated at '12 V, 36 W'.

a Calculate the energy stored by the capacitor.

b Estimate the time the lamp stays fully lit. Assume that energy is dissipated in the lamp at a steady rate.

10 In a simple photographic flashgun, a 0.20 F capacitor is charged by a 9.0 V battery. It is then discharged in a flash of duration 0.01 s. Calculate:

a the charge on and energy stored by the capacitor

b the average power dissipated during the flash

c the average current in the flash bulb

d the approximate resistance of the bulb.

23.3 Capacitors in parallel

Capacitors are used in electric circuits to store energy. Situations often arise where two or more capacitors are connected together in a circuit. In this topic, we will look at capacitors connected in parallel. The next topic deals with capacitors in series.

When two capacitors are connected in parallel (Figure 23.10), their combined or total capacitance C_{total} is simply the sum of their individual capacitances C_1 and C_2:

$$C_{total} = C_1 + C_2$$

This is because, when two capacitors are connected together, they are equivalent to a single capacitor with larger plates. The bigger the plates, the more charge that can be stored for a given voltage, and hence the greater the capacitance.

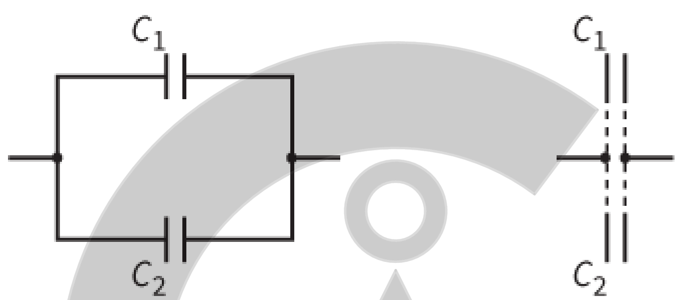


Figure 23.10: Two capacitors connected in parallel are equivalent to a single, larger capacitor.

The total charge Q on two capacitors connected in parallel and charged to a potential difference V is simply given by:

$$Q = C_{total} \times V$$

For three or more capacitors connected in parallel, the equation for their total capacitance becomes:

$$C_{total} = C_1 + C_2 + C_3 + \ldots$$

Capacitors in parallel: deriving the formula

We can derive the equation for capacitors in parallel by thinking about the charge on the two capacitors. As shown in Figure 23.11, C_1 stores charge Q_1 and C_2 stores charge Q_2. Since the p.d. across each capacitor is V, we can write:

$$Q_1 = C_1V \text{ and } Q_2 = C_2V \;|$$

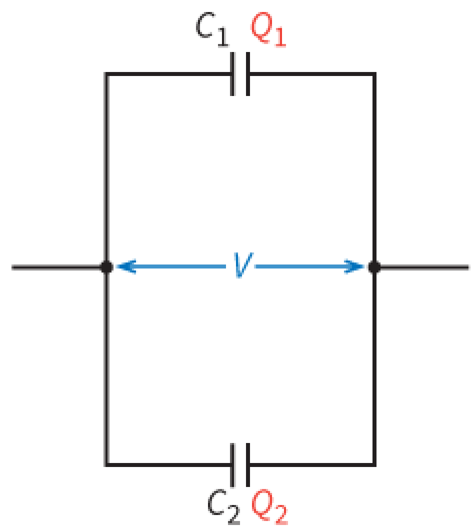


Figure 23.11: Two capacitors connected in parallel have the same p.d. across them, but different amounts of charge.

The total charge is given by the sum of these:

$$Q = Q_1 + Q_2 = C_1V + C_2V$$

Since V is a common factor:

$$Q = (C_1 + C_2)V$$

Comparing this with $Q = C_{total}V$ gives the required $C_{total} = C_1 + C_2$. It follows that for three or more capacitors connected in parallel, we have:

$$C_{total} = C_1 + C_2 + C_3 + \ldots$$

Capacitors in parallel: summary

For capacitors in parallel, the following rules apply:

- The p.d. across each capacitor is the same.
- The total charge on the capacitors is equal to the sum of the charges:

$$Q_{total} = Q_1 + Q_2 + Q_3 + \ldots$$

- The total capacitance C_{total} is given by:

$$C_{total} = C_1 + C_2 + C_3 + \ldots$$

KEY EQUATION

$$C_{total} = C_1 + C_2 + C_3 + \ldots$$

The combined capacitance of capacitors in parallel.

You must learn how to derive this equation.

Questions

11 **a** Calculate the total capacitance of two 100 μF capacitors connected in parallel.

b Calculate the total charge they store when charged to a p.d. of 20 V.

12 A capacitor of capacitance 50 μF is required, but the only values available to you are 10 μF, 20 μF and 100 μF (you may use more than one of each value). How would you achieve the required value by connecting capacitors in parallel? Give at least two answers.

23.4 Capacitors in series

In a similar way to the case of capacitors connected in parallel, we can consider two or more capacitors connected in series (Figure 23.12). The total capacitance C_{total} of two capacitors of capacitances C_1 and C_2 is given by:

$$\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2}$$

Here, it is the reciprocals of the capacitances that must be added to give the reciprocal of the total capacitance. For three or more capacitors connected in series, the equation for their total capacitance is:

$$\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$$

Figure 23.12: Two capacitors connected in series.

Capacitors in series: deriving the formula

The same principles apply here as for the case of capacitors in parallel. Figure 23.13 shows the situation. C_1 and C_2 are connected in series, and there is a p.d. V across them. This p.d. is divided (it is shared between the two capacitors), so that the p.d. across C_1 is V_1 and the p.d. across C_2 is V_2. It follows that:

$$V = V_1 + V_2$$

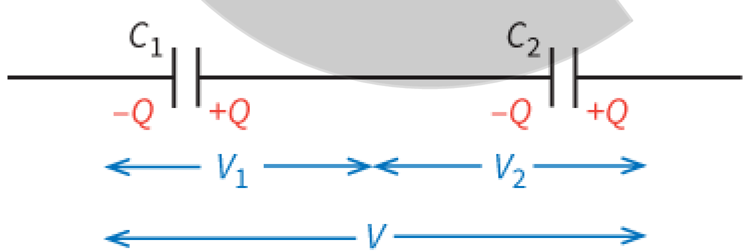


Figure 23.13: Capacitors connected in series store the same charge, but they have different p.d.s across them.

Now we must think about the charge stored by the combination of capacitors. In Figure 23.13, you will see that both capacitors are shown as storing the same charge Q. How does this come about? When the voltage is first applied, charge $-Q$ arrives on the left-hand plate of C_1. This repels charge $-Q$ off the right-hand plate, leaving it with charge $+Q$. Charge $-Q$ now arrives on the left-hand plate of C_2, and this in turn results in charge $+Q$ on the right-hand plate.

KEY EQUATION

$$\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$$

The combined capacitance of capacitors in series.

You must learn how to derive this equation.

Note that charge is not arbitrarily created or destroyed in this process – the total amount of charge in the system is constant. This is an example of the conservation of charge.

Notice also that there is a central isolated section of the circuit between the two capacitors. Since this is initially uncharged, it must remain so at the end. This requirement is satisfied, because there is charge $-Q$ at one end and $+Q$ at the other. Hence, we conclude that capacitors connected in series store the same charge. This allows us to write equations for V_1 and V_2:

$$V_2 = \frac{Q}{C_1} \text{ and } V_2 = \frac{Q}{C_2}$$

The combination of capacitors stores charge Q when charged to p.d. V, and so we can write:

$$V = \frac{Q}{C_{total}}$$

Substituting these in $V = V_1 + V_2$ gives:

$$\frac{Q}{C_{total}} + \frac{Q}{C_1} + \frac{Q}{C_2}$$

Cancelling the common factor of Q gives the required equation:

$$\frac{1}{C_{total}} + \frac{1}{C_1} + \frac{1}{C_2}$$

Worked example 2 shows how to use this relationship.

WORKED EXAMPLE

2 Calculate the total capacitance of a 300 μF capacitor and a 600 μF capacitor connected in series.

Step 1 The calculation should be done in two steps; this is relatively simple using a calculator with a $\frac{1}{x}$ or x^{-1} key.

Substitute the values into the equation:

$$\frac{1}{C_{total}} + \frac{1}{C_1} + \frac{1}{C_2}$$

This gives:

$$\frac{1}{C_{total}} = \frac{1}{300} + \frac{1}{600}$$

$$\frac{1}{C_{total}} = 0.005\mu F^{-1}$$

Step 2 Now take the reciprocal of this value to determine the capacitance in μF:

$$C_{total} = \frac{1}{0.005} = 200 \text{ F}$$

Notice that the total capacitance of two capacitors in series is less than either of the individual capacitances.

Using the x^{-1} key on your calculator, you can also do this calculation in one step:

$$C_{total} = (300^{-1} + 600^{-1})^{-1} = 200\ \mu\text{F}$$

Questions

13 Calculate the total capacitance of three capacitors of capacitances 200 μF, 300 μF and 600 μF, connected in series.

14 You have a number of identical capacitors, each of capacitance *C*. Determine the total capacitance when:

- **a** two capacitors are connected in series
- **b** *n* capacitors are connected in series
- **c** two capacitors are connected in parallel
- **d** *n* capacitors are connected in parallel.

23.5 Comparing capacitors and resistors

It is helpful to compare the formulae for capacitors in series and parallel with the corresponding formulae for resistors (Table 23.3).

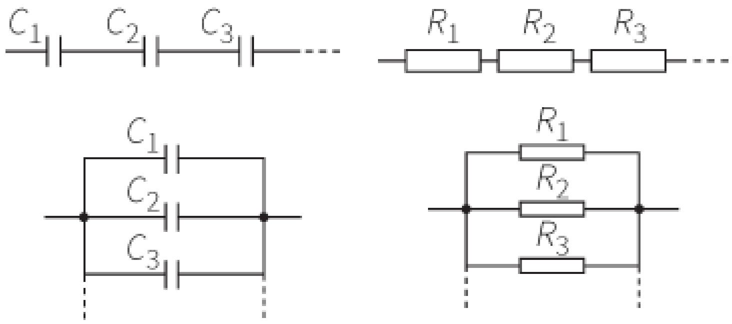


Table 23.3 Capacitors and resistors compared.

Notice that the reciprocal formula applies to capacitors in series but to resistors in parallel. This comes from the definitions of capacitance and resistance. Capacitance indicates how good a capacitor is at storing charge for a given voltage, and resistance indicates how bad a resistor is at letting current through for a given voltage.

23.6 Capacitor networks

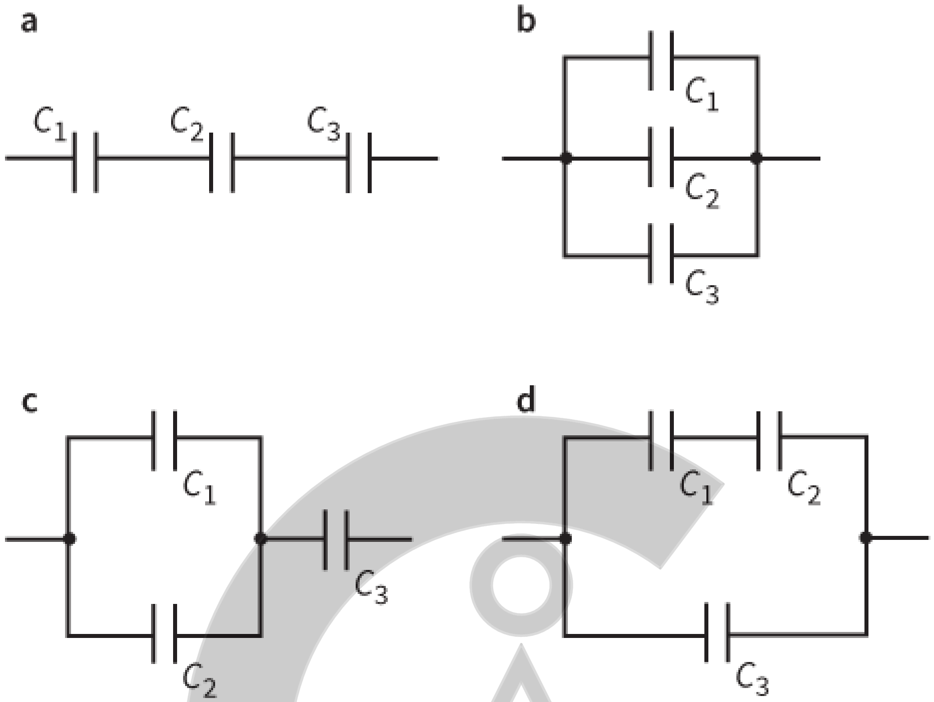


Figure 23.14: Four ways to connect three capacitors.

There are four ways in which three capacitors may be connected together. These are shown in Figure 23.14. The combined capacitance of the first two arrangements (three capacitors in series, three in parallel) can be calculated using the formulae. The other combinations must be dealt with in a different way:

- Figure 23.14a – All in series. Calculate C_{total} as in Table 23.3.
- Figure 23.14b – All in parallel. Calculate C_{total} as in Table 23.3.
- Figure 23.14c – Calculate C_{total} for the two capacitors of capacitances C_1 and C_2, which are connected in parallel, and then take account of the third capacitor of capacitance C_3, which is connected in series.
- Figure 23.14d – Calculate C_{total} for the two capacitors of capacitances C_1 and C_2, which are connected in series, and then take account of the third capacitor of capacitance C_3, which is connected in parallel.

These are the same approaches as would be used for networks of resistors.

Questions

15 For each of the four circuits shown in Figure 23.14, calculate the total capacitance in μF if each capacitor has capacitance 100 μF.

16 Given a number of 100 μF capacitors, how might you connect networks to give the following values of capacitance:

a 400 μF?

b 25 μF?

c 250 μF?

(Note that, in each case, there is more than one correct answer; try to find the answer that requires the minimum number of capacitors.)

17 You have three capacitors of capacitances 100 pF, 200 pF and 600 pF. Determine the maximum and minimum values of capacitance that you can make by connecting them together to form a network. State how they should be connected in each case.

18 Calculate the capacitance in μF of the network of capacitors shown in Figure 23.15.

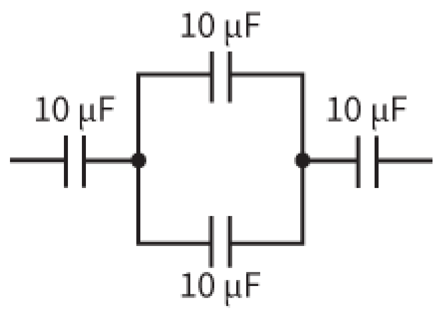


Figure 23.15: A capacitor network. For Question 18.

Sharing charge, sharing energy

If a capacitor is charged and then connected to a second capacitor (Figure 23.16), what happens to the charge and the energy that it stores? Note that, when the capacitors are connected together, they are in parallel, because they have the same p.d. across them. Their combined capacitance C_{total} is equal to the sum of their individual capacitances.

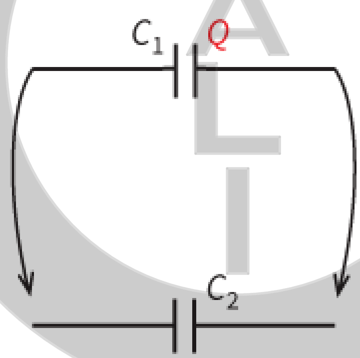


Figure 23.16: Capacitor of capacitance C_1 is charged and then connected across C_2.

Now we can think about the charge stored, Q. This is shared between the two capacitors; the total amount of charge stored must remain the same, since charge is conserved. The charge is shared between the two capacitors in proportion to their capacitances. Now the p.d. can be calculated from $V = \frac{Q}{C}$ and the energy from $W = \frac{1}{2}CV^2$

If we look at a numerical example, we find an interesting result (Worked example 3).

Figure 23.17 shows an analogy to the situation described in Worked example 3.

Capacitors are represented by containers of water. A wide (high capacitance) container is filled to a certain level (p.d.). It is then connected to a container with a smaller capacitance, and the levels equalise. (The p.d. is the same for each.) Notice that the potential energy of the water has decreased, because the height of its centre of gravity above the base level has decreased. Energy is dissipated as heat, as there is friction both within the moving water and between the water and the container.

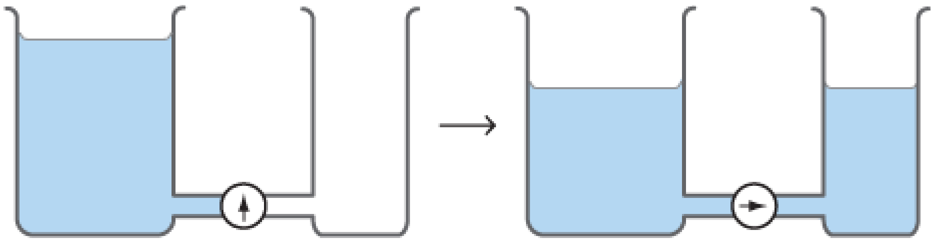


Figure 23.17: An analogy for the sharing of charge between capacitors.

WORKED EXAMPLE

3 Consider two 100 mF capacitors. One is charged to 10 V, disconnected from the power supply, and then connected across the other. Calculate the energy stored by the combination.

Step 1 Calculate the charge and energy stored for the single capacitor.

$$\text{initial charge } Q = VC = 10 \times 0.10 = 1.0\,\text{C}$$

$$\text{initial stored energy} = \tfrac{1}{2}CV^2 = \tfrac{1}{2} \times 0.10 \times 10^2 = 5.0\,\text{J}$$

Step 2 Calculate the final p.d. across the capacitors. The capacitors are in parallel and have a total stored charge of 1.0 C.

$$C_{total} = C_1 + C_2 = 100 + 100 = 200\,\text{mF}$$

The p.d. V can be determined using $Q = VC$.

$$V = \frac{Q}{C} = \frac{1.0}{200} \times 10^{-3} = 5.0\,\text{V}$$

This is because the charge is shared equally, with the original capacitor losing half of its charge.

Step 3 Now calculate the total energy stored by the capacitors.

$$\text{total energy} = \tfrac{1}{2}CV^2 = \tfrac{1}{2} \times 200 \times 10^{-3} \times 5.0^2 = 2.5\,\text{J}$$

The charge stored remains the same, but half of the stored energy is lost. The energy goes to heating the connecting wires as the electrons migrate between the capacitors.

Questions

19 Three capacitors, each of capacitance 120 μF, are connected together in series. This network is then connected to a 10 kV supply. Calculate:

a their combined capacitance in μF

b the charge stored

c the total energy stored.

20 A 20 μF capacitor is charged up to 200 V and then disconnected from the supply. It is then connected across a 5.0 μF capacitor. Calculate:

a the combined capacitance of the two capacitors in μF

b the charge they store

c the p.d. across the combination

d the energy dissipated when they are connected together.

Capacitance of isolated bodies

It is not just capacitors that have capacitance – all bodies have capacitance. Yes, even you have capacitance! You may have noticed that, particularly in dry conditions, you may become charged up, perhaps by rubbing against a synthetic fabric. You are at a high voltage and store a significant amount of charge. Discharging yourself by touching an earthed metal object would produce a spark.

If we consider a conducting sphere of radius r insulated from its surroundings and carrying a charge Q it will have a potential at its surface of V, where

$$V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r}$$

Since $C = \frac{Q}{V}$ it follows that the capacitance of a sphere is $C = 4\pi\varepsilon_0 r$.

Question

21 Estimate the capacitance of the Earth given that it has a radius of 6.4×10^6 m. State any assumptions you make.

23.7 Charge and discharge of capacitors

In Figure 23.18, the capacitor is charged by the battery when the switch is connected to terminal P. When first connected to P, a current is observed in the microammeter. The current starts off quite large and gradually decreases to zero. When connected to terminal Q, the capacitor discharges through the resistor and a current in the opposite direction is observed. As with the previous current, it starts off large and gradually falls to zero.

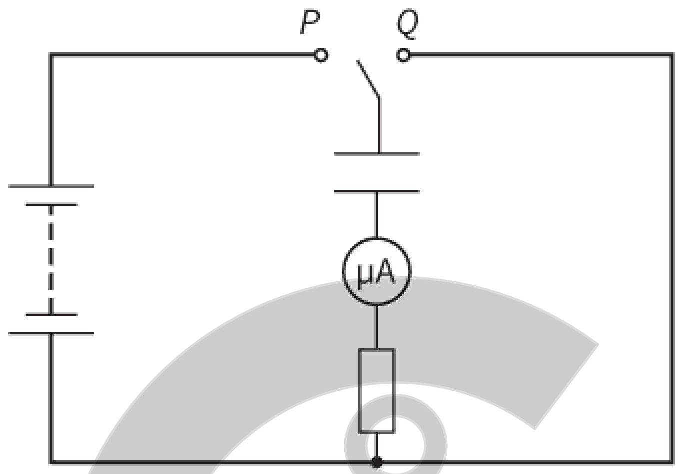


Figure 23.18 A circuit to charge and discharge a capacitor.

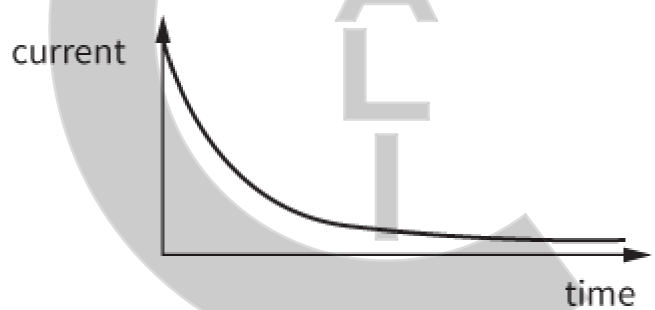


Figure 23.19 A graph showing how the current changes with time when a capacitor discharges through a resistor.

This shape of this graph it is quite common in sciences and it occurs in different situations – you will come across it again in radioactive decay in Chapter 29. In this case, it comes from the fact that, as charge flows off the capacitor, the potential difference reduces and so the current (the charge flowing per unit time) in the circuit also decreases. In radioactive decay, it occurs because as atoms decay, there are fewer atoms left to Charles's law and, therefore, fewer decays per unit time.

This type of decay is called **exponential decay** and is described by the formula:

$$x = x_0 e^{-ky}$$

where x is the dependent variable, y is the independent variable, k and x_0 are constants and e is the exponential function (a naturally occurring number of value 2.7118 28 …).

Question

22 In the circuit in Figure 23.18, the resistance has a resistance of 2000 Ω, the capacitor has a capacitance of 1000 μF and the battery has an e.m.f. of 12 V.

a Calculate:

i the potential difference across the capacitor when it is fully charged by the battery

ii the charge stored by the capacitor when it is fully charged

iii the current in the resistor when the switch is first connected to terminal Q.

b Explain what happens to the amount of charge stored on the plates in the moments after the switch is first connected to terminal Q.

c Based on your answer to part b, explain what effect this has on:

i the potential difference across the capacitor

ii the current in the resistor.

Once you have worked through Question 22, you should understand why the current gradually reduces: it reduces because of the current itself, as it takes charge off the plates.

What is the effect of changing the resistance in the circuit? There will be no change in the initial potential difference across the capacitor, but the initial current through the resistor will be changed. Increased resistance will mean decreased current, so charge flows off the capacitor plates more slowly and, therefore, the capacitor will take longer to discharge. Conversely, decreasing the resistance will cause the capacitor to discharge more quickly.

What is the effect of increasing the capacitance of the capacitor? The initial p.d. across the capacitor is, again, unchanged. So, with an unchanged resistance, the initial current will be unchanged. However, there will be more charge on the capacitor and so it will take longer to discharge.

From this, we can see that the time taken for a capacitor to discharge depends on both the capacitance and the resistance in the circuit. The quantity *RC* is called the **time constant** of the circuit. It is written using the Greek letter tau (τ).

KEY EQUATION

$$\tau = RC$$

Time constant for a capacitor discharging.

Question

23 Show that the unit of the time constant (*RC*) is the second.

The equation for the exponential decay of charge on a capacitor is:

$$I = I_0 \exp\left(-\frac{t}{RC}\right)$$

where I is the current, I_0 is the initial current, t is time and RC is the time constant.

The current at any time is directly proportional to the potential difference across the capacitor, which in turn is directly proportional to charge across the plate. The equation also describes the change in the potential difference and the charge on the capacitor.

So:

$$V = V_0 \exp\left(-\frac{t}{RC}\right)$$

where V is the p.d, and V_0 is the initial p.d.

And:

$$Q = Q_0 \exp\left(-\frac{t}{RC}\right)$$

where Q is the charge and Q_0 is the initial charge.

KEY EQUATIONS

Exponential decay of charge on a capacitor:

$$I = I_0 \exp\left(-\frac{t}{RC}\right)$$

$$Q = Q_0 \exp\left(-\frac{t}{RC}\right)$$

$$V = V_0 \exp\left(-\frac{t}{RC}\right)$$

WORKED EXAMPLE

4 The potential difference across the plates of a capacitor of capacitance 500 μF is 240 V. The capacitor is connected across the terminals of a 600 Ω resistor.

Find the time taken for the current to fall to 0.10 A.

Step 1 Calculate the initial current:

$$I_0 = \frac{V}{R} = \frac{240}{600} = 0.40 \text{ A}$$

Step 2 Calculate the time constant:

$$\tau = RC = 600 \times 500 \times 10^{-6} = 0.30 \text{ s}$$

Step 3 Substitute into the equation:

$$I = I_0 \exp\left(-\frac{t}{RC}\right)$$

$$0.10 = 0.40 \exp\left(-\frac{t}{0.30}\right)$$

$$\frac{0.10}{0.40} = \exp\left(-\frac{t}{0.30}\right)$$

$$0.25 = \exp\left(-\frac{t}{0.30}\right)$$

Step 4 *e* comes from the antilog of the natural logarithm (ln) such that ln (e^x) = x

Taking ln of both sides:

$$\ln 0.25 = -\frac{t}{0.30}$$

$$-1.386 = -\frac{t}{0.30}$$

$$t = 1.386 \times 0.30 = 0.41 \text{ s}$$

Question

24 A 400 μF capacitor is charged using a 20 V battery. It is connected across the ends of a 600 Ω resistor with 20 V potential difference across its plates.

- **a** Calculate the charge stored on the capacitor.
- **b** Calculate the time constant for the discharging circuit.
- **c** Calculate the time it takes the charge on the capacitor to fall to 2.0 mC.
- **d** State the potential difference across the plates when the charge has fallen to 2.0 mC.

REFLECTION

In Worked example 3, we showed that when a charged capacitor is connected to an identical uncharged capacitor, half the energy is dissipated in driving the charge through the circuit and is transformed to thermal energy. If we had superconducting connectors – ones that conduct electricity without any energy losses – what would happen? Discuss with a partner.

What did you find satisfying about discussing this problem?

SUMMARY

Capacitors are constructed from two metal sheets ('plates'), separated by an insulating material. A capacitor stores equal and opposite amounts of charge on its plates.

For a capacitor, the charge stored is directly proportional to the p.d. between the plates:

$$Q = VC$$

Capacitance is the charge stored per unit of p.d.

A farad is a coulomb per volt: 1 F = 1 C V^{-1}.

Capacitors store energy. The energy W stored at p.d. V is:

$$W = \frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C}$$

The formula $W = \frac{1}{2}QV$ is deduced from the area under a graph of potential difference against charge.

For capacitors connected in parallel and in series, the combined capacitances are:

parallel: $C_{total} = C_1 + C_2 + C_3 + \ldots$

series: $\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$

These formulae are derived from conservation of charge and addition of p.d.s.

The graphs for the discharge current, charge stored and potential difference across a capacitor are all examples of exponential decay.

The time constant for circuits containing capacitance and resistance is: $\tau = CR$

The graphs of discharge current, charge stored and potential difference across a capacitor are all of the form:

$$x = x_0 \exp\left(-\frac{t}{RC}\right)$$

EXAM-STYLE QUESTIONS

1 A capacitor has a potential difference of 6.0 V across its plates and stores 9.0 mJ of energy.

Which row in the table gives the capacitance of the capacitor and the charge on its plates? **[1]**

	Capacitance / μF	Charge / mC
A	500	3.0
B	500	18
C	3000	3.0
D	3000	18

Table 23.3

2 A capacitor in an electronic circuit is designed to slowly discharge through an indicator lamp.

It is decided that the time taken for the capacitor to discharge needs to be increased. Four changes are suggested:

1 Connect a second capacitor in parallel with the original capacitor.

2 Connect a second capacitor in series with the original capacitor.

3 Connect a resistor in parallel with the lamp.

4 Connect a resistor in series with the lamp.

Which suggestions would lead to the discharge time being increased? **[1]**

A 1 and 3 only

B 1 and 4 only

C 2 and 3 only

D 2 and 4 only

3 A 470 μF capacitor is connected across the terminals of a battery of e.m.f. 9 V. Calculate the charge on the plates of the capacitor. **[1]**

4 Calculate the p.d. across the terminals of a 2200 μF capacitor when it has a charge of 0.033 C on its plates. **[1]**

5 Calculate the capacitance of a capacitor if it stores a charge of 2.0 C when there is a potential difference of 5000 V across its plates. **[1]**

6 Calculate the energy stored when a 470 μF capacitor has a potential difference of 12 V across its plates. **[1]**

7 Calculate the energy stored on a capacitor if it stores 1.5 mC of charge when there is a potential difference of 50 V across it. **[1]**

8 A 5000 μF capacitor has a p.d. of 24 V across its plates.

a Calculate the energy stored on the capacitor. **[1]**

b The capacitor is briefly connected across a bulb and half the charge flows off the capacitor. Calculate the energy dissipated in the lamp. **[3]**

[Total: 4]

9 A 4700 μF capacitor has a p.d. of 12 V across its terminals. It is connected to a resistor and the charge leaks away through the resistor in 2.5 s.

a Calculate the energy stored on the capacitor. **[1]**

b Calculate the charge stored on the capacitor. **[1]**

c Estimate the average current through the resistor. **[1]**

d Estimate the resistance of the resistor. **[2]**

e Suggest why the last two quantities can only be estimates. **[1]**

[Total: 6]

10 An electronics engineer is designing a circuit in which a capacitor of capacitance of 4700 μF is to be connected across a potential difference of 9.0 V. He has four 4700 μF, 6 V capacitors available. Draw a diagram to show how the four capacitors could be used for this purpose. **[1]**

11 Calculate the different capacitances that can be made from three 100 μF capacitors. For each value, draw the network that is used. **[4]**

12 This diagram shows three capacitors connected in series with a cell of e.m.f. 1.5 V.

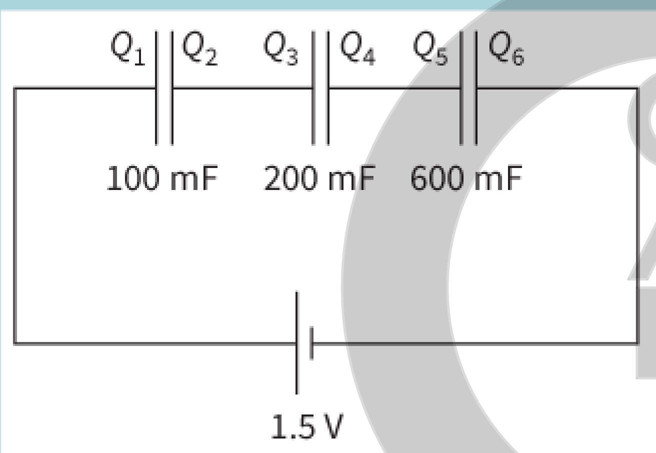


Figure 23.20

a Calculate the charges Q_1 to Q_6 on each of the plates. **[5]**

b Calculate the p.d. across each capacitor. **[3]**

[Total: 8]

13 a State one use of a capacitor in a simple electric circuit. **[1]**

b This is a circuit used to investigate the discharge of a capacitor, and a graph showing the change in current with time when the capacitor is discharged.

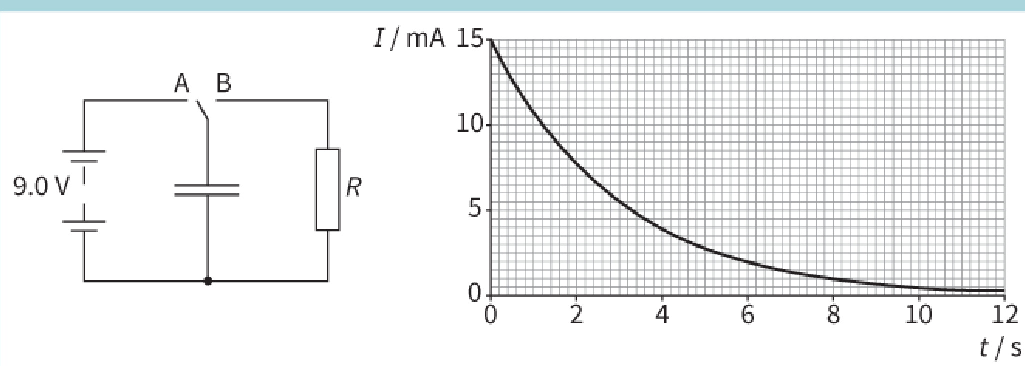


Figure 23.21

i Deduce the resistance R of the resistor. **[2]**

ii Explain why the current decreases as the capacitor discharges. **[2]**

iii The charge on the capacitor is equal to the area under the graph. Estimate the charge on the capacitor when the potential difference across it is 9.0 V. **[2]**

iv Calculate the capacitance of the capacitor. **[2]**

[Total: 9]

14 The spherical dome on a Van de Graaff generator has a diameter of 40 cm and the potential at its surface is 5.4 kV.

a **i** Calculate the charge on the dome. **[2]**

ii Calculate the capacitance of the dome. **[2]**

An earthed metal plate is moved slowly towards the sphere but does not touch it. The sphere discharges through the air to the plate. This graph shows how the potential at the surface of the sphere changes during the discharge.

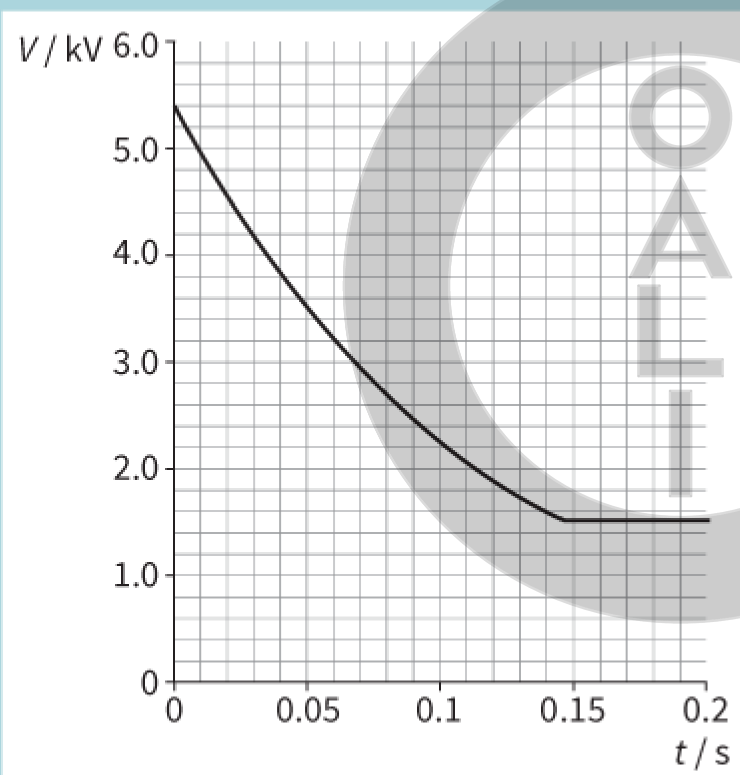


Figure 23.22

b Calculate the energy that is dissipated during the discharge. **[5]**

c Suggest why the discharge ceases while there is still some charge on the dome. **[2]**

[Total: 11]

15 a Show that the capacitance C of an isolated conducting sphere of radius r is given by the formula:

$C = 4\pi\varepsilon_0 r$ **[2]**

This diagram shows two identical conducting brass spheres of radius 10Cm mounted on insulating stands. Sphere A has a charge of +5.0 × 10^{-8} C and sphere B is uncharged.

Figure 23.23

b **i** Calculate the potential at the surface of sphere A. **[2]**

ii Calculate the energy stored on sphere A. **[2]**

Sphere B is brought up to sphere A and is touched to it so that the charge is shared between the two spheres, before being taken back to its original position.

c **i** Calculate the energy stored on each sphere. **[3]**

ii Suggest why there is a change in the total energy of the system. **[1]**

[Total: 10]

16 **a** Define the term capacitance of a capacitor. **[2]**

b This is a circuit that can be used to measure the capacitance of a capacitor.

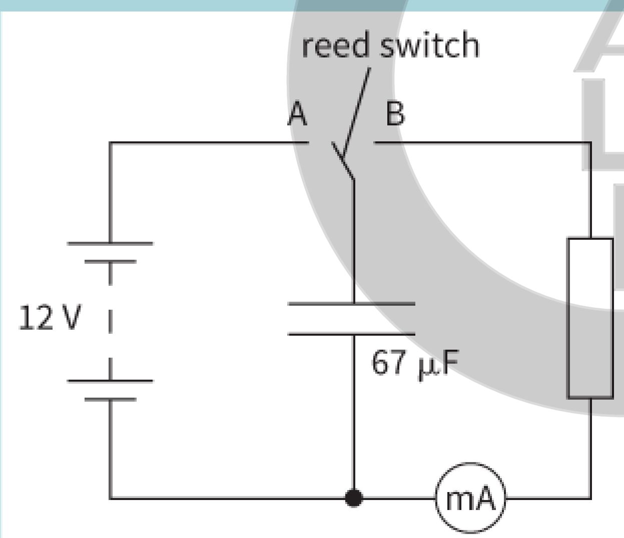


Figure 23.24

The reed switch vibrates back and forth at a frequency of 50 Hz. Each time it makes contact with A, the capacitor is charged by the battery so that there is a p.d. of 12 V across it. Each time it makes contact with B, it is fully discharged through the resistor.

i Calculate the charge that is stored on the capacitor when there is a p.d. of 12 V across it. **[2]**

ii Calculate the average current in the resistor. **[2]**

iii Calculate the average power dissipated in the resistor. **[3]**

c A second capacitor of the same value is connected in series with the first capacitor.

Discuss the effect on both the current recorded and the power dissipated in the resistor. **[4]**

[Total: 13]

17 a Explain what is meant by the time constant of a circuit containing capacitance and resistance. **[2]**

b A circuit contains capacitors of capacitance 500 μF and 2000 μF in series with each other and in series with a resistance of 2.5 kΩ.

i Calculate the effective capacitance of the capacitors in series. **[2]**

ii Calculate the charge on the capacitor plates when there is a potential difference of 50 V across the plates. **[2]**

iii Calculate the time taken for the charge on the plates to fall to 5% of the charge when there was a p.d. of 50 V across the plates. **[2]**

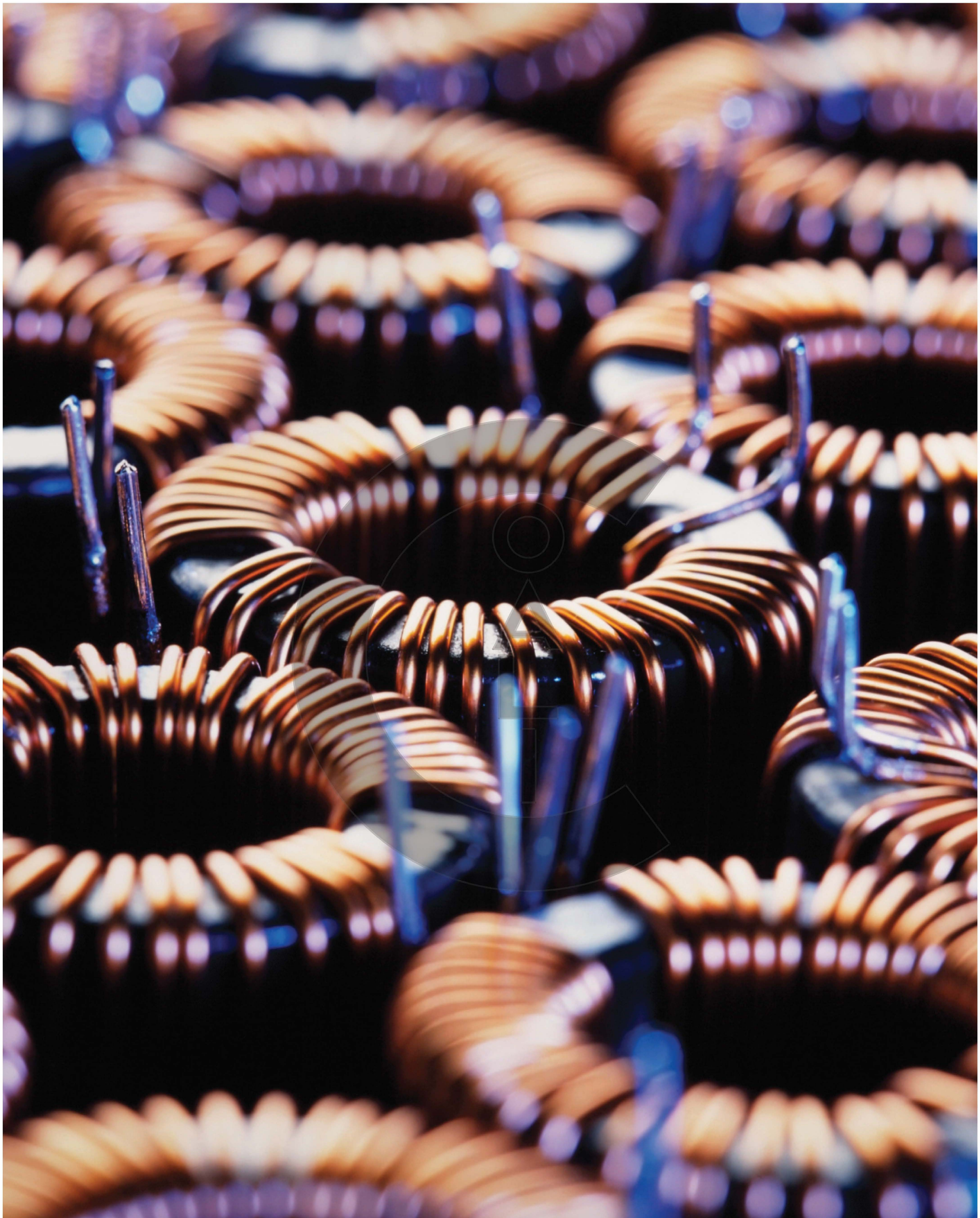
[Total: 8]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define capacitance of both capacitors and spherical conductors	23.1			
recall and use the formula: $C = \frac{Q}{V}$	23.1			
recognise that the unit of capacitance is the farad (F)	23.1			
derive and use the formula: $C_{total} = C_1 + C_2 + C_3 + \ldots$ for capacitors in parallel	23.3			
recognise that the time constant for circuits containing capacitance and resistance is: $\tau = CR$	23.7			
derive and use the formula: $\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$ for capacitors in series	23.4			
determine the energy stored in a capacitor from the potential–charge graph	23.2			
recall and use the formulae: $\frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C}$	23.2			
understand that the decay of charge on a capacitor, the discharge current and the potential difference across the plates are exponential decays	23.7			
recognise and understand that the rate of discharge is dependent on the time constant of the circuit	23.7			

I can	See topic…	Needs more work	Almost there	Ready to move on
recall and use the equation: $x = x_0 \exp\left(-\frac{t}{RC}\right)$ for potential difference, discharge current and charge on the plates of a capacitor.	23.7			



〉Chapter 24

Magnetic fields and electromagnetism

LEARNING INTENTIONS

In this chapter you will learn how to:

- describe a magnetic field as an example of a field of force caused by moving charges or permanent magnets
- use field lines to represent a field and sketch various patterns
- determine the size and direction of the force on a current-carrying conductor in a magnetic field
- define magnetic flux density and know how it can be measured
- explain the origin of the forces between current-carrying conductors and find the direction of these forces.

BEFORE YOU START

- Can you describe how to plot a magnetic field using a compass and iron filings?
- What effect do like poles and unlike poles have on each other?
- Write down definitions for gravitational field and an electric field. Swap your definitions with a partner.
- With your partner, discuss how charge and current are related.

MAGNETS AND CURRENTS

The patient shown in Figure 24.1 is about to undergo a magnetic resonance imaging (MRI) scan. The patient is placed in a magnetic field created by solenoid, or long coil, containing many turns of wire. A very strong magnetic field is created in these coils by a high current. Most of these coils are made from superconducting materials (materials with zero resistivity).

Why do you think that iron objects, such as scissors and gas cylinders, must not be taken into the same room as this machine?

What advantages are there in using a superconducting material for the wires in the coil?

In this chapter, we will look at magnetic forces and fields, how they arise and how they interact.

Figure 24.1: A patient about to have an MRI scan.

24.1 Producing and representing magnetic fields

A magnetic field exists wherever there is force on a magnetic pole. As we saw with electric and gravitational fields, a magnetic field is a field of force.

You can make a magnetic field in two ways: using a permanent magnet, or using the movement of electric charges, usually by having an electric current. You should be familiar with the magnetic field patterns of bar magnets (Figure 24.2). These can be shown using iron filings or plotting compasses.

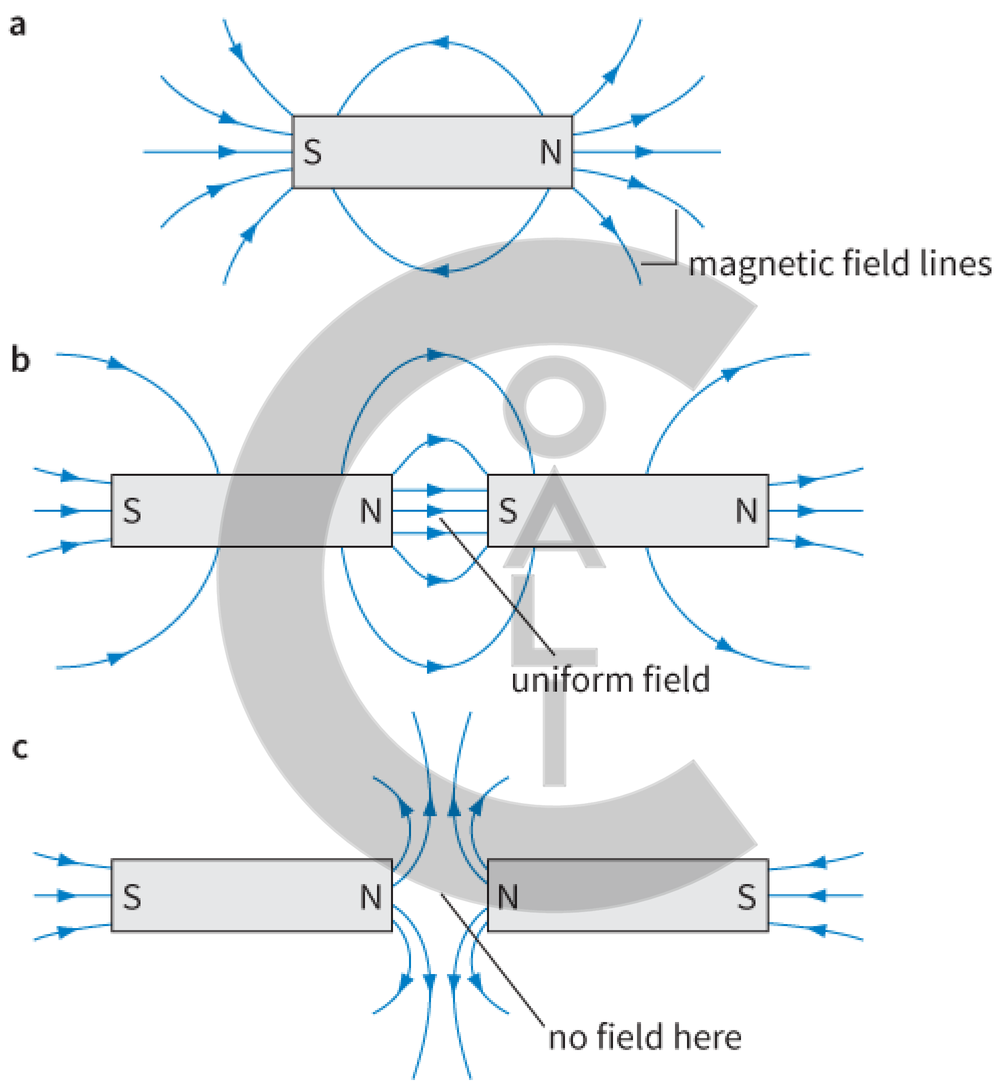


Figure 24.2: Magnetic field patterns: **a** for a bar magnet; **b** for two attracting bar magnets and **c** for two repelling bar magnets.

We represent magnetic field patterns by drawing magnetic field lines.

- The magnetic field lines come out of north poles and go into south poles.
- The direction of a field line at any point in the field shows the direction of the force that a ‘free’ magnetic north pole would experience at that point.
- The field is strongest where the field lines are closest together.

An electromagnet makes use of the magnetic field created by an electric current (Figure 24.3a). A coil is used because this concentrates the magnetic field. One end becomes a north pole (field lines emerging), while the other end is the south pole. Another name for a coil like this is a **solenoid**.

The field pattern for the solenoid looks very similar to that of a bar magnet (see Figure 24.2a), with field lines emerging from a north pole at one end and returning to a south pole at the other. The strength of the magnetic field of a solenoid can be greatly increased by adding a core made of a ferrous (iron-rich) material. For example, an iron rod placed inside the solenoid can act as a core; when the current flows through the solenoid, the iron core itself becomes magnetised and this produces a much stronger field. A flat coil (Figure 24.3b) has a similar field to that of a solenoid.

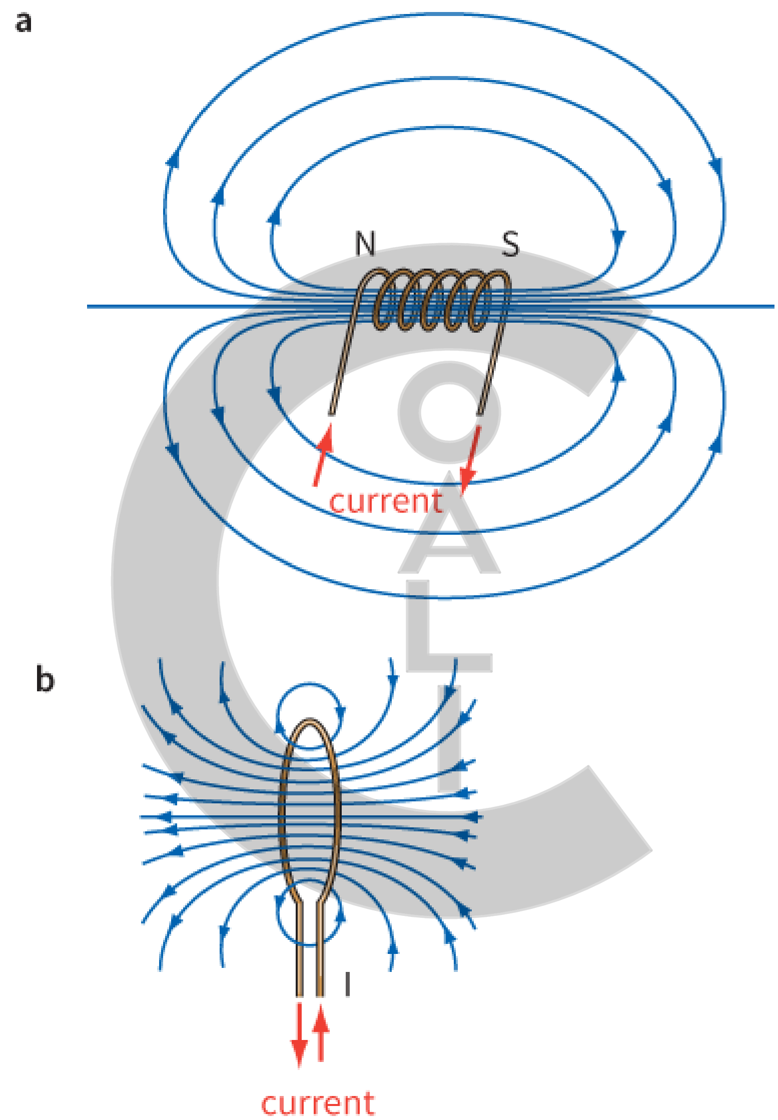


Figure 24.3: Magnetic field patterns for **a** a solenoid, and **b** a flat circular coil.

If we unravel an electromagnet, we get a weaker field. This, too, can be investigated using iron filings or compasses. The magnetic field pattern for a long current-carrying wire is very different from that of a solenoid. The magnetic field lines shown in Figure 24.4 are circular, centred on the long current-carrying wire. Further away from the wire, the field lines are drawn further apart, representing the weaker field at this distance. Reversing the current reverses the direction of the field.

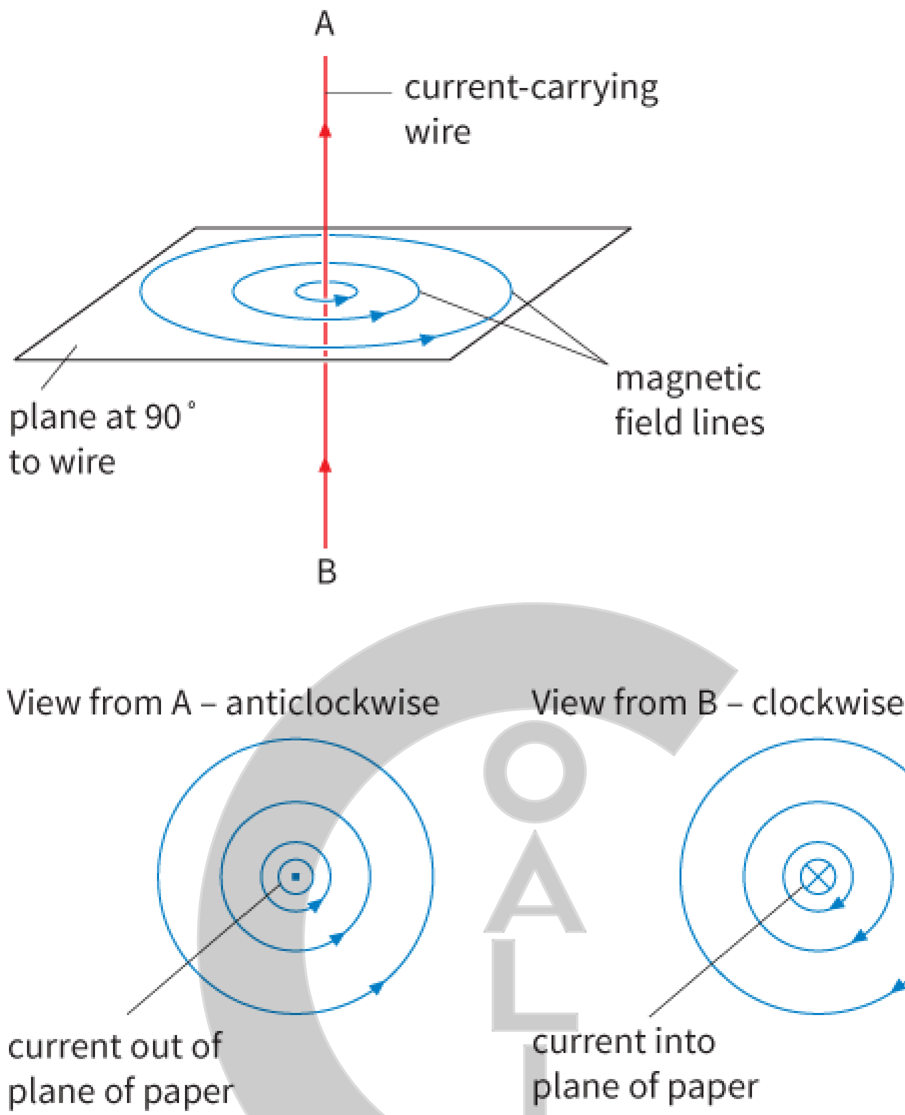


Figure 24.4: The magnetic field pattern around a current-carrying wire. The diagram also shows the convention used to indicate the direction of current.

All magnetic fields are created by **moving** charges. (In the case of a wire, the moving charges are free electrons.) This is even true for a permanent bar magnet. In a permanent magnet, the magnetic field is produced by the movement of electrons within the atoms of the magnet. Each electron represents a tiny current as it circulates around within its atom, and this current sets up a magnetic field. In a ferrous material, such as iron, the weak fields due to all the electrons combine together to make a strong field, which spreads out into the space beyond the magnet. In non-magnetic materials, the fields produced by the electrons cancel each other out.

Field direction

The idea that magnetic field lines emerge from north poles and go into south poles is simply a convention. Figure 24.5 shows some useful rules for remembering the direction of the magnetic field produced by a current.

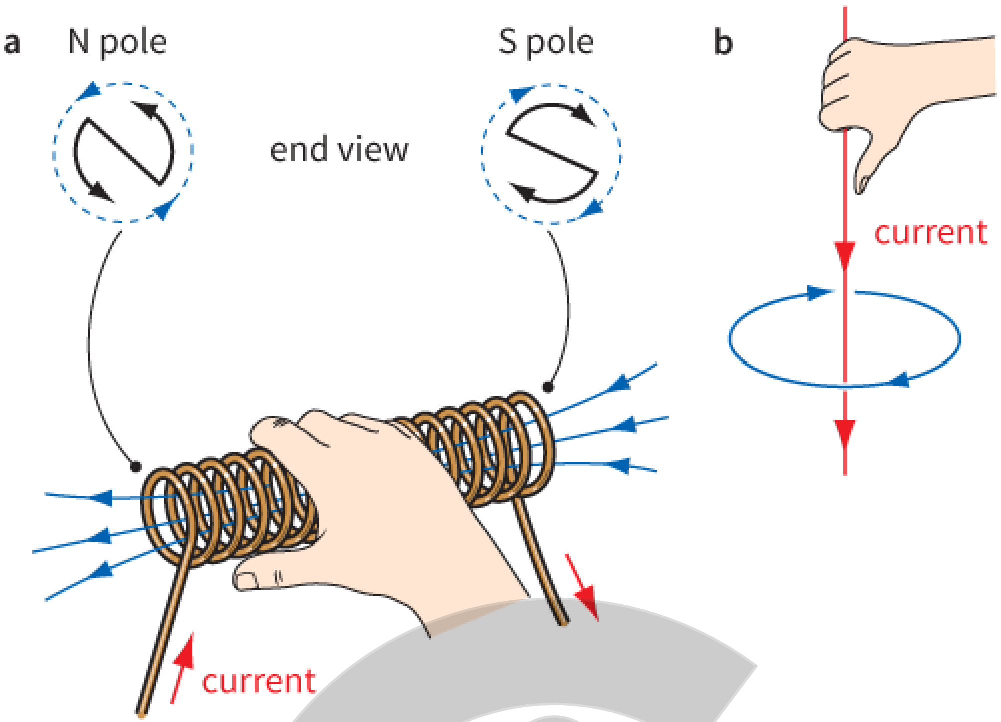


Figure 24.5: Two rules for determining the direction of a magnetic field, **a** inside a solenoid and **b** around a current-carrying wire.

The **right-hand grip rule** gives the direction of magnetic field lines in an electromagnet. Grip the coil so that your fingers go around it following the direction of the current. Your thumb now points in the direction of the field lines inside the coil; that is, it points towards the electromagnet's north pole.

Another way to identify the poles of an electromagnet is to look at it end on, and decide which way round the current is flowing. Figure 24.5a show how you can remember that clockwise is a south pole, anticlockwise is a north pole.

The circular field around a wire carrying a current does not have magnetic poles. To find the direction of the magnetic field you need to use another rule, the **right-hand rule**. Grip the wire with your right hand, pointing your thumb in the direction of the current. Your fingers curl around in the direction of the magnetic field.

Note that these two rules are slightly different. The right-hand grip rule applies to a *solenoid*; the fingers are curled in the direction of the current and the thumb then gives the direction of the field. The right-hand rule applies to a *current in a straight wire*; the thumb is pointed in the direction of the current and the fingers then give the direction of the field lines.

Questions

1 Sketch the magnetic field pattern around a long straight wire carrying an electric current. Now, alongside this first sketch, draw a second sketch to show the field pattern if the current flowing is doubled and its direction reversed. How does the pattern show that the field is stronger nearer the wire?

2 Sketch the diagram in Figure 24.6, and label the north and south poles of the electromagnet. Show on your sketch the direction of the magnetic field (as shown by the needle of a plotting compass) at each of the positions A, B, C and D.

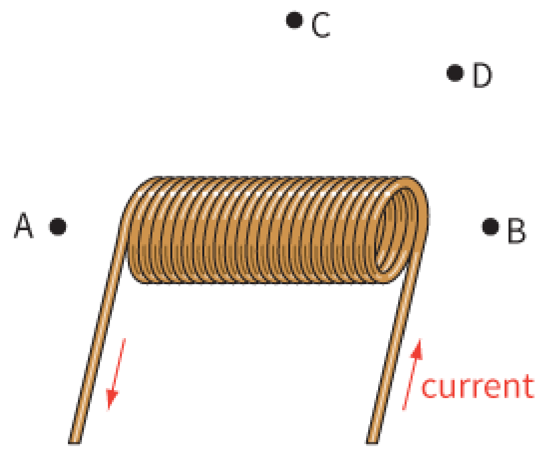


Figure 24.6: A current-carrying solenoid. For Question 2.

3 State which of the pairs of electromagnets shown in Figure 24.7 attract one another, and which repel.

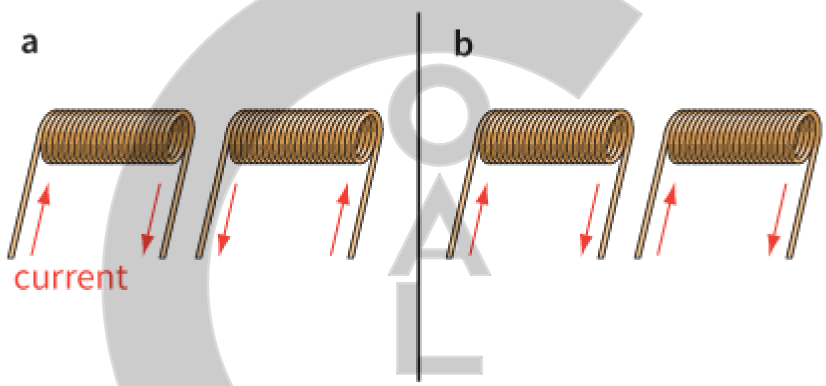


Figure 24.7: Two pairs of solenoids. For Question 3.

24.2 Magnetic force

A current-carrying wire is surrounded by a magnetic field. This magnetic field will interact with an external magnetic field, giving rise to a force on the conductor, just like the fields of two interacting magnets. A simple situation is shown in Figure 24.8.

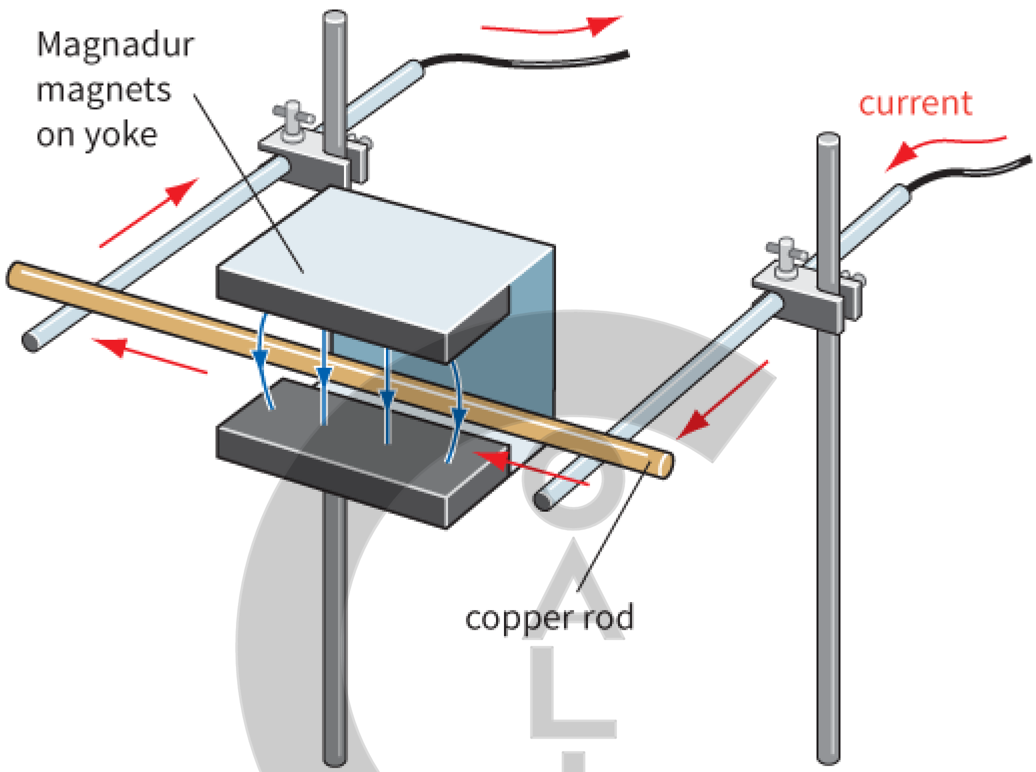


Figure 24.8: The copper rod is free to roll along the two horizontal aluminium 'rails'.

The magnets create a fairly uniform magnetic field. As soon as the current in the copper rod is switched on, the rod starts to roll, showing that a force is acting on it. We use **Fleming's left-hand (motor) rule** to predict the direction of the force on the current-carrying conductor, as explained in Practical Activity 24.1.

PRACTICAL ACTIVITY 24.1

Using Fleming's left-hand rule

Look at Figure 24.9. There are three things here, all of which are mutually at right angles to each other – the magnetic field, the current in the rod and the force on the rod. These can be represented by holding the thumb and the first two fingers of your left hand so that they are mutually at right angles (Figure 24.9). Your thumb and fingers then represent:

- thu**M**b – direction of **M**otion
- **F**irst finger – direction of external magnetic **F**ield

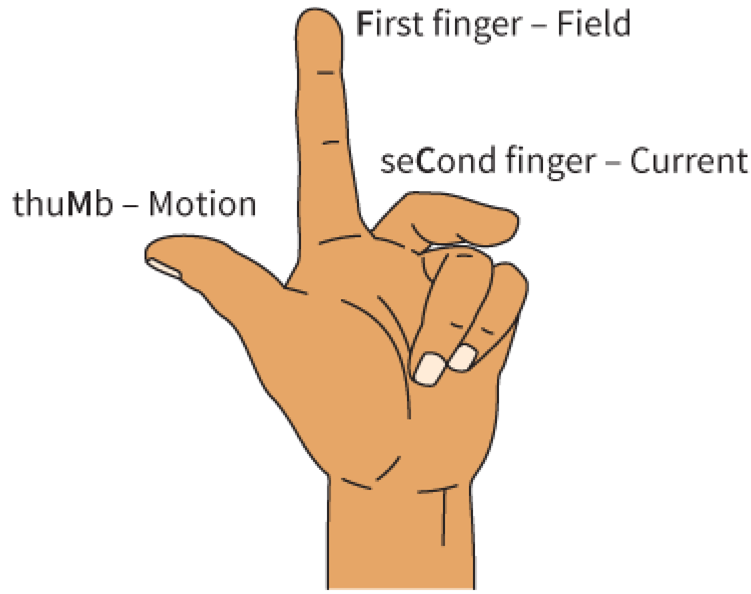


Figure 24.9: Fleming's left-hand (motor) rule.

- seCond finger – direction of conventional **C**urrent.

If the thumb and first two fingers of the left hand are held at right angles to one another, with the **F**irst finger pointing in the direction of the **F**ield and se**C**ond finger in the direction of the **C**urrent, then the thu**M**b points in the direction of the **M**otion or force.

You should practise using your left hand to check that the rule correctly predicts these directions.

Explaining the magnetic force

We can explain this force by thinking about the magnetic fields of the magnets and the current-carrying conductor. These fields combine or interact to produce the force on the rod.

Figure 24.10 shows:

- the external magnetic field of the magnets
- the magnetic field of the current-carrying conductor
- the combined fields of the current-carrying conductor and the magnets.

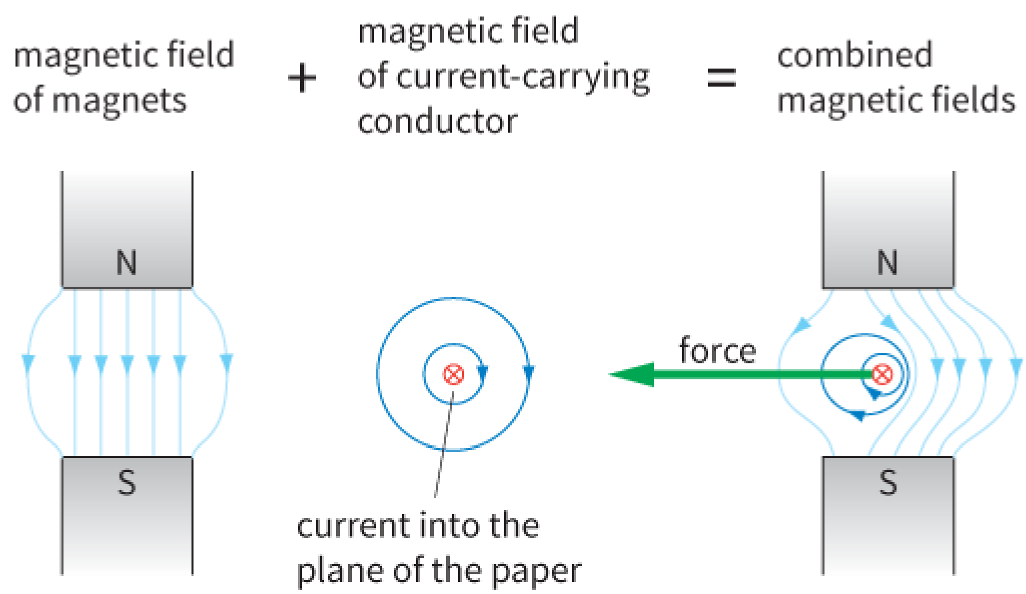


Figure 24.10: In the field of a permanent magnet, a current-carrying conductor experiences a force in accordance with Fleming's left-hand rule. The fields due to the permanent magnet and the current (left and centre) combine as shown on the right.

If you think of the magnetic field lines as elastic bands then you can see why the wire is pushed out in the direction shown.

The production of this force is known as the **motor effect**, because this force is used in electric motors. In a simple motor, a current in a coil produces a magnetic field; this field interacts with a second field produced by a permanent magnet.

Question

4 Figure 24.11 shows three examples of current-carrying conductors in magnetic fields. For each example, decide whether there will be a magnetic force on the conductor. If there is a force, in what direction will it act? Note the cross in the circle shows the current is into the plane of the paper, as in Figure 24.4.

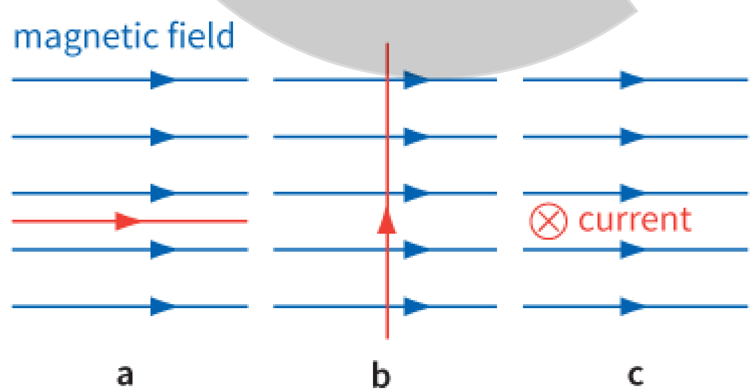


Figure 24.11: Three conductors in a magnetic field.

24.3 Magnetic flux density

In electric or gravitational field diagrams, the strength of the field is indicated by the separation between the field lines. The field is strongest where the field lines are closest together. The same is true for magnetic fields. The **strength** of a magnetic field is known as its **magnetic flux density**, with symbol *B*. Sometimes it is known as the magnetic field strength. (You can imagine this quantity to represent the number of magnetic field lines passing through a region per unit area.) The magnetic flux density is greater close to the pole of a bar magnet, and gets smaller as you move away from it.

We define gravitational field strength *g* at a point as the force per unit mass:

$$g = \frac{F}{m}$$

Electric field strength *E* is defined as the force per unit positive charge:

$$E = \frac{F}{Q}$$

In a similar way, magnetic flux density is defined in terms of the magnetic force experienced by a current-carrying conductor placed at **right angles** to a magnetic field. For a uniform magnetic field, the flux density *B* is defined by the equation:

$$B = \frac{F}{IL}$$

where *F* is the force experienced by a current-carrying conductor, *I* is the current in the conductor and *L* is the length of the conductor in the uniform magnetic field of flux density *B*. The direction of the force *F* is given by Fleming's left-hand rule.

The magnetic flux density at a point in space is the force experienced per unit length by a long straight conductor carrying unit current and placed at right angles to the field at that point.

The unit for magnetic flux density is the tesla, T. It follows from the equation for *B* that $1\ \text{T} = 1\ \text{N A}^{-1}\ \text{m}^{-1}$.

The force on the conductor is given by the equation:

$$F = BIL$$

Note that you can only use this equation when the field is at right angles to the current.

KEY EQUATION

$$F = BIL$$

Force on the conductor (when the conductor is at right angles to the magnetic field).

Questions

5 A current of 0.20 A flows in a wire of length 2.50 m placed at right angles to a magnetic field of flux density 0.060 T. Calculate the force on the wire.

6 A 20 cm length of wire is placed at right angles to a magnetic field. When a current of 1.5 A flows in the wire, a force of 0.015 N acts on it. Determine the flux density of the field.

7 A wire of length 50 cm carrying a current lies at right angles to a magnetic field of flux density 5.0 mT.

- **a** If 10^{18} electrons pass a point in the wire each second, what current is flowing? (electron charge $e = 1.60 \times 10^{-19}$ C.)

b What force acts on the wire?

24.4 Measuring magnetic flux density

Practical Activity 24.2 looks at two practical methods for measuring magnetic flux density.

PRACTICAL ACTIVITY 24.2 MEASURING MAGNETIC FLUX DENSITY

Measuring *B* with a Hall probe

The simplest device for measuring magnetic flux density *B* is a Hall probe (Figure 24.12). When the probe is held so that the field lines are passing at right angles through the flat face of the probe, the meter gives a reading of the value of *B*. Some instruments are calibrated so that they give readings in microteslas (μT) or milliteslas (mT). Others are not calibrated, so you must either calibrate them or use them to obtain relative measurements of *B*.

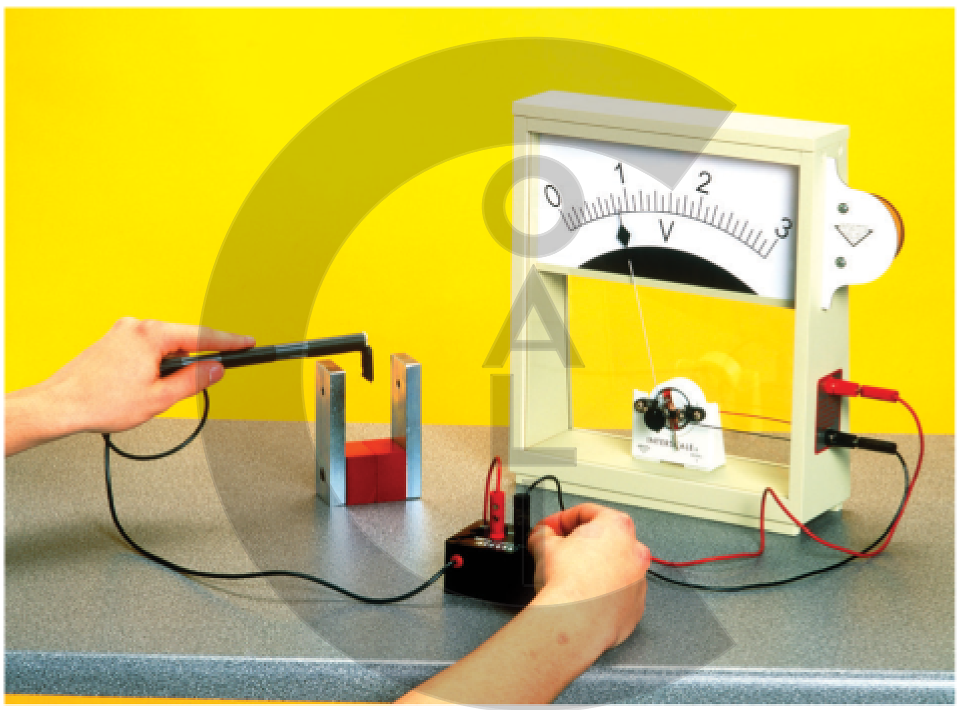


Figure 24.12: Using a Hall probe to measure the flux density between two magnets.

A Hall probe must be held so that the field lines are passing directly through it, at right angles to the flat surface of the probe (Figure 24.13). If the probe is not held in the correct orientation, the reading on the meter will be reduced.

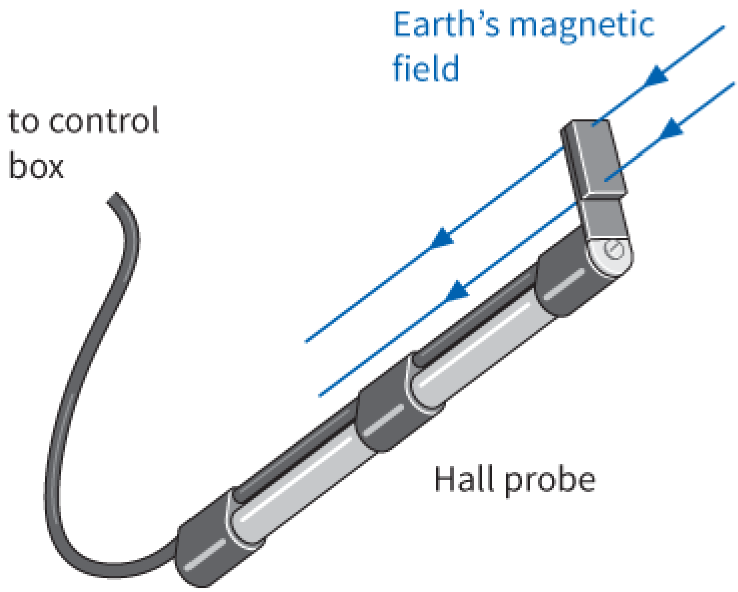


Figure 24.13: Magnetic flux lines must pass through the probe at 90° to the surface.

A Hall probe is sensitive enough to measure the Earth's magnetic flux density. The probe is first held so that the Earth's field lines are passing directly through it, as shown in Figure 24.13. In this orientation, the reading on the voltmeter will be a maximum. The probe is then rotated through 180° so that the magnetic field lines are passing through it in the opposite direction. The change in the reading of the meter is twice the Earth's magnetic flux density. There is more about how the Hall probe works in Chapter 25.

Measuring *B* with a current balance

Figure 24.14 shows a simple arrangement that can be used to determine the flux density between two magnets. The magnetic field between these magnets is (roughly) uniform. The length L of the current-carrying wire in the uniform magnetic field can be measured using a ruler.

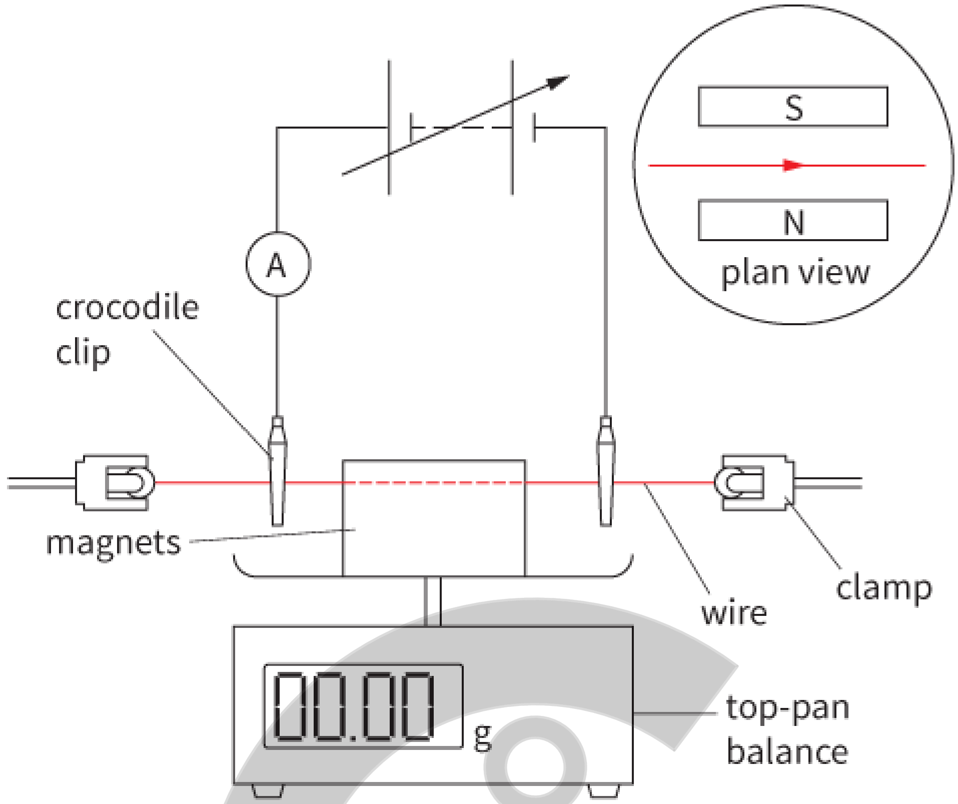


Figure 24.14: An arrangement to determine magnetic flux density in the laboratory.

When there is no current in the wire, the magnet arrangement is placed on the top pan and the balance is zeroed. Now, when a current *I* flows in the wire, its value is shown by the ammeter. The wire experiences an upward force and, according to Newton's third law of motion, there is an equal and opposite force on the magnets. The magnets are pushed downwards and a reading appears on the scale of the balance. The force *F* is given by *mg*, where *m* is the mass indicated on the balance in kilograms and *g* is the acceleration of free fall (9.81 m s^{-2}).

Knowing *F*, *I* and *L*, the magnetic flux density *B* between the magnets can be determined using the equation:

$$B = \frac{F}{IL}$$

You can also use the arrangement in Figure 24.14 to show that the force is directly proportional to the current.

A system like this in effect 'weighs' the force on the current-carrying conductor, and is an example of a current balance. Another version of a current balance is shown in Figure 24.15. This consists of a wire frame that is balanced on two pivots. When a current flows through the frame, the magnetic field pushes the frame downwards. By adding small weights to the other side of the frame, you can restore it to a balanced position.

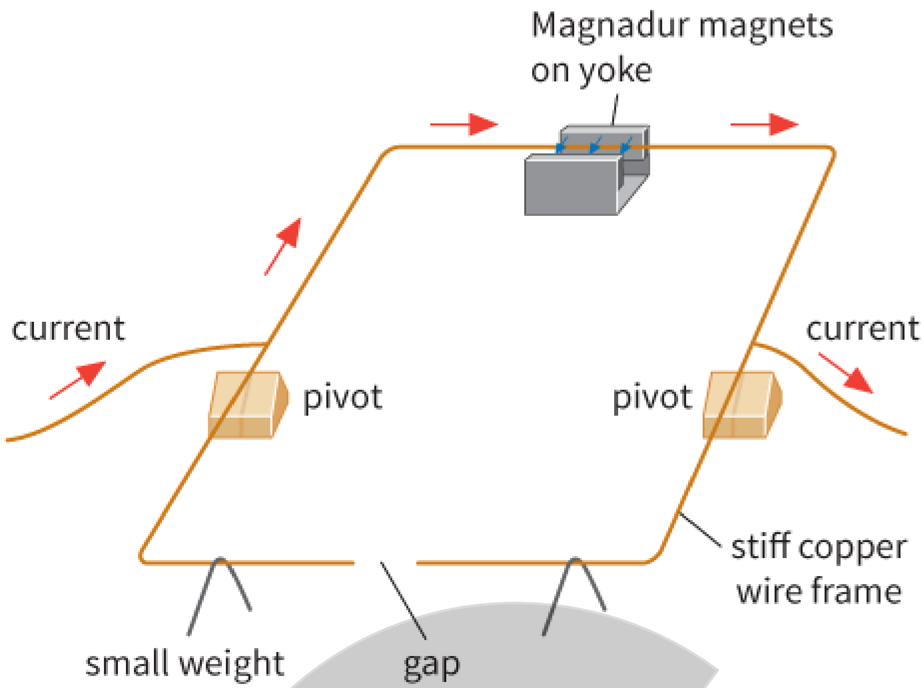


Figure 24.15: A simple laboratory current balance.

Questions

8 In the examples shown in the diagrams in Figure 24.16, which current balances will tilt? Will the side carrying the current tilt upwards or downwards?

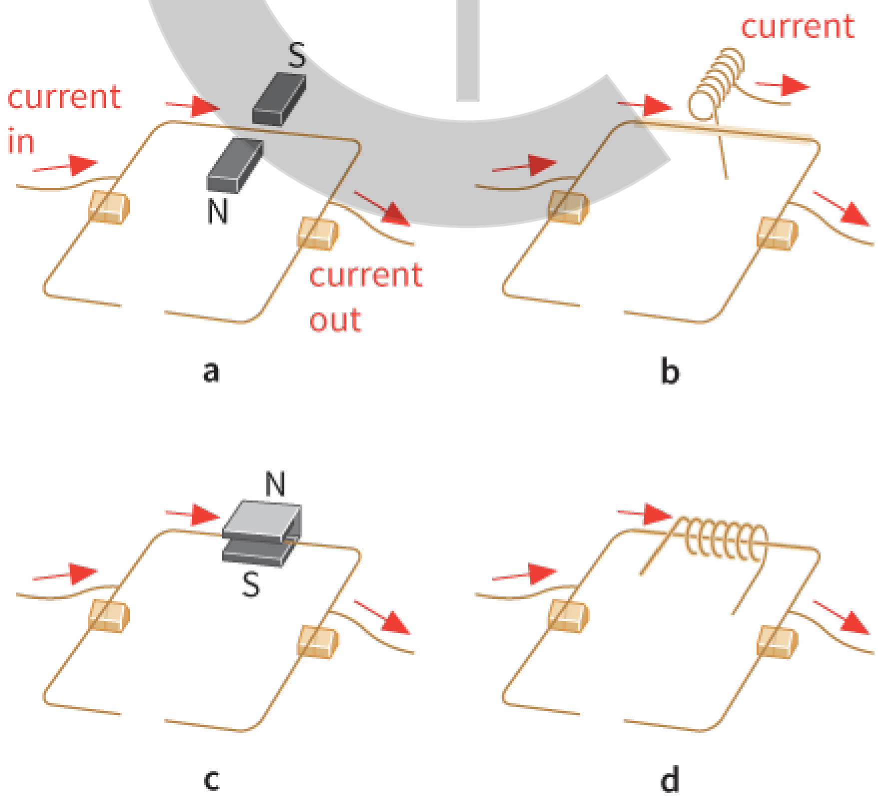


Figure 24.16: Four current balances – will they tip? For Question 8.

9 In the arrangement shown in Figure 24.17, the balance reading changes from 102.48 g to 104.48 g when the current is switched on. Explain why this happens and give the direction and the size of the force on the wire when the current is on. What is the direction of the current in the wire?

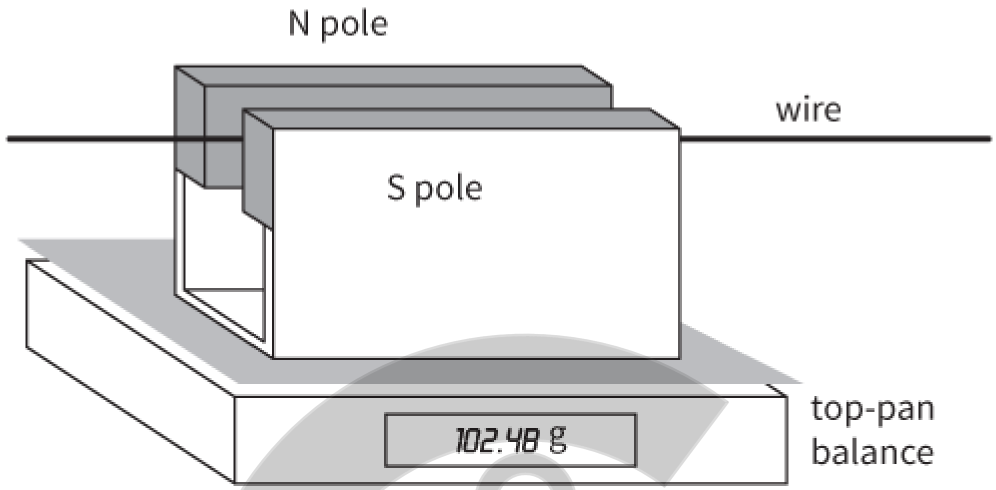


Figure 24.17: Using an electronic balance. For Question 9.

24.5 Currents crossing fields

At right angles

We explained the force on a current-carrying conductor in a field in terms of the interaction of the two magnetic fields: the field due to the current and the external field. Here is another, more abstract, way of thinking about this.

Whenever an electric current cuts across magnetic field lines (Figure 24.18), a force is exerted on the current-carrying conductor. This helps us to remember that a conductor experiences no force when the current is parallel to the field.

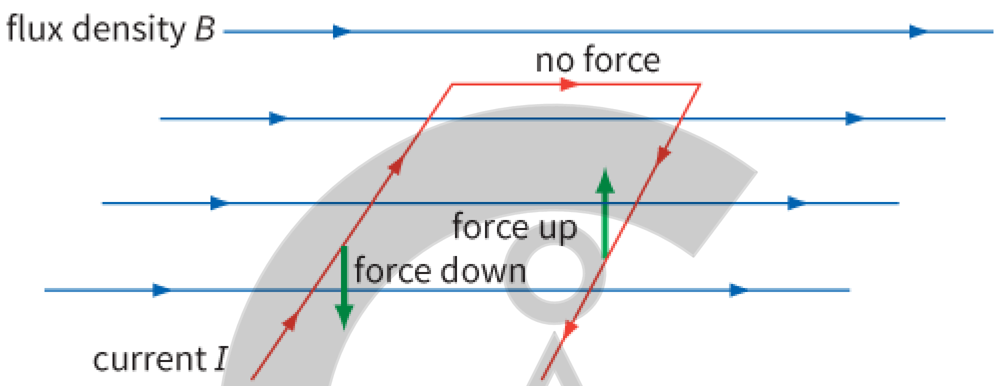


Figure 24.18: The force on a current-carrying conductor crossing a magnetic field.

This is a useful idea, because it saves us thinking about the field due to the current. In Figure 24.18, we can see that there is only a force when the current cuts across the magnetic field lines.

This force is very important – it is the basis of electric motors. Worked example 1 shows why a current-carrying coil placed in a magnetic field rotates.

WORKED EXAMPLE

1 An electric motor has a rectangular loop of wire with the dimensions shown in Figure 24.19. The loop is in a magnetic field of flux density 0.10 T. The current in the loop is 2.0 A. Calculate the torque (moment) that acts on the loop in the position shown.

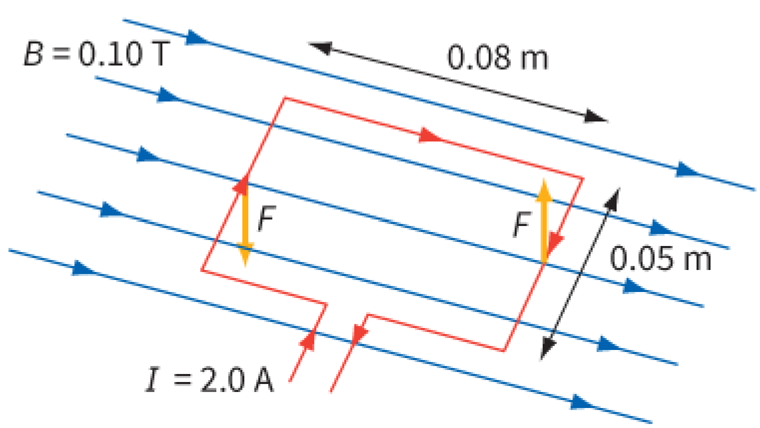


Figure 24.19: A simple electric motor – a current-carrying loop in a magnetic field.

Step 1 The quantities we know are:

$B = 0.10$ T, $I = 2.0$ A and $L = 0.05$ m

Step 2 Now we can calculate the force on one side of the loop using the equation $F = BIL$:

$$\begin{aligned} F &= 0.10 \times 2.0 \times 0.05 \\ &= 0.01 \text{ N} \end{aligned}$$

Step 3 The two forces on opposite sides of the loop are equal and anti-parallel. In other words, they form a couple. From Chapter 4, you should recall that the torque (moment) of a couple is equal to the magnitude of one of the forces times the perpendicular distance between them. The two forces are separated by 0.08 m, so:

$$\begin{aligned} \text{torque} &= \text{force} \times \text{seperation} \\ &= 0.01 \times 0.08 \\ &= 8.0 \times 10^{-4} \text{ N m} \end{aligned}$$

Questions

10 A wire of length 50 cm carrying a current of 2.4 A lies at right angles to a magnetic field of flux density 5.0 mT. Calculate the force on the wire.

11 The coil of an electric motor is made up of 200 turns of wire carrying a current of 1.0 A. The coil is square, with sides of length 20 cm, and it is placed in a magnetic field of flux density 0.05 T.

- **a** Determine the maximum force exerted on the side of the coil.
- **b** In what position must the coil be for this force to have its greatest turning effect?
- **c** List four ways in which the motor could be made more 'powerful' – that is, have greater torque.

At an angle other than 90°

Now we must consider the situation where the current-carrying conductor cuts across a magnetic field at an angle other than a right angle. In Figure 24.20, the force gets weaker as the conductor is moved round from OA to OB, to OC and finally to OD. In the position OD, there is no force on the conductor. To calculate the force, we need to find the component of the magnetic flux density B at right angles to the current. This component is $B \sin \theta$, where θ is the angle between the magnetic field and the current or the conductor. Substituting this into the equation $F = BIL$ gives:

$$F = (B \sin\theta) IL$$

KEY EQUATION

$$F = BIL \sin\theta$$

Force on a current-carrying conductor.

or simply:

$$F = BIL \sin\theta$$

Now look at Worked example 2.

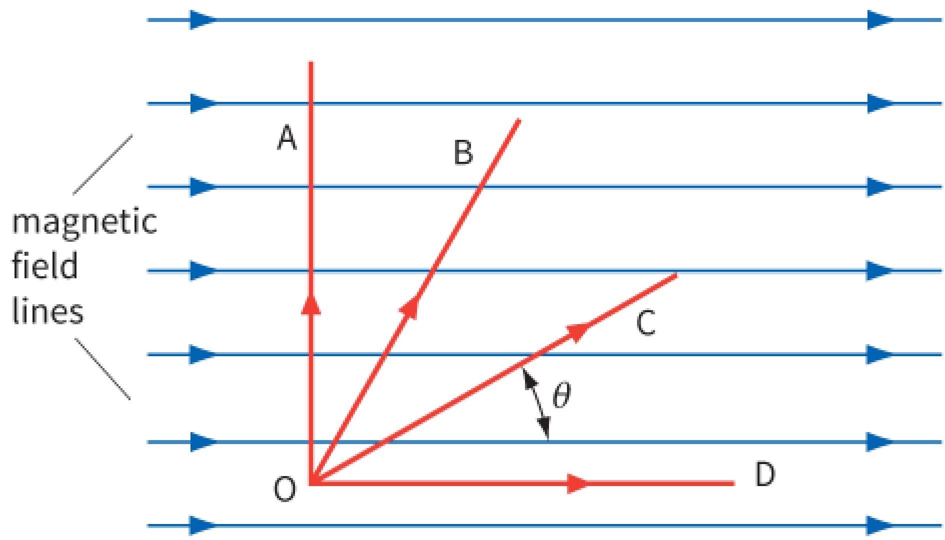


Figure 24.20: The force on a current-carrying conductor depends on the angle it makes with the magnetic field lines.

WORKED EXAMPLE

2 A conductor OC (see Figure 24.20) of length 0.20 m lies at an angle θ of 25° to a magnetic field of flux density 0.050 T. Calculate the force on the conductor when it carries a current of 400 mA.

Step 1 Write down what you know, and what you want to know:

$$B = 0.050\ \text{T}$$
$$L = 0.20\ \text{m}$$
$$I = 400\ \text{mA} (= 0.40\ \text{A})$$
$$\theta = 25°$$
$$F = ?$$

Step 2 Write down the equation, substitute values and solve:

$$F = BIL \sin\theta$$
$$= 0.050 \times 0.40 \times 0.20 \times \sin 25°$$
$$\approx 1.7 \times 10^{-3}\ \text{N}$$

Step 3 Give the direction of the force. The force acts at 90° to the field and the current, i.e. perpendicular to the page. The left-hand rule shows that it acts downwards into the plane of the paper.

Note that the component of *B* parallel to the current is *B* cos θ, but this does not contribute to the force; there is no force when the field and current are parallel. The force *F* is at right angles to both the current and the field.

Question

12 What force is exerted on each of the currents shown in Figure 24.21, and in what direction does each force act?

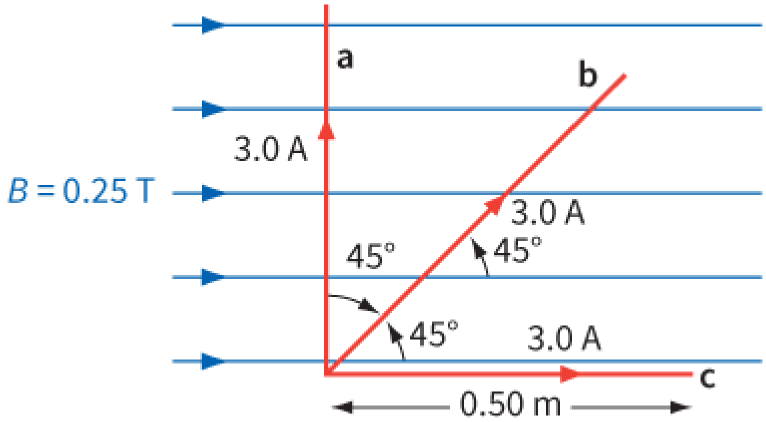


Figure 24.21: Three currents in a magnetic field. For Question 12.

24.6 Forces between currents

Any electric current has a magnetic field around it. If we have two currents, each will have its own magnetic field and we might expect these to interact.

Explaining the forces

There are two ways to understand the origin of the forces between current-carrying conductors. In the first, we draw the magnetic fields around two current-carrying conductors (Figure 24.22a). Figure 24.22a shows two unlike (anti-parallel) currents, one flowing into the page, the other flowing out of the page. Their magnetic fields circle round, and in the space between the wires there is an extra-strong field. We imagine the field lines squashed together, and the result is that they push the wires apart. The diagram shows the resultant field, and the repulsive forces on the two wires.

Figure 24.22b shows the same idea, but for two like (parallel) currents. In the space between the two wires, the magnetic fields cancel out. The wires are pushed together.

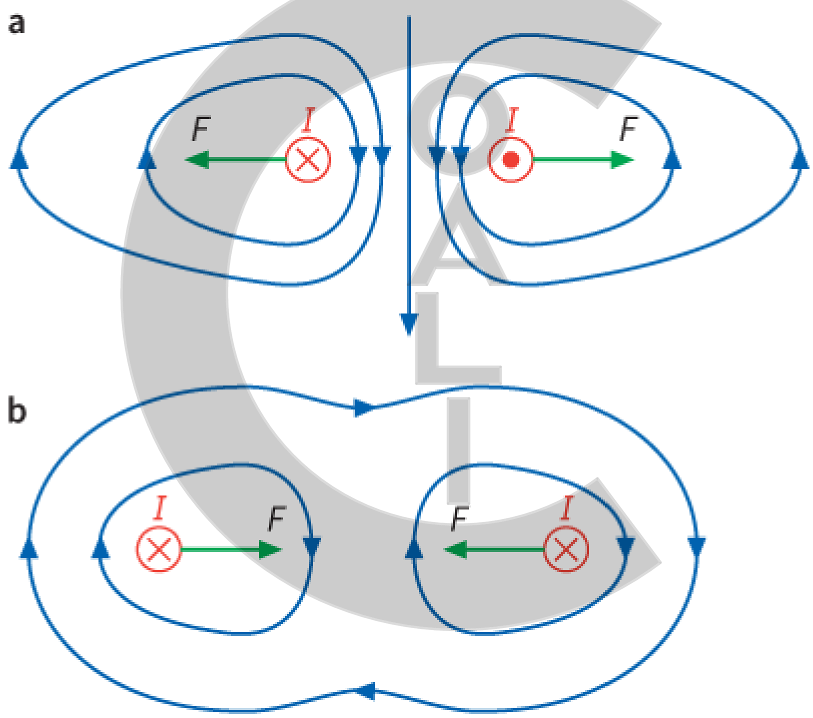


Figure 24.22: The forces between current-carrying wires.

The other way to explain the forces between the currents is to use the idea of the motor effect. Figure 24.23 again shows two like currents, I_1 and I_2, but this time we only consider the magnetic field B of one of them, I_1. The second current I_2 is flowing across the magnetic field of I_1; from the diagram, you can see that B is at right angles to I_2. Hence, there will be a force on I_2 (the BIL force), and we can find its direction using Fleming's left-hand rule. The arrow shows the direction of the force, which is towards I_1. Similarly, there will be a BIL force on I_1, directed towards I_2.

These two forces are equal and opposite to one another. They are an example of an action and reaction pair, as described by Newton's third law of motion.

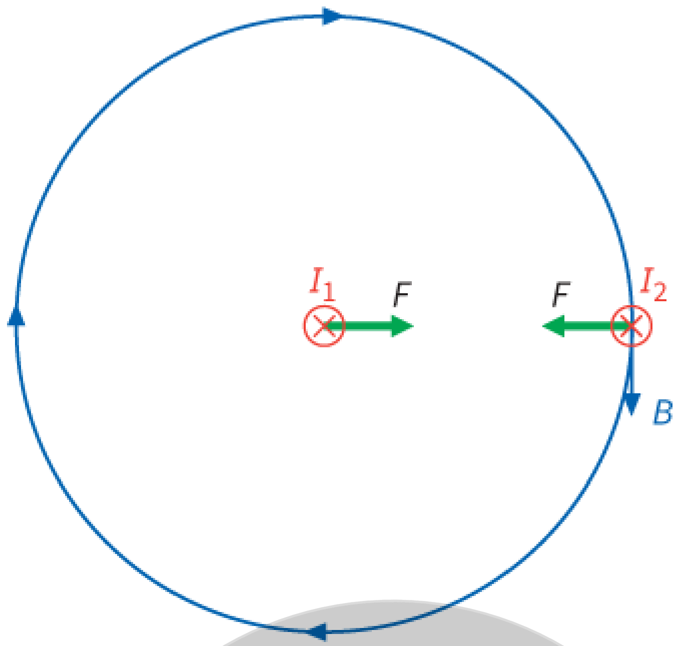


Figure 24.23: Explaining the force between two currents.

PRACTICAL ACTIVITY 24.3

Observing the forces between currents

You can observe the attraction and repulsion between two parallel currents using the equipment shown in Figure 24.24.

Two long thin strips of aluminium foil are mounted so that they are parallel and a small distance apart. By connecting them in series with a power supply, you can make a current occur in both of them. By changing the connections, you can make the current first in the same direction through both strips (parallel currents) and then in opposite directions (anti-parallel currents).

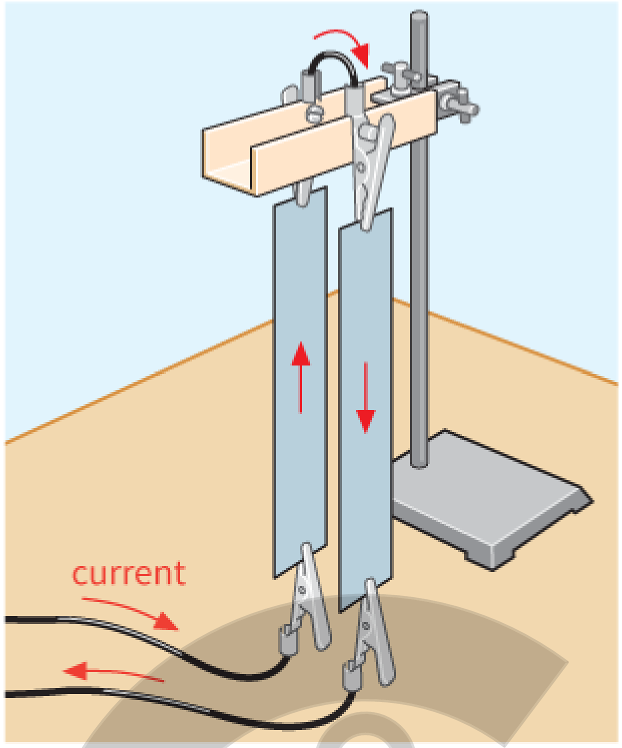


Figure 24.24: Current in two aluminium strips – their magnetic fields interact.

If you try this out, you will observe the strips of foil either bending towards each other or away from each other. (Foil is used because it is much more flexible than wire.)

You should find that parallel currents attract one another, while anti-parallel currents repel. This may seem surprising, since we are used to opposite charges attracting, and opposite magnetic poles attracting. Now we have found that opposite currents repel one another.

Question

13 Two flat circular coils of wire are set up side by side, as shown in Figure 24.25. They are connected in series so that the same current flows around each, and in the same direction. Will the coils attract or repel one another? Explain your answer, first by describing the coils as electromagnets, and second by considering the forces between parallel currents. What will happen if the current is reversed in both coils?

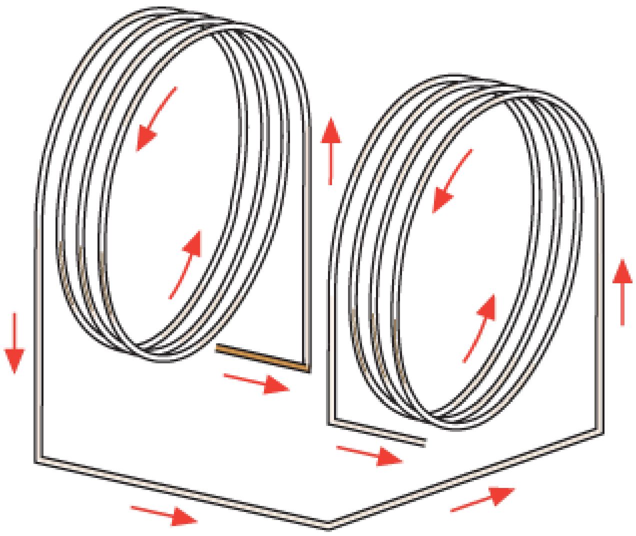


Figure 24.25: Two coils carrying the same current. For Question 13.

24.7 Relating SI units

In this chapter, we have seen how one SI unit, the tesla, is defined in terms of three others, the amp, the metre and the newton. It is an essential feature of the SI system that all units are carefully defined; in particular, derived units such as the newton and tesla must be defined in terms of a set of more fundamental units called **base units**.

We met the idea of base units in Chapter 3. The SI system of units has seven base units, of which you have met six. These are:

m kg s A K mol

(The seventh is the candela, cd, the unit of luminous intensity.) Each base unit is carefully defined; for example, the ampere can be defined in terms of the magnetic force between two parallel wires carrying a current. The exact definition is not required, but you should know that the ampere is itself a base unit. Other units are known as **derived units**, and can be deduced from the base units. For example, as shown in Chapter 3, the newton is given by:

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

Similarly, in this chapter, you have learned about the tesla, the unit of magnetic flux density, given by:

$$1\ \text{T} = 1\ \text{N A}^{-1}\ \text{m}^{-1} \quad \text{or} \quad 1\ \text{T} = 1\ \text{kg A}^{-1}\ \text{s}^{-2}$$

If you learn formulae relating physical quantities, you can replace the quantities by their units to see how the units are defined. For example:

$$\text{force} = \text{mass} \times \text{acceleration} \qquad F = ma \qquad \text{N} = \text{kg m s}^{-2}$$

You should be able to picture how the different derived units form a logical sequence, as shown in Table 24.1.

Base units	Derived units	Because
m, kg, s	newton N = kg m s^{-2}	$F = ma$
	joule J = kg m^2 s^{-2}	$W = Fd$
	watt W = kg m^2 s^{-3}	$P = \frac{E}{t}$
m, kg, s, A	coulomb C = A s	$Q = It$
	volt V = kg m^2 A^{-1} s^{-3}	$V = \frac{W}{Q}$
	tesla T = kg A^{-1} s^{-2}	$B = \frac{F}{IL}$

Table 24.1 How derived units relate to base units in the SI system.

24.8 Comparing forces in magnetic, electric and gravitational fields

We have now considered three types of field: electric (Chapter 21), gravitational (Chapter 17) and magnetic (this chapter). What are the similarities and differences between these three types of field?

Modern physics sees magnetic fields and electric fields as two parts of a combined whole, an electromagnetic field. Gravitational fields, however, are different in nature to electromagnetic fields.

Gravitational and electric fields are defined in terms of placing a test mass or a test charge at a point to measure the field strength. Similarly, a test wire carrying a current can be placed at a point to measure the magnetic field strength. Therefore, all fields are defined in terms of the force on a unit mass, charge or current.

Other features that all fields share include:

- action at a distance, between masses, between charges or between wires carrying currents
- decreasing strength with distance from the source of the field
- representation by field lines, the direction of which show the direction of the force at points along the line; the density of field lines indicates the relative strength of the field.

How do the forces arising from these fields compare? The answer depends on the exact situation. Using ideas that you have studied earlier, you should be able to confirm each of the following values:

- The force between two 1 kg masses 1 m apart = 6.7×10^{-11} N
- The force between two charges of 1 C placed 1 m apart = 9.0×10^{9} N
- The force per metre on two wires carrying a current of 1 A placed 1 m apart = 2.0×10^{-7} N

This might suggest that the electric force is strongest and gravity is the weakest. Certainly, if you consider an electron in a hydrogen atom moving in a circular orbit around a proton, the electrical force is 10^{39} times the gravitational force. So for an electron, or any other small charged object, electric forces are the most significant. However, over larger distances and with objects of large mass, the gravitational field becomes the most significant. For example, the motions of planets in the Solar System are affected by the gravitational field but the electromagnetic field is comparatively insignificant.

REFLECTION

Without looking at your textbook, make a list of the definitions for measuring the strength of a magnetic, an electric and a gravitational field. Compare your definitions with other students in your class.

Write a list of situations where magnetic fields are used in modern life.

How will you use what you have learned in the future?

SUMMARY

Moving charges produce a magnetic field; this is electromagnetism.

A current-carrying conductor has concentric magnetic field lines. The magnetic field pattern for a solenoid or flat coil resembles that of a bar magnet.

The separation between magnetic field lines is an indication of the field's strength.

Magnetic flux density B is defined by the equation:

$$B = \frac{F}{IL}$$

where F is the force experienced by a current-carrying conductor, I is the current in the conductor and L is the length of the conductor in the uniform magnetic field.

The unit of magnetic flux density is the tesla (T). 1 T = 1 N A^{-1} m^{-1}.

The magnetic force on a current-carrying conductor is given by:

$$F = BIL\sin\theta$$

The force on a current-carrying conductor can be used to measure the flux density of a magnetic field.

A force acts between current-carrying conductors due to the interaction of their magnetic fields.

EXAM-STYLE QUESTIONS

1 A wire carrying a current is placed at right angles to a uniform magnetic field of magnetic flux density *B*. When the current in the wire is *I*, the magnetic force that acts on the wire is *F*.

What is the force on the wire, placed in the same orientation, when the magnetic field strength is 2*B* and the current is $\frac{I}{4}$? **[1]**

A $\frac{F}{4}$

B $\frac{F}{2}$

C F

D $2F$

2 There is an electric current in a wire of mass per unit length 40 g m^{-1}. The wire is placed in a magnetic field of strength 0.50 T and the current is gradually increased until the wire just lifts off the ground.

What is the value of the current when this happens? **[1]**

A 0.080 A

B 0.20 A

C 0.78 A

D 780 A

3 A current-carrying wire is placed in a uniform magnetic field.

a Describe how the wire should be placed to experience the maximum force due to the magnetic field. **[1]**

b Describe how the wire should be placed to experience no force due to the magnetic field. **[1]**

[Total: 2]

4 A current-carrying conductor placed at right angles to a uniform magnetic field experiences a force of 4.70×10^{-3} N. Determine the force on the wire when, separately:

a the current in the wire is increased by a factor of 3.0 **[2]**

b the magnetic flux density is halved **[2]**

c the length of the wire in the magnetic field is reduced to 40% of its original length. **[2]**

[Total: 6]

5 A copper wire carrying a current of 1.2 A has 3.0 cm of its length placed in a uniform magnetic field, as shown.

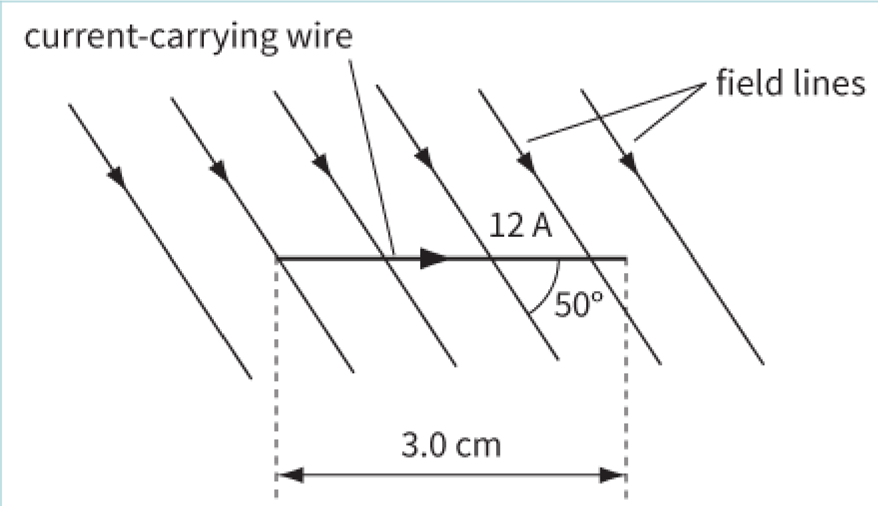


Figure 24.26

The force experienced by the wire is 3.8×10^{-3} N when the angle between the wire and the magnetic field is 50°.

a Calculate the magnetic flux density. **[3]**

b State the direction of the force experienced by the wire. **[1]**

[Total: 4]

6 This diagram shows a view from above of two long, parallel strips of aluminium foil, A and B, carrying a current downwards into the paper.

Figure 24.27

a On a copy of the diagram, draw the magnetic field around and between the two strips. **[2]**

b State and explain the direction of the forces caused by the current in the strips. **[4]**

[Total: 6]

7 This diagram shows a wire XY that carries a constant direct current. Plotting compass R, placed alongside the wire, points due north. Compass P is placed below the wire and compass Q is placed above the wire.

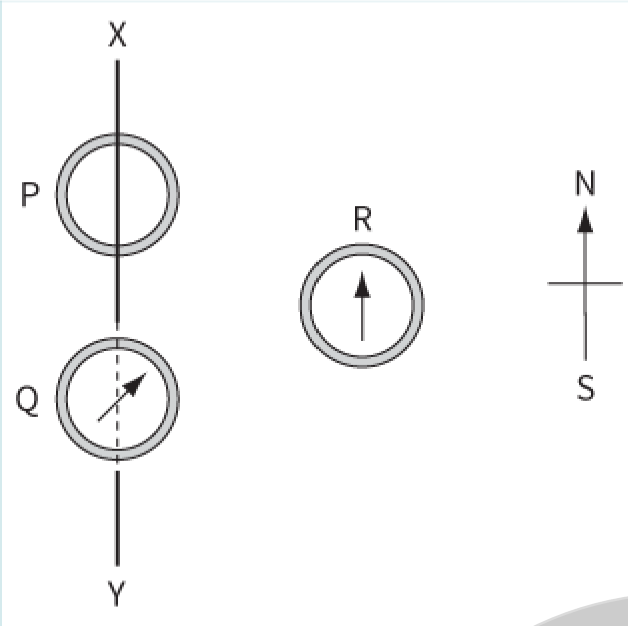


Figure 24.28

a State the direction of the current in the wire. **[1]**

b State in which direction compass P points. **[1]**

c State in which direction compass Q points if the current in the wire is reversed. **[1]**

[Total: 3]

8 This diagram shows a rectangular metal frame PQRS placed in a uniform magnetic field.

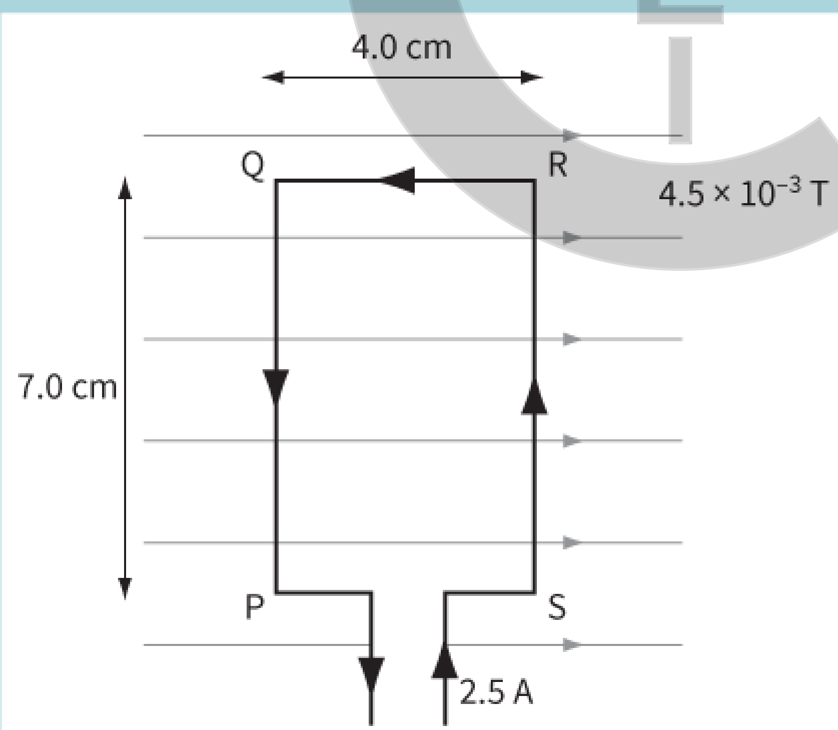


Figure 24.29

The magnetic flux density is 4.5×10^{-3} T and the current in the metal frame is 2.5

A.

a Calculate the force experienced by side PQ of the frame. **[3]**

b Suggest why side QR does not experience a force. **[1]**

c Describe the motion of the frame immediately after the current in the frame is switched on. **[2]**

d Calculate the maximum torque (moment) exerted about an axis parallel to side PQ. **[2]**

[Total: 8]

9 This diagram shows a current-carrying wire frame placed between a pair of Magnadur magnets on a yoke. A pointer is attached to the wire.

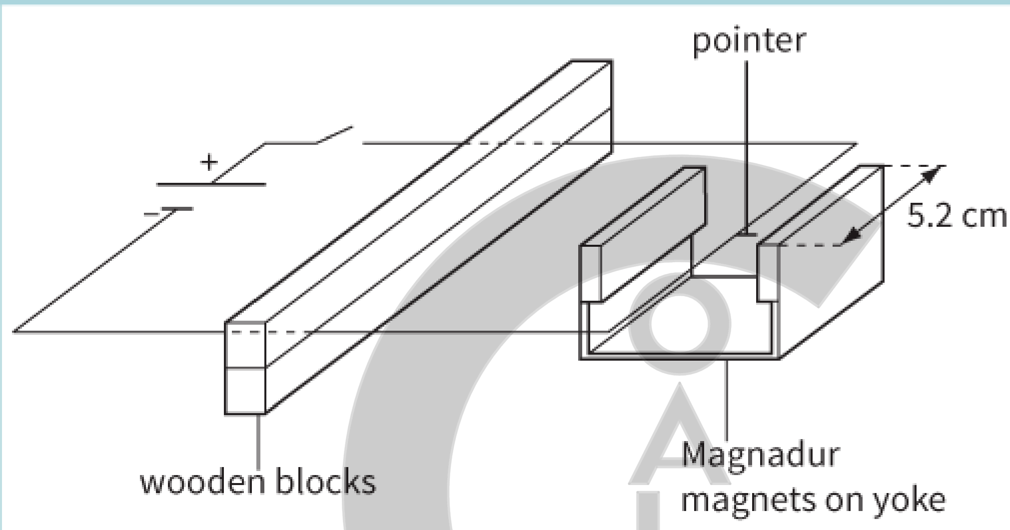


Figure 24.30

A current of 8.5 A in the wire causes the pointer to move vertically upwards. A small paper tape is attached to the pointer and the current is adjusted until the weight of the paper tape causes the pointer to return to its initial position (with no current and no paper tape). The mass of the paper tape is 60 mg. The section of the wire between the poles of the magnetic has a length of 5.2 cm.

a State the direction of the magnetic field. **[1]**

b Calculate the force on the wire due to the magnetic field when it carries a current of 8.5 A. **[2]**

c Calculate the magnetic flux density of the magnetic field between the poles of the magnet. **[3]**

d Describe what happens to the frame if low-frequency alternating current passes through the wire. **[1]**

[Total: 7]

10 a The size of the force acting on a wire carrying a current in a magnetic field is proportional to the size of the current in the wire. With the aid of a diagram, describe how this can be demonstrated in a school laboratory. **[5]**

b At a certain point on the Earth's surface, the horizontal component of the Earth's magnetic field is 1.6×10^{-5} T. A piece of wire 3.0 m long and weight 0.020 N lies in an east–west direction on a laboratory bench. When a large current flows in the wire, the wire just lifts off the surface of the bench.

i State the direction of the current in the wire. **[1]**

ii Calculate the minimum current needed to lift the wire from the bench. **[3]**

[Total: 9]

11 This diagram shows a fixed horizontal wire passing centrally between the poles of a permanent magnet that is placed on a top-pan balance.

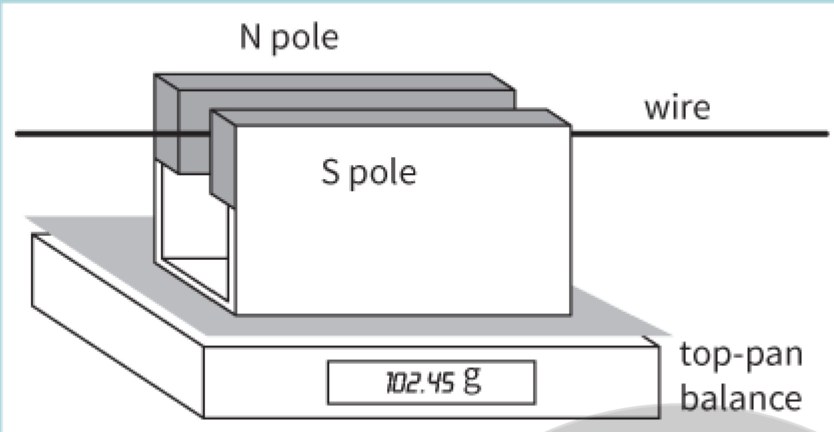


Figure 24.31

With no current flowing, the balance records a mass of 102.45 g. When a current of 4.0 A flows in the wire, the balance records a mass of 101.06 g.

a Explain why the reading on the top-pan balance decreases when the current is switched on. **[2]**

b State and explain the direction of the current flow in the wire. **[2]**

c The length of the wire in the magnetic field is 5.0 cm. Calculate the average magnetic flux density between the poles of the magnet. **[2]**

d Sketch a graph, with balance reading on the vertical axis and current on the horizontal axis, to show how the balance reading changes when the current is altered. **[2]**

[Total: 8]

12 a Define **magnetic flux density** and explain the similarity with the definition of electric field strength. **[3]**

b Two thin horizontal wires are placed in a north–south direction. One wire is placed on a bench and the other wire is held 3.0 cm directly above the first wire.

i When equal currents flow in the two wires, the force exerted on the bench by the lower wire decreases. Explain why this is so. What can you say about the directions of the currents in the two wires? **[4]**

ii The magnetic flux density B at a distance x from a long straight wire carrying a current I is given by the expression $B = 2.0 \times 10^{-7} \frac{I}{x}$ where x is in metres and I is in amps. When the current in each wire is 4.0 A, calculate the force per unit length on one wire due to the current in the other. **[3]**

[Total: 10]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic...	Needs more work	Almost there	Ready to move on
understand that a magnetic field is a field of force produced by moving charges or permanent magnets and represented by field lines	24.1			
sketch magnetic field patterns due to the currents in a long straight wire, a flat circular coil and a long solenoid	24.1			
understand that the magnetic field due to the current in a solenoid is increased by a ferrous core	24.1			
understand forces on a current-carrying conductor in a magnetic field	24.3			
recall and use the equation $F = BIL\sin\theta$ and use Fleming's left-hand rule to find directions	24.2, 24.5			
define magnetic flux density	24.4			
explain the origin of the forces between current-carrying conductors and determine the direction of the forces.	24.6			

〉 Chapter 25

Motion of charged particles

LEARNING INTENTIONS

In this chapter you will learn how to:

- determine the direction of the force on a charge moving in a magnetic field
- recall and use $F = BQv \sin\theta$
- describe the motion of a charged particle moving in a uniform magnetic field perpendicular to the direction of motion of the particle
- explain how electric and magnetic fields can be used in velocity selection
- understand the origin of the Hall voltage and derive and use the expression $V_H = \frac{BI}{ntq}$
- understand the use of a Hall probe to measure magnetic flux density.

BEFORE YOU START

- A current-carrying conductor in a uniform magnetic field experiences a magnetic force *F*. Write down the factors that affect this force *F* and how you can determine the direction of the force.
- You can get a uniform electric field between two oppositely charged parallel plates. Can you recall and write down the definition for electric field strength *E*?

MOVING PARTICLES

The world of atomic physics is populated by a great variety of particles – electrons, protons, neutrons, positrons, and many more. Many of these particles are electrically charged, and so their motion is influenced by electric and magnetic fields. Indeed, we use this fact to help us to distinguish one particle from another. Figure 25.1 shows the tracks of particles in a detector called a bubble chamber. A photon (no track) has entered from the top and collided with a proton; the resulting spray of nine particles shows up as the gently curving tracks moving downwards. The tracks curve because the particles are charged and are moving in a magnetic field. The tightly wound spiral tracks are produced by electrons that, because their mass is small, are more dramatically affected by the field.

In this chapter, we will look at how charged particles behave in electric and magnetic fields and how this knowledge can be used to control beams of charged particles. At the end of the chapter, we will look at how this knowledge was used to discover the electron and to measure its charge and mass.

Figure 25.1: A bubble chamber image of the tracks of sub-atomic particles. The tracks curve because the charged particles are affected by the presence of a magnetic field.

25.1 Observing the force

You can use your knowledge of how charged particles and electric currents are affected by fields to interpret diagrams of moving particles. You must always remember that, by convention, the direction of conventional electric current is the direction of flow of positive charge. When electrons are moving, the conventional current is regarded as flowing in the opposite direction.

An electron beam tube (Figure 25.2) can be used to demonstrate the magnetic force on moving charged particles. A beam of electrons is produced by an 'electron gun', and magnets or electromagnets are used to apply a magnetic field.

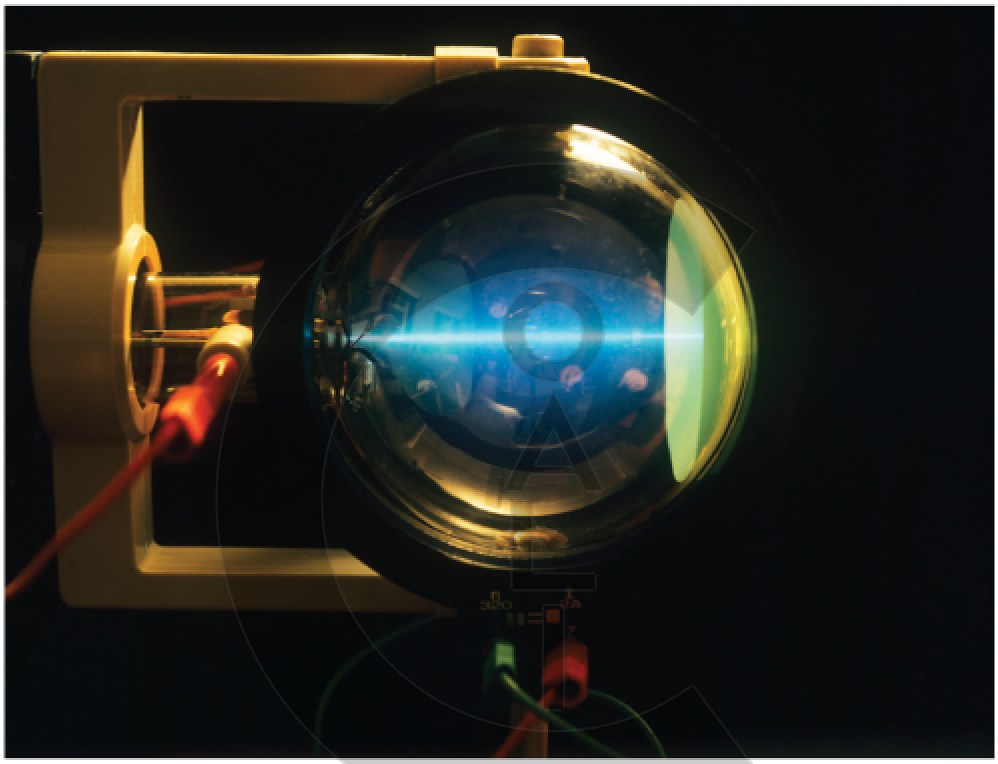


Figure 25.2: An electron beam tube.

You can use such an arrangement to observe the effect of changing the strength and direction of the magnetic field, and the effect of reversing the field.

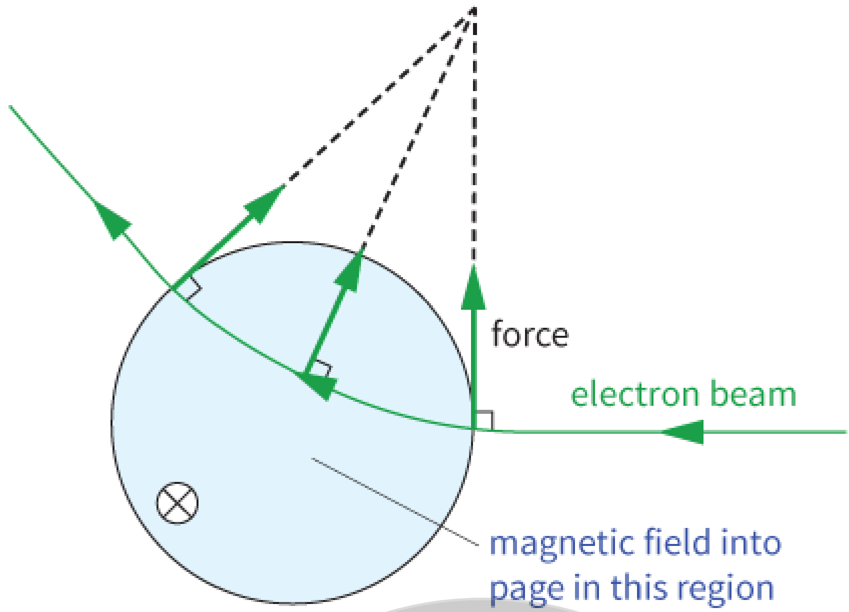


Figure 25.3: A beam of electrons is deflected as it crosses a magnetic field. The magnetic field into the plane of the paper is represented by the cross in the circle.

If you are able to observe a beam of electrons like this, you should find that the force on the electrons moving through the magnetic field can be predicted using Fleming's left-hand rule (see Chapter 24). In Figure 25.3, a beam of electrons is moving from right to left, into a region where a magnetic field is directed into the plane of the paper. Since electrons are negatively charged, they represent a conventional current from left to right. Fleming's left-hand rule predicts that, as the electrons enter the field, the force on them will be upwards and so the beam will be deflected up the page. As the direction of the beam changes, so does the direction of the force. The force due to the magnetic field is always at 90° to the velocity of the electrons. It is this force that gives rise to the motor effect. The electrons in a wire experience a force when they flow across a magnetic field, and they transfer the force to the wire itself. In the past, most oscilloscopes, monitors and television sets made use of beams of electrons. The beams were moved about using magnetic and electric fields, and the result was a rapidly changing image on the screen.

PRACTICAL ACTIVITY 25.1

Electron beam tubes

Figure 25.4 shows the construction of a typical tube. The electron gun has a heated cathode. The electrons have sufficient thermal energy to be released from the surface of the heated cathode. These electrons form a cloud around the cathode. The positively charged anode attracts these electrons, and they pass through the anode to form a narrow beam in the space beyond. The direction of the beam can be changed using an electric field between two plates (as in Figure 25.4), or a magnetic field created by electromagnetic coils.

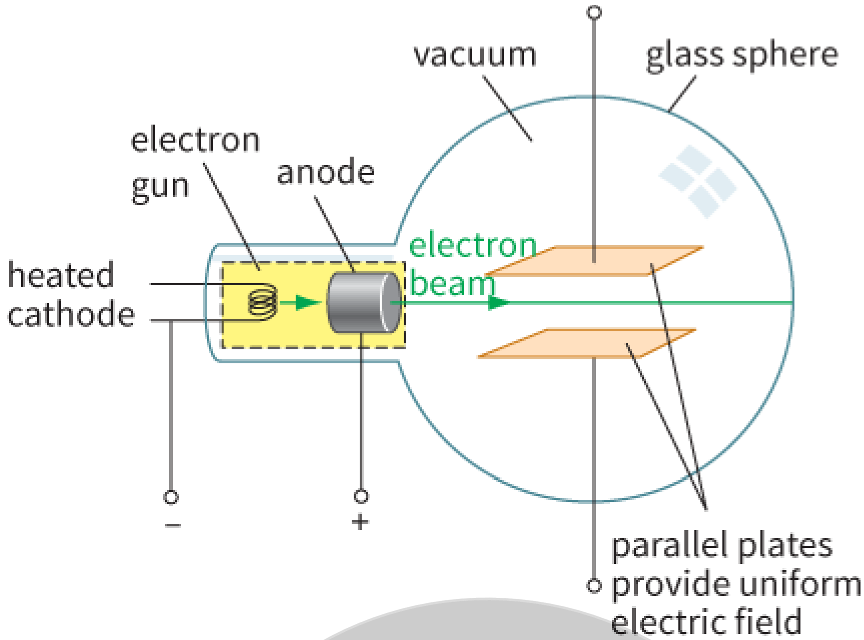


Figure 25.4: The construction of an electron beam tube.

Question

1 Figure 25.5 shows how radiation from a radioactive material passes through a region of uniform magnetic field.

State and explain whether each type of radiation has positive or negative charge, or is uncharged.

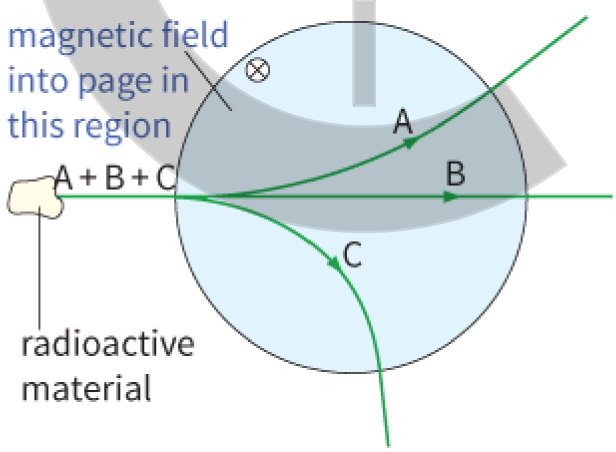


Figure 25.5: Three types of radiation passing through a magnetic field.

Magnetic force on a moving charged particle

Imagine a charged particle moving in a region of uniform magnetic field, with the particle's velocity at right angles to the field. We can make an intelligent guess about the factors that determine the size of the force on the particle (Figure 25.6). It will depend on:

- the magnetic flux density B (strength of the magnetic field)
- the charge Q on the particle

- the speed v of the particle.

The magnetic force F on a moving particle at right angles to a magnetic field is given by the equation:

$$F = BQv$$

The direction of the force can be determined from Fleming's left-hand rule. The force F is always at 90° to the velocity of the particle. Consequently, the path described by the particle will be an arc of a circle.

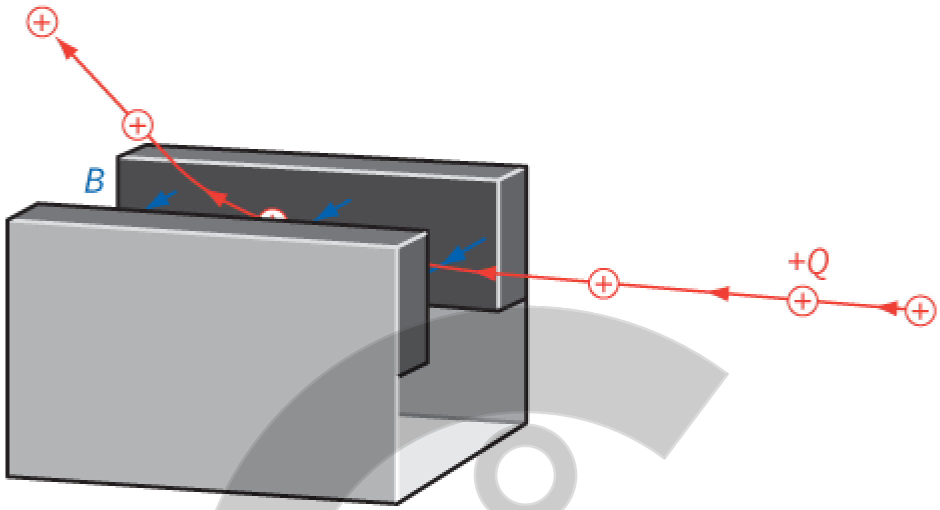


Figure 25.6: The path of a charged particle is curved in a magnetic field.

If the charged particle is moving at an angle θ to the magnetic field, the component of its velocity at right angles to B is $v \sin \theta$. Hence, the equation becomes:

$$F = BQv\sin\theta$$

where B is the magnetic flux density, Q is the charge on the particle, v is the speed of the particle and θ is the angle between the magnetic field and the velocity of the particle.

KEY EQUATION

$$F = BQv\sin\theta$$

Magnetic force F experienced by a charged particle.

We can show that the two equations $F = BIL$ and $F = BQv$ are consistent with one another, as follows.

Since current I is the rate of flow of charge, we can write:

$$I = \frac{Q}{t}$$

Substituting in $F = BIL$ gives:

$$F = \frac{BQL}{t}$$

Now, $\frac{L}{t}$ is the speed v of the moving particle, so we can write:

$$F = BQv$$

For an electron, with a charge of $-e$, the magnitude of the force is:

$$F = Bev \ (e = 1.60 \times 10^{-19} \text{ C})$$

The force on a moving charged particle is sometimes called the '*Bev* force'; it is this force acting on all the electrons in a wire that gives rise to the '*BIL* force'.

Here is an important reminder: the force F is always at right angles to the particle's velocity v, and its direction can be found using Fleming's left-hand rule (Figure 25.7).

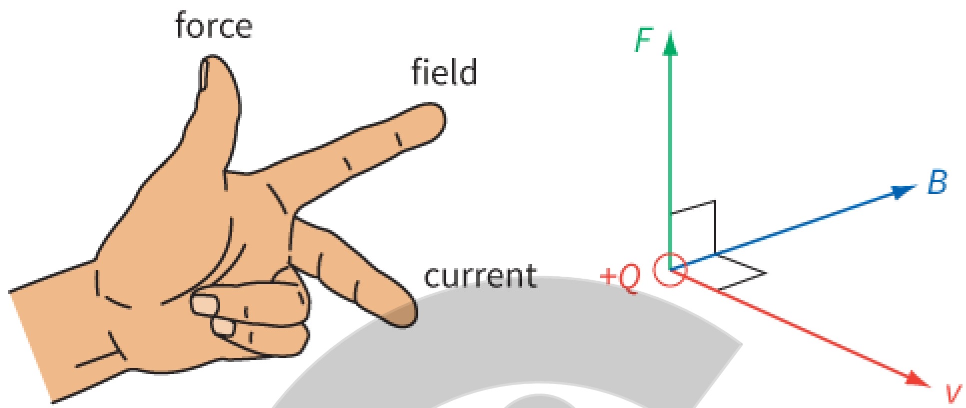


Figure 25.7: Fleming's left-hand rule, applied to a moving positive charge.

Questions

2 An electron is moving at 1.0×10^{6} m s^{-1} in a uniform magnetic field of flux density 0.50 T.

Calculate the force on the electron when it is moving:

a at right angles to the field

b at an angle of 45° to the field.

3 Positrons are particles identical to electrons, except that their charge is positive ($+e$). Use a diagram to explain how a magnetic field could be used to separate a mixed beam consisting of both positrons and electrons.

25.2 Orbiting charged particles

Consider a charged particle of mass m and charge Q moving at right angles to a uniform magnetic field. It will describe a circular path because the magnetic force F is always perpendicular to its velocity. The magnetic force F, provides the **centripetal force** on the particle – the direction of the force is always towards the centre of the circle.

Figure 25.8: In this fine-beam tube, a beam of electrons is bent around into a circular orbit by a uniform magnetic field. The beam is shown up by the presence of a small amount of gas in the tube. (The electrons travel in an anticlockwise direction.)

Figure 25.8 shows a fine-beam tube. In this tube, a beam of fast-moving electrons is produced by an electron gun. This is similar to the cathode and anode shown in Figure 25.4, but in this case the beam is directed vertically downwards as it emerges from the gun. It enters the spherical tube, which has a uniform horizontal magnetic field. The beam is at right angles to the magnetic field and the Bev force pushes it round in a circle.

The fact that the centripetal force is provided by the magnetic force BQv, gives us a clue as to how we can calculate the radius r of the orbit of a charged particle in a uniform magnetic field. The centripetal force is given by:

$$\text{centripetal force} = \frac{mv^2}{r}$$

Therefore

$$BQv = \frac{mv^2}{r}$$

Cancelling and rearranging, you get:

$$r = \frac{mv}{BQ}$$

If the charged particles are electrons, then *Q* is numerically equal to *e*. The equation then becomes:

$$r = \frac{mv}{Be}$$

The momentum *p* of the particle is *mv.* You can therefore write the equation as:

$$p = Ber$$

The equation $r = \frac{mv}{Be}$ shows that:

- faster-moving particles move in bigger circles because $r \propto v$
- particles with greater masses also move in bigger circles because $r \propto m$
- particles with greater charge move in tighter (smaller) circles because $r \propto \frac{1}{Q}$
- a stronger field (greater magnetic flux density) makes the particles move in tighter circles because $r \propto \frac{1}{B}$

These ideas have a variety of scientific applications, such as particle accelerators and mass spectrometers. They can also be used to find the charge-to-mass ratio $\frac{e}{m_e}$ of an electron.

The charge-to-mass ratio of an lectron

Experiments to find the mass of an electron first involve finding the charge-to-mass ratio $\frac{e}{m_e}$. This is known as the specific charge on the electron – the word 'specific' here means 'per unit mass'.

Using the equation for an electron travelling in a circle in a magnetic field, we have $\frac{e}{m_e} = \frac{v}{Br}$. Clearly, measurements of *v*, *B* and *r* are needed to determine $\frac{e}{m_e}$.

There are difficulties in measuring *B* and *r.* For example, it is difficult to directly measure *r* with a ruler outside the tube in Figure 25.8 because of parallax error. Also, *v* must be measured, and you need to know how this is done. One way is to use the potential difference (p.d.) V_{ca} between the cathode and the anode. This p.d. causes each electron to accelerate as it moves from the cathode to the anode. An individual electron has charge $-e$, therefore an amount of work is done on each electron is $e \times V_{ca}$. This is equivalent to the kinetic energy of the electron as it leaves the anode – we assume that the electron has zero kinetic energy at the cathode. Therefore:

$$eV_{ca} = \frac{1}{2}m_e v^2$$

where m_e is the mass of the electron and *v* is the final speed of the electron.

Eliminating *v* from the equations:

$$eV_{ca} = \frac{1}{2}m_e v^2 \text{ and } r = \frac{m_e v}{Be}$$

gives: $\frac{e}{m_e} = \frac{2V_{ca}}{r^2B^2}$

A voltmeter can be used to measure V_{ca}, and if *r* and *B* are known, we can calculate the ratio $\frac{e}{m_e}$. As you shall see shortly, the charge on the electron *e* can be measured more directly, and this allows physicists to calculate the electron mass m_e from the value of $\frac{e}{m_e}$.

WORKED EXAMPLE

1 An electron is travelling at right angles to a uniform magnetic field of flux density 1.2 mT. The speed of the electron is 8.0×10^6 m s^{-1}.

Calculate the radius of circle described by this electron.

$e = 1.60 \times 10^{-19}$ C and $m_e = 9.11 \times 10^{-31}$ kg

Step 1 Calculate the magnetic force on the electron.

$$F = Bev = 1.2 \times 10^{-3} \times 1.60 \times 10^{-19} \times 8.0 \times 10^{6}$$
$$F = 1.536 \times 10^{-15}\ \text{N}$$

Step 2 Use your knowledge of motion in a circle to determine the radius *r*.

$$F = \frac{m_e v^2}{r}$$

Therefore:

$$r = \frac{m_e v^2}{F} = \frac{9.11 \times 10^{-31} \times (8.0 \times 10^{6})^2}{1.536 \times 10^{-15}}$$
$$r = 3.8 \times 10^{-2}\ \text{m}\ (3.8\ \text{cm})$$

Note: You can get the same result by using the equation:

$$r = \frac{m_e v}{Be}$$

Questions

4 Look at the photograph of the electron beam in the fine-beam tube (Figure 25.8).
State the direction is the magnetic field (into or out of the plane of the photograph).

5 The particles in the circular beam shown in Figure 25.8 all travel round in the same orbit.
State what can you deduce about their mass, charge and speed.

6 An electron beam in a vacuum tube is directed at right angles to a magnetic field, so that it travels along a circular path.
Predict the effect on the size and shape of the path that would be produced (separately) by each of the following changes:

- **a** increasing the magnetic flux density
- **b** reversing the direction of the magnetic field
- **c** slowing down the electrons
- **d** tilting the beam, so that the electrons have a component of velocity along the magnetic field.

PRACTICAL ACTIVITY 25.2

The deflection tube

A deflection tube (Figure 25.9) is designed to show a beam of electrons passing through a combination of electric and magnetic fields.

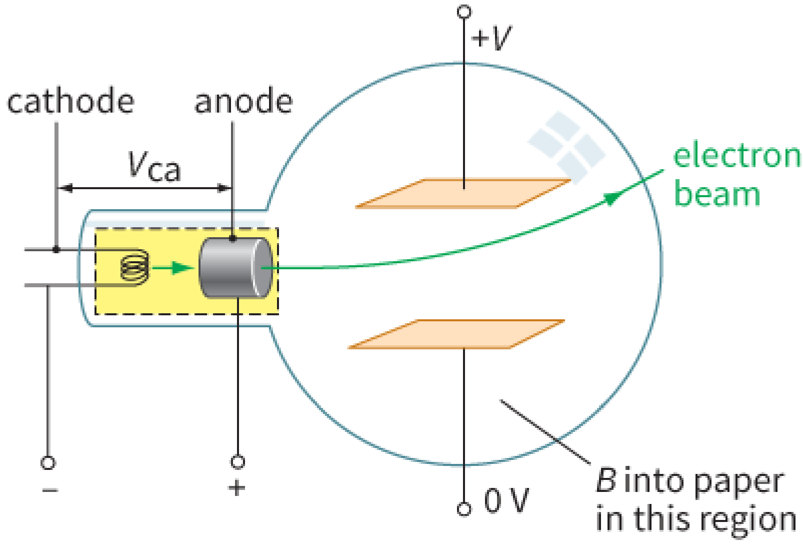


Figure 25.9: The path of an electron beam in a deflection tube.

By adjusting the strengths of the electric and magnetic fields, you can balance the two forces on the electrons, and the beam will remain horizontal. The magnetic field is provided by two vertical coils, called Helmholtz coils (Figure 25.10), which give a very uniform field in the space between them.

When the electron beam remains straight, it follows that the electric and magnetic forces on each electron must have the same magnitude and act in opposite directions.

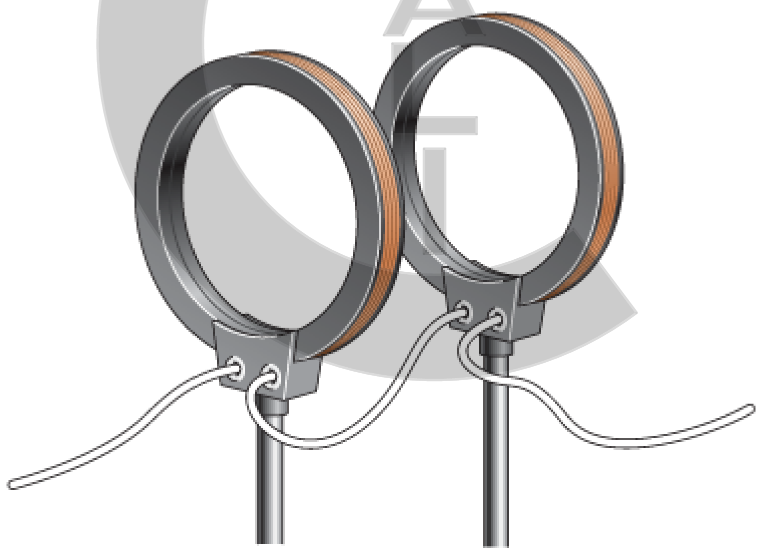


Figure 25.10: A pair of Helmholtz coils is used to give a uniform magnetic field.

Therefore: electric force (upwards) = magnetic force (downwards)

$$eE = Bev$$

where E is the electric field strength between the parallel horizontal plates. The speed v of the electrons is simply related to E and B because e in the expression cancels out. Therefore:

$$v = \frac{E}{B}$$

The electric field strength *E* is given by:

$$E = \frac{V}{d}$$

where *V* is the p.d. between the plates and *d* is the distance between the plates. Therefore:

$$v = \frac{V}{Bd}$$

25.3 Electric and magnetic fields

Now we will consider in detail what happens when an electron beam passes through an electric field and a magnetic field at the same time.

Velocity selection

In a device called a **velocity selector***,* charged particles of a specific velocity are selected using both electric and magnetic fields. This is used in devices such as mass spectrometers where it is essential to produce a beam of charged particles all moving with the same velocity. The construction of a velocity selector is shown in Figure 25.11.

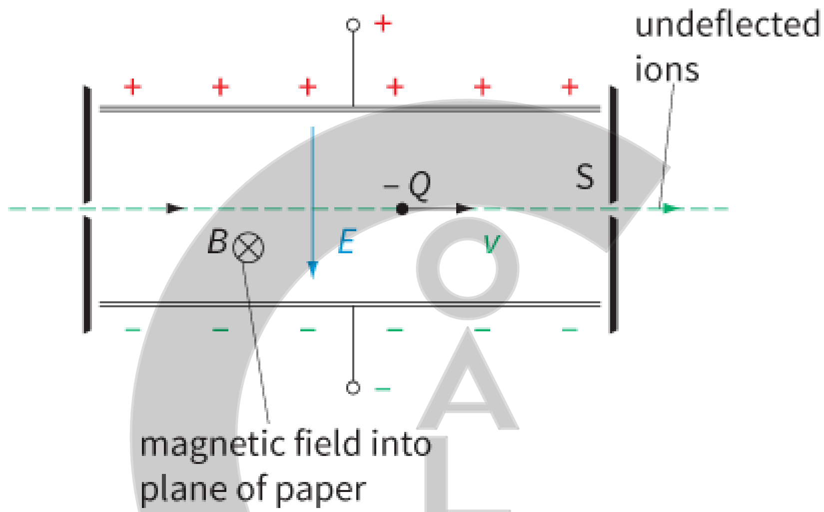


Figure 25.11: A velocity selector – only particles with the correct velocity will emerge through the slit S.

The apparatus is very similar to the deflection tube in Figure 25.9. Two oppositely charged horizontal plates are situated in an evacuated chamber. These plates provide a uniform electric field of strength *E* in the space between the plates.

The region between the plates is also occupied by a uniform magnetic field of flux density *B* that is at right angles to the electric field. Negatively charged particles (electrons or ions) enter from the left. They all have the same charge −*Q* but are travelling at **different speeds**. The magnitude of the electric force *EQ* will be the same on all particles as it does not depend on their speed. However, the magnitude of the magnetic force *BQv* will be greater for those particles that are travelling faster. Hence, for particles travelling at the desired speed *v*, the electric force and the magnetic force must have the same value, but be in opposite directions. The resultant force on the charged particles in the vertical direction must be zero, and all the charged particles with the speed *v* will emerge undeflected from the slit S. Therefore:

$$QE = BQv$$
$$v = \frac{E}{B}$$

If a charged particle has a speed greater than $\frac{E}{B}$, the downward magnetic force on it will be greater than the upward electric force. Thus, it will be deflected downwards and it will hit below slit S.

Note that we do not have to concern ourselves with the gravitational force *mg* acting on the charged particles as this will be negligible compared with the electric and magnetic forces.

Question

7 This question is about the velocity selector shown in Figure 25.11.

a State the directions of the magnetic and electric forces on a positively charged ion travelling towards the slit S.

b Calculate the speed of an ion emerging from the slit S when the magnetic flux density is 0.30 T and the electric field strength is 1.5×10^3 V m^{-1}.

c Explain why ions travelling with a speed greater than your answer to part b will not emerge from the slit.

25.4 The Hall effect

In Chapter 24, you saw how to use a Hall probe to measure magnetic flux density. This probe works on the basis of the **Hall effect**. The Hall effect is the production of a potential difference across an electrical conductor when an external magnetic field is applied in a direction perpendicular to the current.

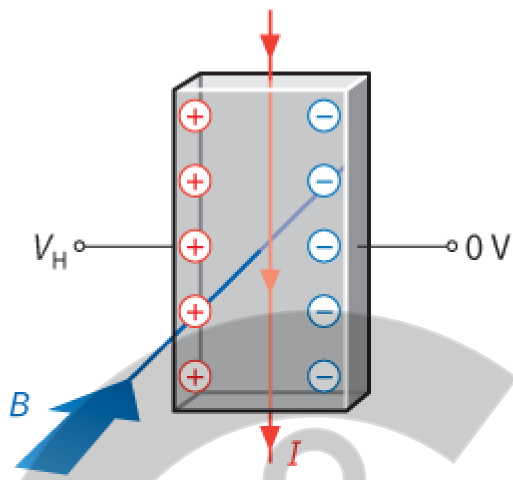


Figure 25.12: A Hall voltage is produced across the sides of the slice of conductor (metal).

Consider a slice of conductor with an external magnetic field applied perpendicular to the direction of the current. If the conductor is a **metal**, then the current is due to the flow of electrons. These electrons will experience a magnetic force, which will make them drift towards one side of the conductor, where they will gather. The opposite side of the slice is deficient of electrons. A potential difference, known as the **Hall voltage**, will be developed across the conductor (Figure 25.12). As you will see later, the Hall voltage V_H for the slice is constant for a given current and is directly proportional to the magnetic flux density *B* of the external magnetic field.

An equation for the Hall voltage

Using what we know about electric current and the forces on electric charges produced by electric and magnetic fields, we can derive an expression for the Hall voltage V_H. Figure 25.13 shows a current-carrying slice of a metal. The Hall voltage is the voltage that appears between the two opposite sides of the slice.

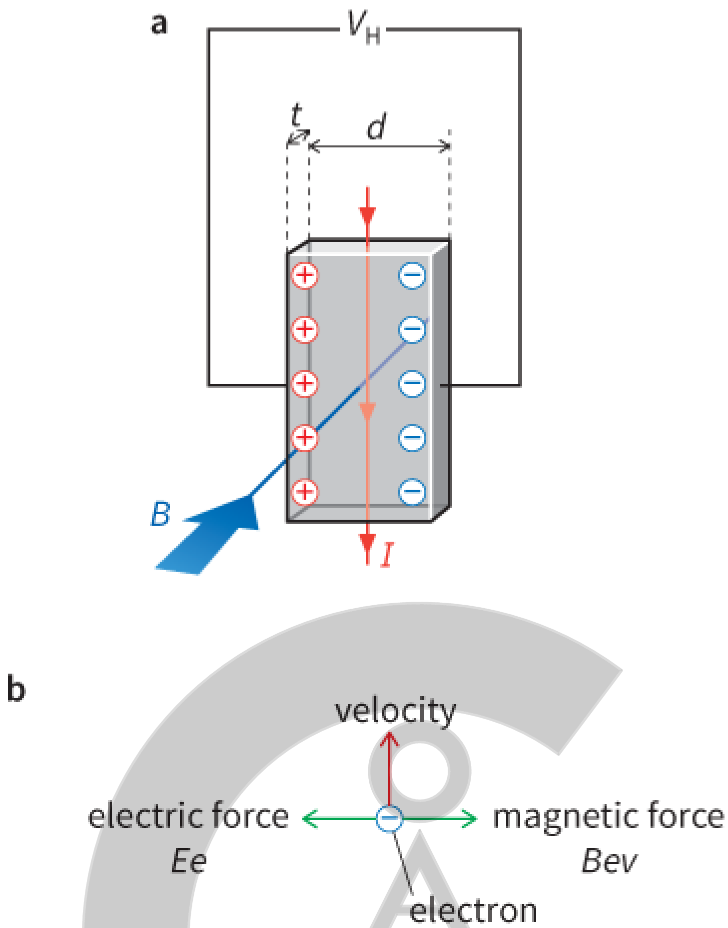


Figure 25.13: a The Hall voltage is measured across the slice of metal. **b** The forces on an electron when the electric and magnetic forces are equal and opposite.

As we have seen, this voltage arises because electrons accumulate on one side of the slice. There is a corresponding lack of electrons on the opposite side – this opposite side may be considered to have a positive charge. As a result, there is an electric field set up within the slice between the two sides. The two charged sides may be treated as oppositely charged parallel plates – see Chapter 21. Therefore, the electric field strength E is related to the Hall voltage V_H by:

$$E = \frac{V_H}{d}$$

where d is the width of the slice.

Now, imagine a single electron as it travels with drift velocity v through the slice. The magnetic field is into the plane of the paper, so this electron will experience a magnetic force Bev to the right. It will also experience an electric force Ee to the left.

When the current first starts to flow, there is no Hall voltage and so electrons are pushed to the right by the magnetic field. However, as the charge on the right-hand side builds up, so does the internal electric field and this pushes the electrons in the opposite direction to the magnetic force. Soon, an equilibrium situation is reached, the resultant force on each electron is zero and the electrons are undeflected. Now we can equate the two forces:

$$eE = Bev$$

Substituting for E we have:

$$\frac{eV_H}{d} = Bev$$

Now recall from Chapter 8 that the current I is related to the mean drift velocity v of the electrons by $I = nAve$, where A is the cross-sectional area of the conductor and n is the number density of charge carriers (in this case, electrons). So, we can substitute for v to get:

$$\frac{eV_H}{d} = \frac{BeI}{nAe}$$

Making V_H the subject of the equation (and cancelling e) gives:

$$V_H = \frac{BId}{nAe}$$

The cross-sectional area A of each side-face of the slice is:

$$A = d \times t$$

where t is the thickness of the slice.

Substituting and cancelling gives:

$$V_H = \frac{BI}{nte}$$

This equation for the Hall voltage shows that V_H is directly proportional to the magnetic flux density B for a given slice and current. That is what makes the Hall effect so useful for measuring *B.*

To get a large voltage, it would be desirable to have a material with a smaller value for n compared with metals. Hall probes use a very thin slice of semiconductor. Semiconductors have a number density many thousands of times smaller than metals, hence the Hall voltage will be thousands of times larger.

In some semiconductors, the charge carriers are not electrons, but positively charged particles referred to as 'holes'. We can write a more general equation for the Hall voltage replacing e with q, where q is the charge of an individual charge carrier. This gives $V_H = \frac{BI}{ntq}$

KEY EQUATION

Hall voltage:

$$VH = BI/(ntq)$$

You must learn how to derive this equation.

Positive charges will be deflected in the opposite direction to negative charges, and so we can determine whether the charge carriers are positive or negative by the sign of the Hall voltage.

Questions

8 A Hall probe is designed to operate with a steady current of 0.020 A in a semiconductor slice of thickness 0.05 mm. The number density of charge carriers (electrons) in the semiconductor is 1.5×10^{23} m^{-3}.

- **a** Calculate the Hall voltage that will result when the probe is placed at right angles to a magnetic field of flux density 0.10 T.

  (Elementary charge $e = 1.60 \times 10^{-19}$ C.)
- **b** Explain why the current in the Hall probe must be maintained at a constant value.

9 Suggest how the Hall effect could be used to determine the number density of charge carriers n in a semiconducting material.

25.5 Discovering the electron

Today, we know a lot about electrons and we use the idea of electrons to explain all sorts of phenomena, including electric current and chemical bonding. However, at the end of the 19th century, physicists were only just beginning to identify the tiny particles that make up matter.

One of the leaders in this field was the English physicist J.J. Thomson (Figure 25.14). In the photograph, he is shown with the deflection tube that he used in his discovery of the electron.

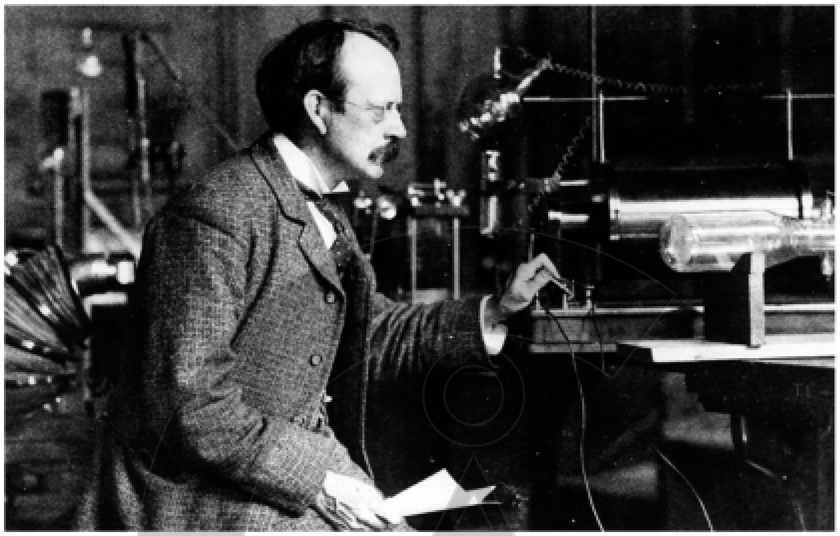


Figure 25.14: J.J. Thomson – in 1897, he discovered the electron using the vacuum tube shown here.

His tube was similar in construction to the deflection tube shown in Figure 25.9. At one end was an electron gun that produced a beam of electrons (which he called 'cathode rays'). Two metal plates allowed him to apply an electric field to deflect the beam, and he could place magnets outside the tube to apply a magnetic force to the beam. Here is a summary of his observations and what he concluded from them:

- The beam in his tube was deflected towards a positive plate and away from a negative plate, so the particles involved must have negative charge. This was confirmed by the deflection of the beam by a magnetic field.
- When the beam was deflected, it remained as a tight, single beam rather than spreading out into a broad beam. This showed that, if the beam consisted of particles, they must all have the same mass, charge and speed. (Lighter particles would have been deflected more than heavier ones; particles with greater charge would be deflected more, and faster particles would be deflected less.)
- By applying both electric and magnetic fields, Thomson was able to balance the electric and magnetic forces so that the beam in the tube remained straight. He could then calculate the charge-to-mass ratio $\frac{e}{m_e}$ for the particles he had discovered. Although he did not know the value of either e or m_e individually, he was able to show that the particles concerned must be much lighter than atoms. They were the particles that we now know as electrons. In fact, for a while, Thomson thought that atoms were made up of thousands of electrons, although his ideas could not explain how so many negatively charged particles could combine to produce a neutral atom.

The magnitude of the charge e of an electron is very small (1.60×10^{-19} C) and difficult to measure. The American physicist Robert Millikan devised an ingenious way to do it. He observed electrically charged droplets of oil as they moved in electric and gravitational fields and found that they all had a charge that was a small integer multiple of a particular value, which he took to be the magnitude of the charge on a single electron, e.

Having established a value for e, he could combine this with Thomson's value for $\frac{e}{m_e}$ to calculate the electron mass m_e.

Question

10 The charge-to-mass ratio $\frac{e}{m_e}$ for the electron is 1.76×10^{11} C kg^{-1}.

Calculate the mass of the electron using $e = 1.60 \times 10^{-19}$ C.

REFLECTION

Without looking at your textbook, summarise the similarities between a velocity-selector and the Hall effect. Compare your summary with a fellow learner. Did you miss out any key ideas?

Make a list of all the equations leading to $V_H \propto B$ for a Hall probe.

What things might you want more help with in this topic?

SUMMARY

The magnetic force on a charged particle moving at right angles to a magnetic field is given by the equation: $F = BQv$. For an electron, $Q = e$.

For charged particle travelling at an angle θ to the magnetic field, the force is given by the equation:

$$F = BQv \sin\theta$$

The direction of the force experienced by a charge moving in a uniform magnetic field can be determined using Fleming's left-hand rule.

A charged particle entering at right angles to a uniform magnetic field describes a circular path because the magnetic force is perpendicular to the velocity.

For an electron describing a circular path in a uniform magnetic field, the centripetal force is provided by Bev. Therefore:

$$Bev = \frac{mv^2}{r}$$

In a velocity selector, the speed of an undeflected charged particle in a region where electric and magnetic fields are at right angles is given by the equation:

$$v = \frac{E}{B}$$

This speed is independent of the charge of the particle.

In the Hall effect, a potential difference is produced across an electrical conductor when an external magnetic field is applied in a direction perpendicular to the direction of the current.

The Hall voltage is given by:

$$V_H = \frac{BI}{ntq}$$

Note:

$V_H \propto$ magnetic flux density B

$V_H \propto$ current I

$$V_H \propto \frac{1}{\text{number density of charce carriers } n}$$

$$V_H \propto \frac{1}{\text{thickness of slice } t}$$

In a Hall probe, $V_H \propto B$ because the current in the slice is constant.

A Hall probe uses a semiconducting material rather than a metal because the smaller number density of charge carriers gives a larger Hall voltage.

EXAM-STYLE QUESTIONS

1 A scientist is doing an experiment on a beam of electrons travelling at right angles to a uniform magnetic field of flux density *B*. The graph shows the variation of the magnetic force *F* acting on an electron with the speed *v* of the electron.

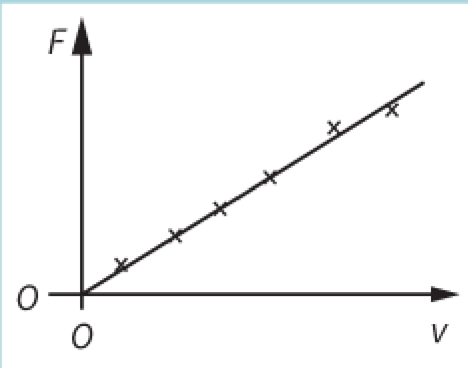


Figure 25.15

The gradient of the graph is *G*. The magnitude of the charge on the electron is *e*.

What is the correct relationship for the magnetic flux density *B*? **[1]**

A $B = G$

B $B = G \times e$

C $B = \frac{G}{e}$

D $B = \frac{e}{G}$

2 The magnetic force *BQv* causes an electron to travel in a circle in a uniform magnetic field.

Explain why this force does not cause an increase in the speed of the electron. **[3]**

3 An electron beam is produced from an electron gun in which each electron is accelerated through a potential difference (p.d.) of 1.6 kV. When these electrons pass at right angles through a magnetic field of flux density 8.0 mT, the radius of curvature of the electron beam is 0.017 m.

Calculate the ratio $\frac{e}{m_e}$ (known as the specific charge of the electron). **[4]**

4 Two particles, an α-particle and a β^--particle, are travelling through a uniform magnetic field. They have the same speed and their velocities are at right angles to the field. Determine the ratio of:

a the mass of the α-particle to the mass of the β^--particle **[2]**

b the charge of the α-particle to the charge of the β^--particle **[2]**

c the force on the α-particle to the force on the β^--particle **[2]**

d the radius of the α-particle's orbit to the radius of the β^--particle's orbit. **[2]**

[Total: 8]

5 A moving charged particle experiences a force in an electric field and also in a magnetic field. State two differences between the forces experienced in the two types of field. **[2]**

6 This diagram shows the path of an electron as it travels in air. The electron rotates

clockwise around a uniform magnetic field into the plane of the paper, but the radius of the orbit decreases in size.

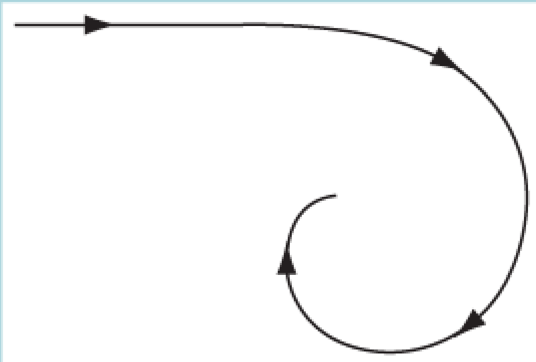


Figure 25.16

a **i** Explain the origin of the force that causes the electron to spiral in this manner. **[2]**

ii Explain why the radius of the circle gradually decreases. **[2]**

b At one point in the path, the speed of the electron is 1.0×10^7 m s^{-1} and the magnetic flux density is 0.25 T. Calculate:

i the force on an electron at this point due to the magnetic field **[2]**

ii the radius of the arc of the circular path at this point. **[2]**

[Total: 8]

7 This diagram shows an arrangement to deflect protons from a source to a detector using a magnetic field. The charge on each proton is $+e$. A uniform magnetic field exists only within the area shown. Protons move from the source to the detector in the plane of the paper.

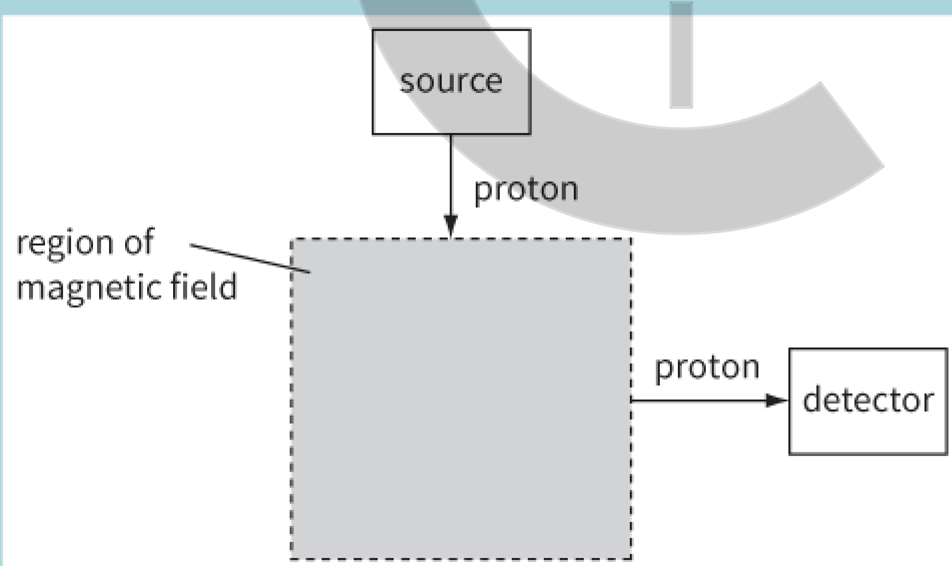


Figure 25.17

a **i** Copy the diagram and sketch the path of a proton from the source to the detector. Draw an arrow at two points on the path to show the direction of the force on the proton produced by the magnetic field. **[3]**

ii State the direction of the magnetic field within the area shown. **[1]**

b The speed of a proton as it enters the magnetic field is 4.0×10^6 m s^{-1}. The magnetic flux density is 0.25 T. Calculate:

i the magnitude of the force on the proton caused by the magnetic field **[1]**

ii the radius of curvature of the path of the proton in the magnetic field. **[2]**

c Two changes to the magnetic field in the area shown are made. These changes allow an electron with the same speed as the proton to be deflected along the same path as the proton. State the two changes made. **[2]**

[Total: 9]

8 This diagram shows a thin slice of semiconductor material carrying a current in a magnetic field at right angles to the current.

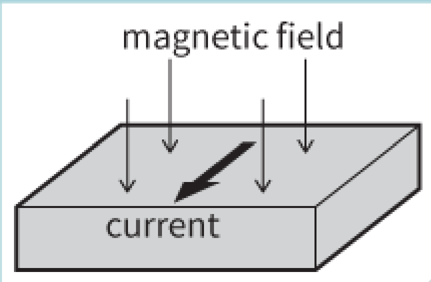


Figure 25.18

a The current in the slice is due to the movement of free electrons.

i Add + and − signs to the diagram to show the charge separation caused by the Hall effect. Explain why the charges separate. **[3]**

ii Explain how an electron can eventually move in a straight line along the slice. **[1]**

b The Hall voltage is measured using the same slice of semiconductor, the same current and the same magnetic field, but with the laboratory at two temperatures, one significantly higher than the other.

Describe and explain the changes in the magnitude of the number density, the drift velocity of the charge carriers and the Hall voltage in the two experiments. **[5]**

[Total: 9]

9 This diagram shows an electron tube. Electrons emitted from the cathode accelerate towards the anode and then pass into a uniform electric field created by two oppositely charged horizontal metal plates.

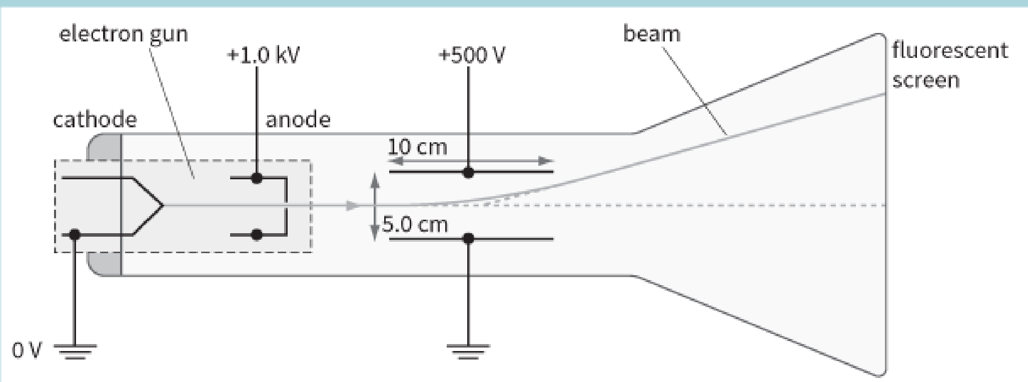


Figure 25.19

a **i** Explain why the beam curves upwards between the plates. **[2]**

ii Explain how the pattern formed on the fluorescent screen shows that all the electrons have the same speed as they leave the anode. **[2]**

b Write down an equation relating the speed of the electrons v to the potential difference V_{ac} between the anode and the cathode. **[1]**

c The deflection of the beam upwards can be cancelled by applying a suitable uniform magnetic field in the space between the parallel plates.

i State the direction of the magnetic field for this to happen. **[1]**

ii Write down an equation relating the speed of the electrons v, the electric field strength E that exists between the plates and the magnetic flux density B needed to make the electrons pass undeflected between the plates. **[2]**

iii Determine the value of B required, using the apparatus shown in the diagram, given that for an electron the ratio $\frac{e}{m_e} = 1.76 \times 10^{11}\ \text{C Kg}^{-1}$. **[4]**

[Total: 12]

10 Protons and helium nuclei from the Sun pass into the Earth's atmosphere above the poles, where the magnetic flux density is 6.0×10^{-5} T. The particles are moving at a speed of 1.0×10^{6} m s^{-1} at right angles to the magnetic field in this region. The magnetic field can be assumed to be uniform.

a Calculate the radius of the path of a proton as it passes above the Earth's pole. **[3]**

mass of a helium nucleus = 6.8×10^{-27} kg

charge on a helium nucleus = 3.2×10^{-19} C

b Sketch a diagram to show the deflection caused by the magnetic field to the paths of a proton and of a helium nucleus that both have the same initial velocity as they enter the magnetic field. **[2]**

[Total: 5]

11 This diagram shows a thin slice of metal of thickness t and width d. The metal slice is in a magnetic field of flux density B and carries a current I, as shown.

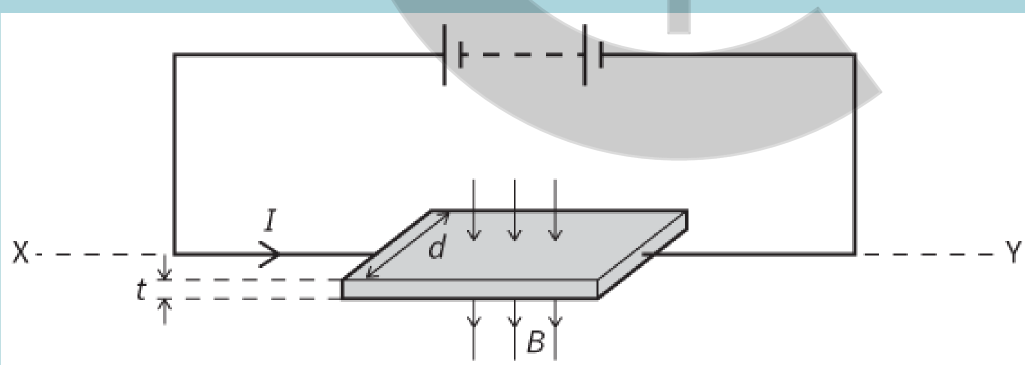


Figure 25.20

a Copy the diagram and mark:

i the side of the slice that becomes negative because of the Hall effect **[1]**

ii where a voltmeter needs to be placed to measure the Hall voltage. **[1]**

b Derive an expression for the Hall voltage in terms of I, B, t, the number density of the charge carriers n in the metal and the charge e on an electron. **[3]**

c Given that I = 40 mA, d = 9.0 mm, t = 0.030 mm, B = 0.60 T, $e = 1.6 \times 10^{-19}$ C and $n = 8.5 \times 10^{28}$ m^{-3}, calculate:

i the mean drift velocity v of the free electrons in the metal **[2]**

ii the Hall voltage across the metal slice **[2]**

iii the percentage uncertainty in the mean drift velocity v of the electrons, assuming the percentage uncertainties in the quantities are as shown. **[1]**

Quantity	**% uncertainty**
Current I	1.3
Width d	2.5
Thickness t	3.0
Number density of charge carriers n	0.2

Table 25.1

d i Explain why, in terms of the movement of electrons, the Hall voltage increases when I increases. **[2]**

ii A Hall probe used to determine the magnetic flux density of a magnetic field uses a thin slice of a semiconductor rather than metal. Explain why a semiconductor is used. **[2]**

e Explain why, when the slice of metal is rotated about the horizontal axis XY, the Hall voltage varies between a maximum positive value and a maximum negative value. **[2]**

[Total: 16]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
use Fleming's left-hand rule to determine the direction of the force on a charge moving in a magnetic field	25.1			
recall and use: $F = BQv \sin\theta$	25.1			
describe the motion of a charged particle moving in a uniform magnetic field perpendicular to the direction of motion of the particle	25.2			
explain how electric and magnetic fields can be used in velocity selection of charged particles	25.3			
understand the Hall effect and the origin of the Hall voltage	25.4			
derive and use the expression for Hall voltage: $V_H = \frac{BI}{ntq}$	25.4			
use a Hall probe to measure magnetic flux density.	25.4			

⟩ Chapter 26

Electromagnetic induction

LEARNING INTENTIONS

In this chapter you will learn how to:

- define magnetic flux as the product of the magnetic flux density and the cross-sectional area perpendicular to the direction of the magnetic flux density
- recall and use $\Phi = BA$
- understand and use the concept of magnetic flux linkage
- understand and explain experiments that demonstrate:
  - that a changing magnetic flux can induce an e.m.f. in a circuit
  - that the direction of the induced e.m.f. opposes the change producing it
  - the factors affecting the magnitude of the induced e.m.f.
- recall and use Faraday's and Lenz's laws of electromagnetic induction.

BEFORE YOU START

- In this chapter, knowledge of magnetic fields is going to be important. Can you work out whether a field is uniform or not from the field pattern? How?
- Physical quantities introduced in this chapter may sound the same, but are very different. Remember the definition for magnetic flux density *B* and its unit (tesla, T).

GENERATING ELECTRICITY

Most of the electricity we use is generated by electromagnetic induction. This process goes on in the generators at work in power stations, in wind turbines (Figure 26.1) and, on a much smaller scale, in bicycle dynamos. It is the process whereby a conductor and a magnetic field are moved relative to each other to induce, or generate, a current or electromotive force (e.m.f.).

One of the most important principles in physics is the idea of conservation of energy. You cannot just produce electrical energy from nowhere. In the case of a generator or a dynamo, how is the electrical energy produced?

Figure 26.1: This giant wind turbine uses electromagnetic induction to produce electricity. Look for the two engineers at work. (You can identify them by their white helmets.) This gives you an idea of the size of the generator.

26.1 Observing induction

You can carry out some simple experiments to observe features of electromagnetic induction. These are described in Practical Activity 26.1.

PRACTICAL ACTIVITY 26.1: OBSERVING INDUCTION

For each experiment, try to predict what you will observe before you try the experiment.

Experiment 1

Connect a small electric motor to a moving-coil voltmeter (Figure 26.2). Spin the shaft of the motor and observe the deflection of the voltmeter. What happens when you spin the motor more slowly? What happens when you stop? Usually, we connect a motor to a power supply and it turns. In this experiment, you have turned the motor and it generates a voltage across its terminals. A generator is like an electric motor working in reverse.

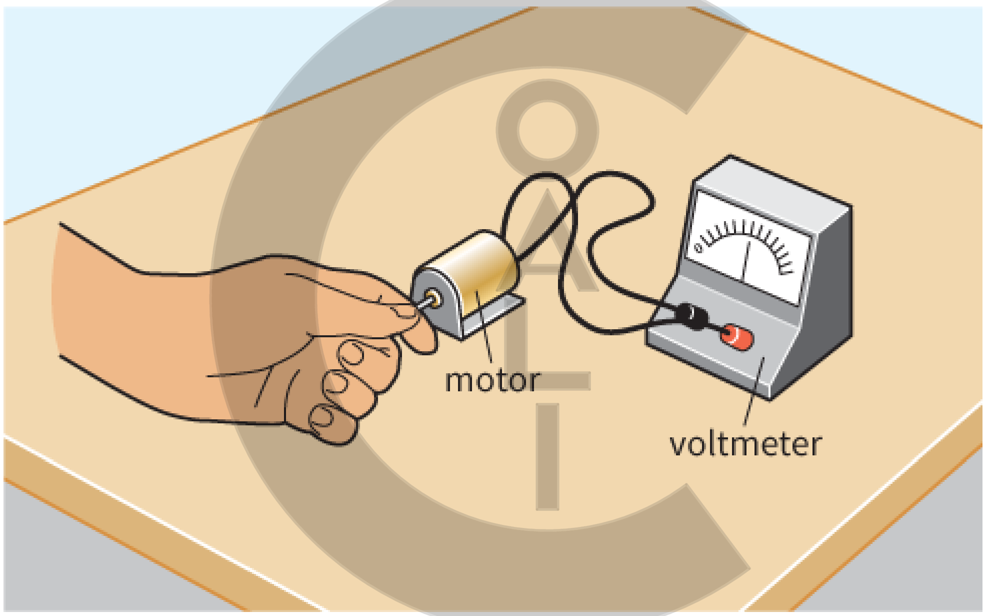


Figure 26.2: A motor works in reverse as a generator.

Experiment 2

Connect a coil to a sensitive microammeter (Figure 26.3). Move a bar magnet in towards the coil. Hold it still, and then remove it. How does the deflection on the meter change? Try different speeds, and the opposite pole of the magnet. Try weak and strong magnets.

With the same equipment, move the coil towards the magnet and observe the deflection of the meter.

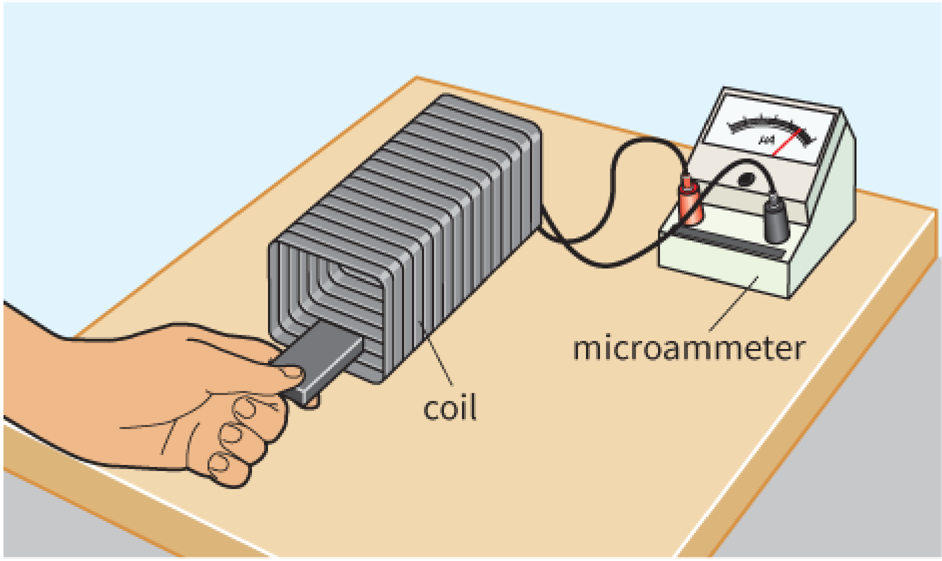


Figure 26.3: A magnet moving near a coil generates a small current.

Experiment 3

Connect a long wire to a sensitive microammeter. Move the middle section of the wire up and down through the magnetic field between the magnets (Figure 26.4). Double up the wire so that twice as much of it passes through the magnetic field. What happens to the meter reading now? How can you form the wire into a loop to give twice the deflection on the meter?

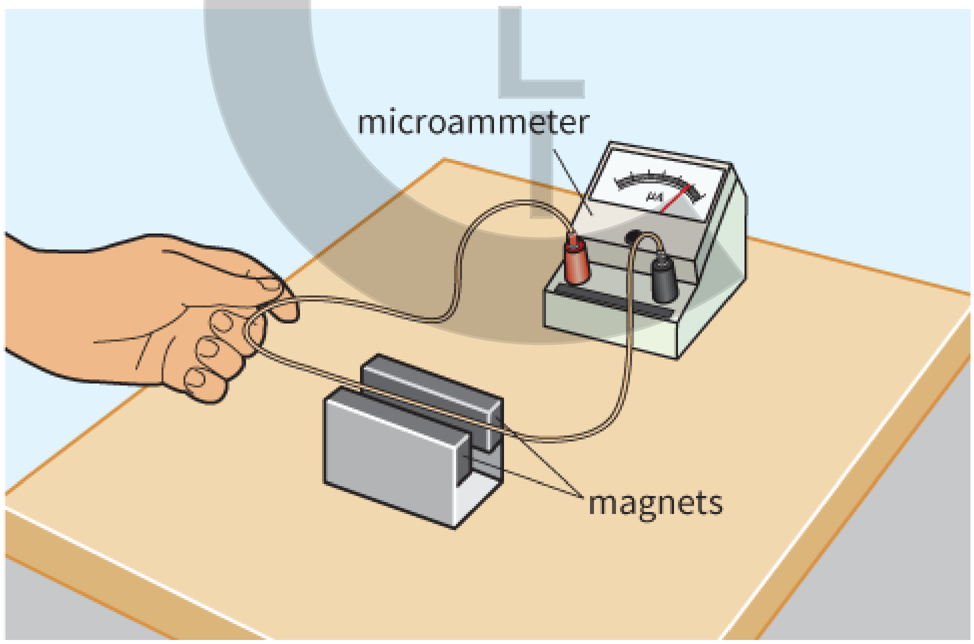


Figure 26.4: Investigating the current induced when a wire moves through a magnetic field.

Factors affecting induced e.m.f

In all the experiments described in Practical Activity 26.1, you have seen an electric current caused by an induced e.m.f. In each case, there is a magnetic field and a conductor. When you move the magnet, or the conductor, there is an induced e.m.f. When you stop, the current stops.

From the three experiments, you should see that the size of the induced e.m.f. depends on several factors.

For a straight wire, the induced e.m.f. depends on the:

- magnitude of the **magnetic flux density**
- length of the wire in the field
- speed of the wire moving across the magnetic field.

For a coil of wire, the induced e.m.f. depends on the:

- magnitude of the magnetic flux density
- cross-sectional area of the coil
- angle between the plane of the coil and the magnetic field
- number of turns of wire
- rate at which the coil turns in the field.

26.2 Explaining electromagnetic induction

You have seen that **relative** movement of a conductor and a magnetic field induces a current in the conductor when it is part of a complete circuit. In the experiments in Practical Activity 26.1, the meter was used to complete the circuit. Now we need to think about how to explain these observations, using what we know about magnetic fields.

Cutting magnetic field lines

Start by thinking about a simple bar magnet. It has a magnetic field in the space around it. We represent this field by magnetic field lines. Now think about what happens when a wire is moved into the magnetic field (Figure 26.5). As it moves, it cuts across the magnetic field. Remove the wire from the field, and again it must cut across the field lines, but in the opposite direction.

We think of this cutting of a magnetic field by a conductor as the effect that gives rise to current caused by induced e.m.f in the conductor. It doesn't matter whether the conductor is moved through the magnetic field or the magnet is moved past the conductor, the result is the same–there will be an induced e.m.f.

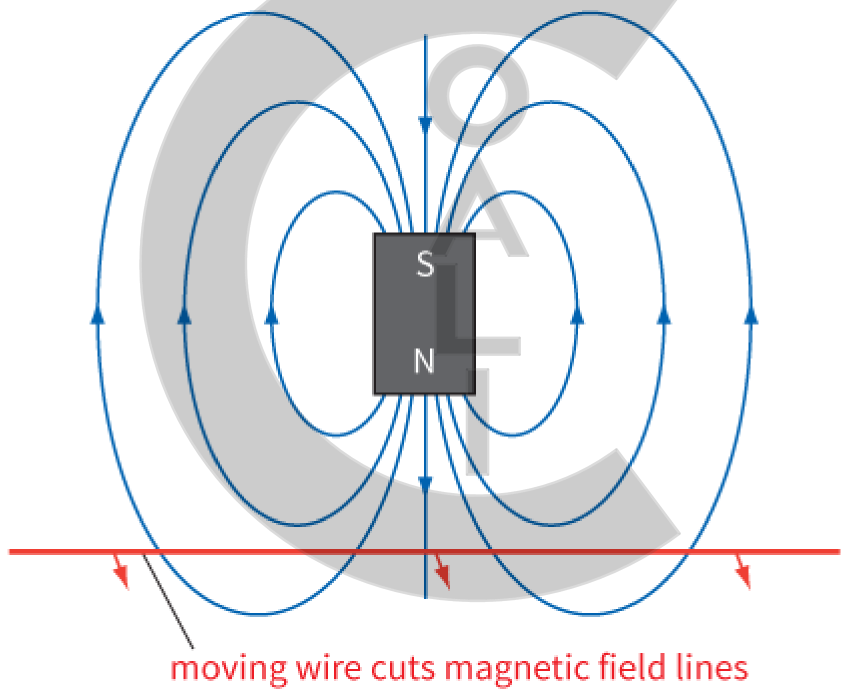


Figure 26.5: Inducing a current by moving a wire through a magnetic field.

The effect is more noticeable if we use a coil of wire. For a coil of *N* turns, the effect is *N* times greater than for a single turn of wire. With a coil, it is helpful to imagine the number of field lines linking the coil. If there is a change in the number of field lines that pass through the coil, an e.m.f. will be induced across the ends of the coil (or there will be a current caused by induced e.m.f if the coil forms part of a complete circuit).

Figure 26.6 shows a coil near a magnet. When the coil is outside the field, there are no magnetic field lines linking the coil. When it is inside the field, field lines link the coil. Moving the coil into or out of the field changes this linkage of field lines, and this induces an e.m.f. across the ends of the coil. Field lines linking the coil is a helpful starting point in our understanding of induced e.m.f. However, as you will see later, a more sophisticated idea of magnetic flux is required for a better understanding of how an e.m.f. is generated in a circuit.

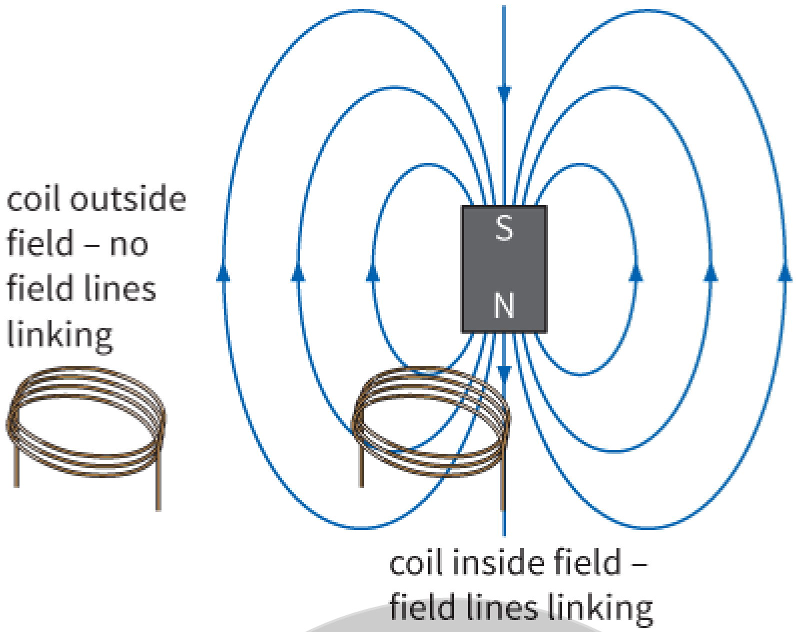


Figure 26.6: The field lines passing through a coil changes as it is moved in and out of a magnetic field.

Question

1 Use the idea of a conductor cutting magnetic field lines to explain how a current is caused by induced e.m.f. in a bicycle generator (Figure 26.7).

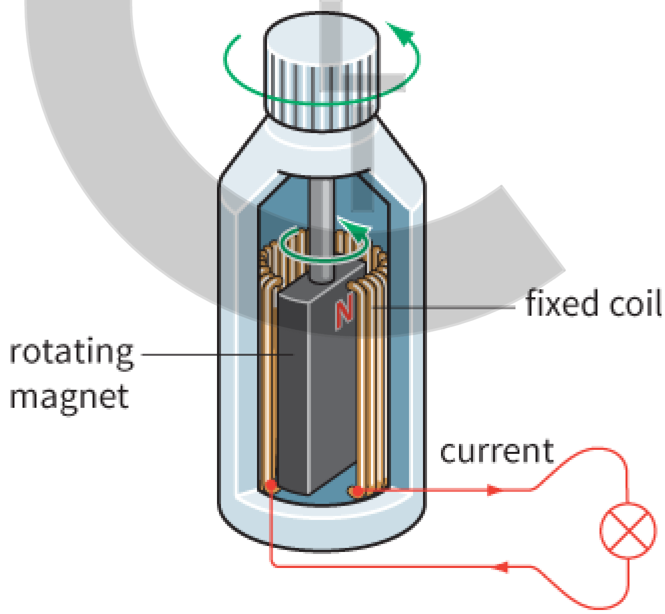


Figure 26.7: In a bicycle generator, a permanent magnet rotates inside a fixed coil of wire. For Question 1.

Current direction (extension)

How can we predict the direction of the current caused by induced e.m.f? For the motor effect in Chapter 24, we used Fleming's left-hand (motor) rule. Electromagnetic induction is like the mirror image of the motor effect. Instead of a current producing a force on a current-carrying conductor in a magnetic field, we provide an external force on a conductor by moving it through a magnetic field and this induces a current in the conductor.

So you should not be too surprised to find that we use the mirror image of the left-hand rule: **Fleming's right-hand (generator) rule**.

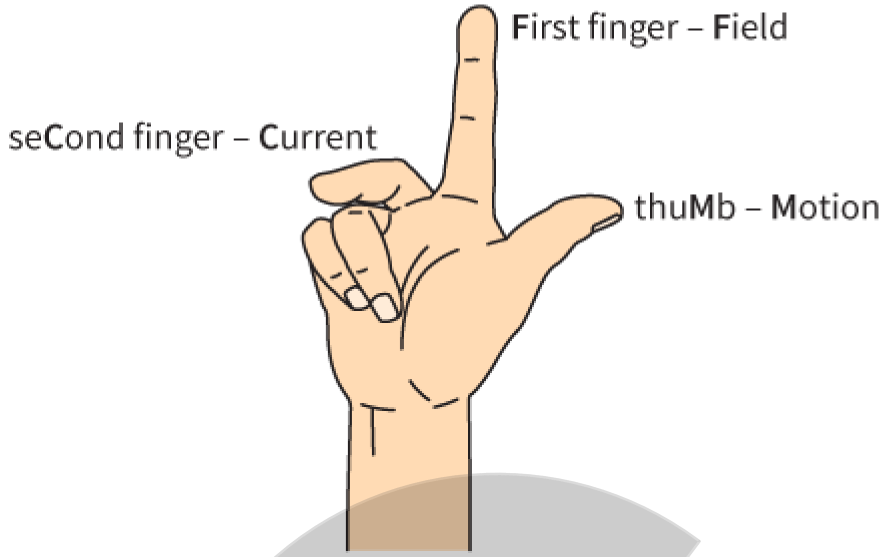


Figure 26.8: Fleming's right-hand (generator) rule.

The three fingers represent the same things again (Figure 26.8):

- thu**M**b–direction of **M**otion
- **F**irst finger–direction of external magnetic **F**ield
- se**C**ond finger–direction of (conventional) **C**urrent caused by induced e.m.f

In the example shown in Figure 26.9, the conductor is being moved downwards across the magnetic field. There is a current caused by induced e.m.f. in the conductor as shown. Check this with your own right hand. You should also check that reversing the movement or the field will result in the current flowing in the opposite direction.

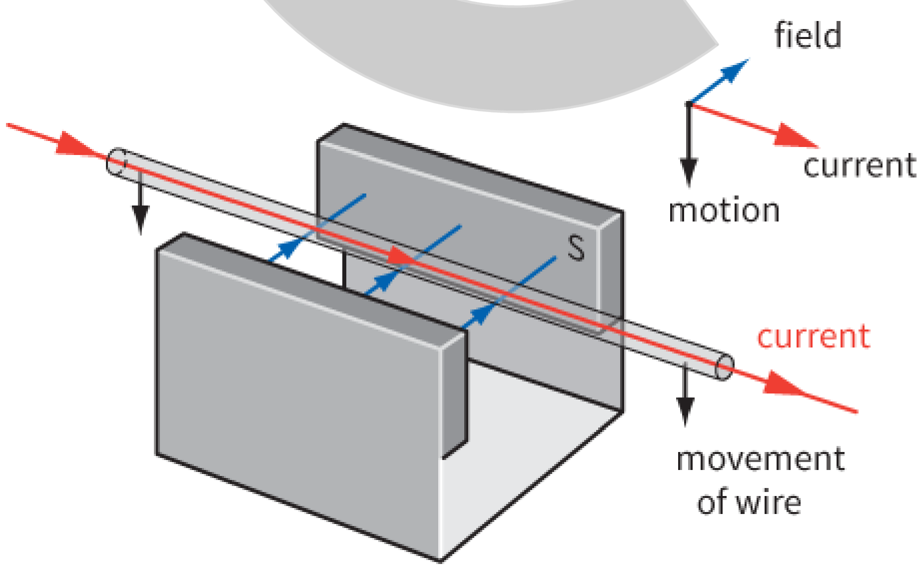


Figure 26.9: Deducing the direction of the current using Fleming's right-hand rule. (The wire shown is a part of a complete circuit or loop.)

Induced e.m.f.

When a conductor is not part of a complete circuit, there cannot be a current induced by e.m.f. Instead, negative charge will accumulate at one end of the conductor, leaving the other end positively charged. We have induced an e.m.f. across the ends of the conductor.

Is e.m.f. the right term? Should it be potential difference (voltage)? In Chapter 8, you saw the distinction between voltage and e.m.f. The term e.m.f. is the correct one here because, by pushing the wire through the magnetic field, work is done and this is transformed into electrical energy. Think of this in another way: since we could connect the ends of the conductor so that there is a current in some other component, such as a lamp, which would light up, it must be an e.m.f. – a source of electrical energy.

Figure 26.10 shows how an e.m.f. is induced. Notice that, within the conductor, conventional current is from negative to positive, in the same way as inside a battery or any other source of e.m.f. In reality, the free electrons within the conductor travel from right to left, making the left-hand side of the conductor negative. What causes these electrons to move? Moving the conductor is equivalent to giving a free electron within the conductor a velocity in the direction of this motion. This electron is in an external magnetic field and hence experiences a magnetic force *Bev* from right to left. Check this out for yourself.

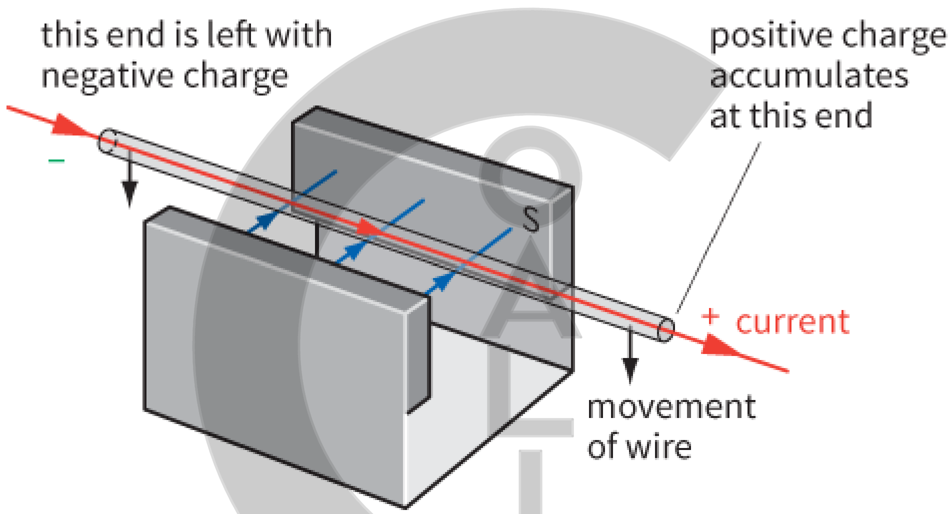


Figure 26.10: An e.m.f. is induced across the ends of the conductor.

Questions

2 The coil in Figure 26.11 is rotating in a uniform magnetic field.

Predict the direction of the current caused by induced e.m.f. in sections AB and CD.

State which terminal, X or Y, will become positive.

3 When an aircraft flies from east to west, its wings are an electrical conductor cutting across the Earth's magnetic flux. In the northern hemisphere, state which wingtip (left or right) will become positive.

State and explain what will happen to this wingtip in the southern hemisphere.

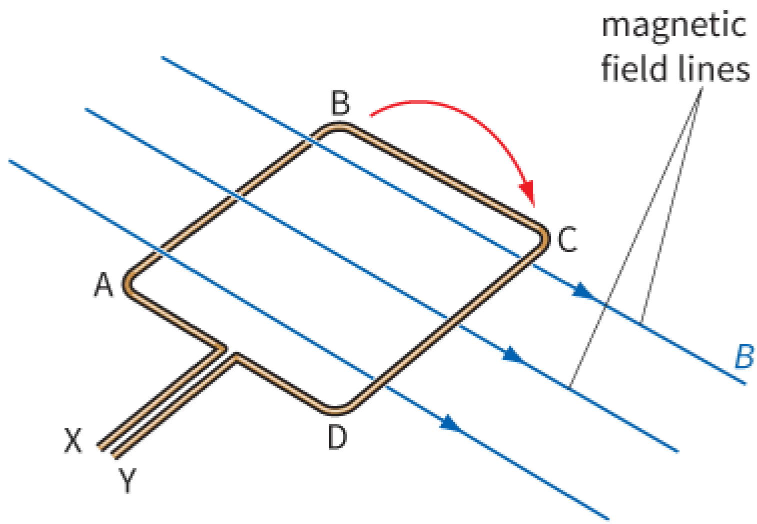


Figure 26.11: A coil rotating in a uniform magnetic field.

Magnetic flux and magnetic flux linkage

So far, in this chapter we have looked at the ideas of electromagnetic induction in a very descriptive manner. Now we will see how to calculate the value of the induced e.m.f. and look at a general way of determining its direction.

In Chapter 24, we saw how magnetic flux density *B* is defined by the equation

$$B = \frac{F}{IL}$$

Now we can go on to define **magnetic flux** as a quantity. We picture magnetic flux density *B* as the number of magnetic field lines passing through a region per unit area. Similarly, we can picture magnetic flux as the total number of magnetic field lines passing through a cross-sectional area *A*. For a magnetic field normal to *A*, the magnetic flux Φ (Greek letter phi) must therefore be equal to the product of magnetic flux density and the area *A* (Figure 26.12a).

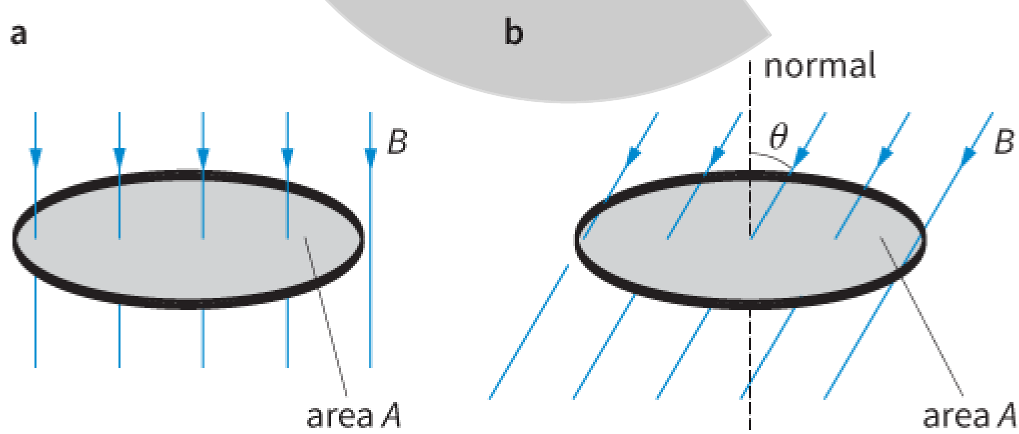


Figure 26.12: **a** The magnetic flux is equal to *BA* when the field is normal to the area. **b** The magnetic flux becomes when the field is at an angle θ to the normal of the area.

The magnetic flux Φ through cross-sectional area *A* is defined as:

$$\Phi = BA$$

where *B* is the component of the magnetic flux density perpendicular to the area.

KEY EQUATION

Magnetic flux:

$$\Phi = BA$$

KEY EQUATION

$$B \cos \theta$$

The component of the magnetic flux density *B* perpendicular to the plane of the cross-sectional area, where θ is the angle between the normal to the area and the magnetic field.

How can we calculate the magnetic flux when *B* is not perpendicular to *A*? You can easily see that when the field is parallel to the plane of the area, the magnetic flux through *A* is zero. To find the magnetic flux in general, we need to find the component of the magnetic flux density perpendicular to the cross-sectional area. Figure 28.12b shows a magnetic field at an angle θ to the normal. In this case:

$$\text{magnetic flux } \Phi = (B\cos\theta) \times A$$

or simply:

$$\text{magnetic flux } \Phi = BA\cos\theta$$

(Note that, when $\theta = 90°$, $\Phi = 0$ and when $\theta = 0°$ $\Phi = BA$)

For a coil with *N* turns, the **magnetic flux linkage** is defined as the product of the magnetic flux and the number of turns; that is:

$$\text{magnetic flux linkage} = N\Phi$$

or

$$\text{magnetic flux linkage} = BAN\cos\theta$$

The unit for magnetic flux, and magnetic flux linkage is the weber (Wb).

One weber (1 Wb) is the magnetic flux that passes perpendicularly through a cross-section of area 1 m^2 when the magnetic flux density is 1 T. 1 Wb = 1 Tm^2.

An e.m.f. is induced in a circuit whenever magnetic flux linking the circuit changes with respect to time. Since magnetic flux is equal to $BA\cos\theta$ there are three ways an e.m.f. can be induced:

- changing the magnetic flux density *B*
- changing the cross-sectional area *A* of the circuit
- changing the angle θ.

Now look at Worked example 1.

WORKED EXAMPLE

1 Figure 26.13 shows a solenoid with a cross-sectional area 0.10 m^2. It is linked by a magnetic field of flux density 2.0×10^{-3} T and has 250 turns.

Determine the magnetic flux and flux linkage for this solenoid.

Step 1 We have $B = 2.0 \times 10^{-3}$ T, $A = 0.10$ m^2, $\theta = 0°$ and $N = 250$ turns.

Hence we can calculate the flux Φ.

$$\begin{aligned} \Phi &= BA \\ &= 2.0 \times 10^{-3} \times 0.10 \\ &= 2.0 \times 10^{-4} \text{ Wb} \end{aligned}$$

Step 2 Now calculate the flux linkage.

$$\begin{aligned} \text{magnetic flux linkage} &= N\Phi \\ &= 2.0 \times 10^{-4} \times 250 \\ &= 5.0 \times 10^{-2} \text{ Wb} \end{aligned}$$

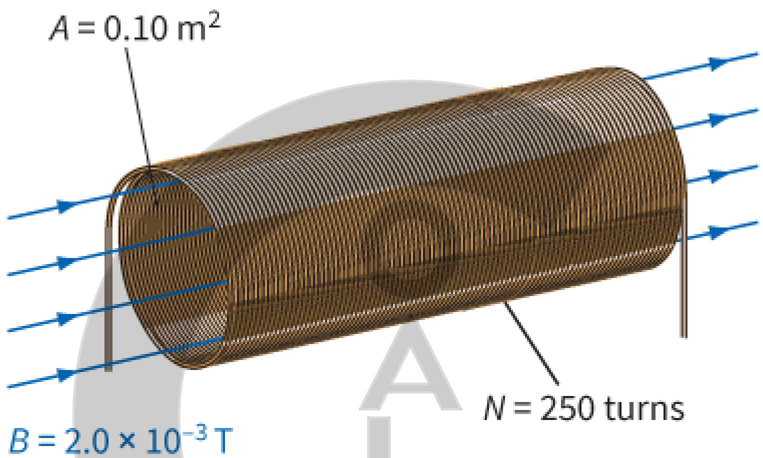


Figure 26.13: A solenoid in a magnetic field.

Questions

4 Use the idea of magnetic flux linkage to explain why, when a magnet is moved into a coil, the e.m.f. induced depends on the strength of the magnet and the speed at which it is moved.

5 In an experiment to investigate the factors that affect the magnitude of an induced e.m.f., a student moves a wire back and forth between two magnets, as shown in Figure 26.14. State why the e.m.f. generated in this way is almost zero.

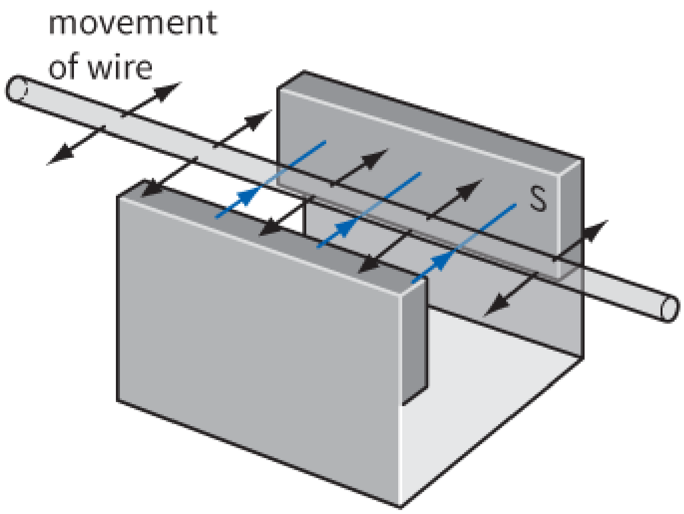


Figure 26.14: A wire is moved horizontally in a horizontal magnetic field. For Question 5.

6 In the type of generator found in a power station (Figure 26.15), a large electromagnet is made to rotate inside a fixed coil. An e.m.f. of 25 kV is induced; this is an alternating voltage of frequency 50 Hz.

- **a** State the factor that determines the frequency.
- **b** Suggest the factors that you think would affect the magnitude of the induced e.m.f.

Figure 26.15: For Question 6. The generators of this power station produce electricity at an induced e.m.f. of 25 kV.

7 At the surface of the north pole of a bar magnet, the magnetic field is uniform with flux density 0.15 T. The pole has dimensions 1.0 cm × 1.5 cm.

Calculate the magnetic flux at this pole.

8 A solenoid has diameter 5.0 cm, length 25 cm and 200 turns of wire (Figure 26.16). A current of 2.0 A creates a uniform magnetic field of flux density 2.0×10^{-5} T through the core of this solenoid.

a Calculate the magnetic flux linkage for this solenoid.

b The diameter of the solenoid is 5.0 ± 0.2 cm. Determine the absolute uncertainty in value calculated in part **a**. You may assume all the other quantities have negligible uncertainties.

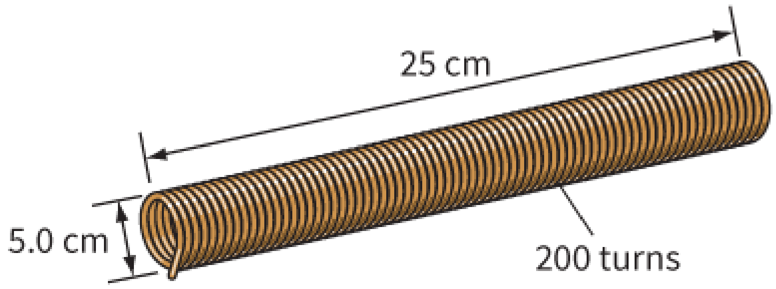


Figure 26.16: A solenoid. For Question 8.

9 A rectangular coil with 120 turns is placed at right angles to a magnetic field of flux density 1.2 T. The coil has dimensions 5.0 cm × 7.5 cm.

Calculate the magnetic flux linkage for this coil.

26.3 Faraday's law of electromagnetic induction

Earlier in this chapter, we saw that electromagnetic induction occurs when magnetic flux linking a circuit changes with time. We can now use **Faraday's law of electromagnetic induction** to determine the magnitude of the induced e.m.f. in a circuit:

The magnitude of the induced e.m.f. is directly proportional to the rate of change of magnetic flux linkage.

Remember that 'rate of change' in physics is equivalent to 'per unit time'. Therefore, we can write this mathematically as:

$$E \propto \left. \frac{\Delta(N\Phi)}{\Delta t} \right|$$

where $\Delta(N\Phi)$ |is the change in the magnetic flux linkage in a time Δt. When working in SI units, the constant of proportionality is equal to 1. (At this level of study, you do not need to worry about why this is the case.)

Therefore:

$$E = \left. \frac{\Delta(N\Phi)}{\Delta t} \right|$$

The equation is a mathematical statement of Faraday's law. Note that it allows us to calculate the **magnitude** of the induced e.m.f.; its **direction** is given by Lenz's law, which is discussed later in topic 26.3 Faraday's law of electromagnetic induction.

Now look at Worked examples 2 and 3.

WORKED EXAMPLES

2 A straight wire of length 0.20 m moves at a steady speed of 3.0 m s^{-1} at right angles to a magnetic field of flux density 0.10 T. Use Faraday's law to determine the magnitude of the induced e.m.f. across the ends of the wire.

Step 1 With a single conductor, $N = 1$. To determine the induced e.m.f. E, we need to find the rate of change of magnetic flux; in other words, the change in magnetic flux per unit time.

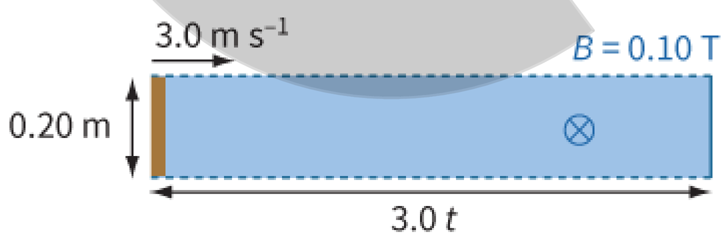


Figure 26.17: A moving wire cuts across the magnetic field.

Figure 26.17 shows that in a time t, the wire travels a distance $3.0t$.

Therefore:

change in magnetic flux = B × change in area

change in magnetic flux = 0.10 × (3.0t × 0.20) = 0.060t

Step 2 Use Faraday's law to determine the magnitude of the induced e.m.f.

E = rate of change of magnetic flux linkage

$$E = \frac{\Delta(N\Phi)}{\Delta t}$$

$\Delta\Phi = 0.060t$, $\Delta t = t$ and $N = 1$

$$E = \frac{0.060t}{t} = 0.060 \text{ V}$$

(The t cancels. You could have done this calculation for any time t, even 1.0 s. The results would still be the same.)

The magnitude of the induced e.m.f. across the ends of the wire is 60 mV.

3 This example illustrates one way in which the flux density of a magnetic field can be measured, shown in Figure 26.18. A search coil is a flat-coil with many turns of very thin insulated wire.

A search coil has 2500 turns and cross-sectional area 1.2 cm^2. It is placed between the poles of a magnet so that the magnetic flux passes perpendicularly through the plane of the coil. The magnetic field between the poles has flux density 0.50 T. The coil is pulled rapidly out of the field in a time of 0.10 s.

Calculate the magnitude of the average induced e.m.f. across the ends of the coil.

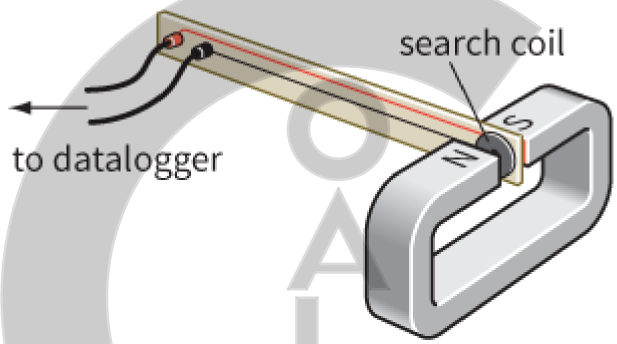


Figure 26.18: An e.m.f. is induced in the search coil when it is moved out of the field between the poles of the magnet. A search coil can be used to detect the presence of magnetic flux.

Step 1 Calculate the change in the magnetic flux linkage, $\Delta(N\Phi)$.

When the coil is pulled out from the field, the final flux linking the coil will be zero. The cross-sectional area A needs to be in m^2. Note: 1 $cm^2 = 10^{-4}$ m^2.

$\Delta(N\Phi)$ = Final $N\Phi$ − initial $N\Phi$

$\Delta(N\Phi) = 0 - [2500 \times 1.2 \times 10^{-4} \times 0.50] = -0.15$ Wb

Step 2 Now calculate the induced e.m.f. using Faraday's law of electromagnetic induction.

$\Delta(N\Phi) = -0.15$ Wb and $\Delta t = 0.10$ s

$$\text{magnitude of e.m.f } E = \frac{\Delta(N\Phi)}{\Delta t} = \frac{0.15}{0.10} = 1.5 \text{ V}$$

(The negative sign is not required; you only need to know the size of the e.m.f.)

Note that, in this example, we have assumed that the flux linking the coil falls steadily to zero during the time interval of 0.10 s. The answer is, therefore, an average value of the induced e.m.f.

Questions

10 A conductor of length L moves at a steady speed v at right angles to a uniform magnetic field of flux density B.

Show that the magnitude of the induced e.m.f. E across the ends of the conductor is given by the equation: $E = BLv$

(You can use Worked example 2 to guide you through Question 10.)

11 A wire of length 10 cm is moved through a distance of 2.0 cm in a direction at right angles to its length in the space between the poles of a magnet, and perpendicular to the magnetic field. The flux density is 1.5 T. If this takes 0.50 s, calculate the magnitude of the average induced e.m.f. across the ends of the wire.

12 Figure 26.19 shows a search coil with 2000 turns and cross-sectional area 1.2 cm^2. It is placed between the poles of a strong magnet. The magnetic field is perpendicular to the plane of the coil. The ends of the coil are connected to a voltmeter. The coil is then pulled out of the magnetic field, and the voltmeter records an average induced e.m.f. of 0.40 V over a time interval of 0.20 s.

Calculate the magnetic flux density between the poles of the magnet.

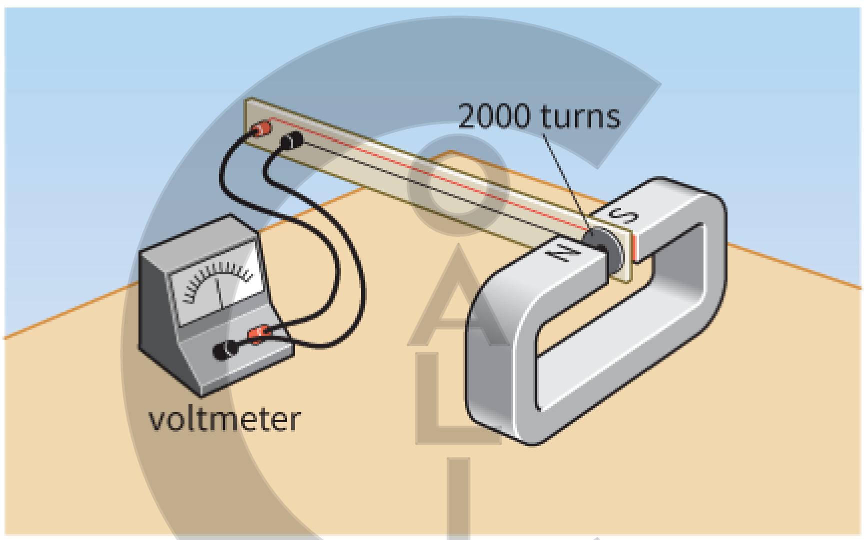


Figure 26.19: Using a search coil to determine the magnetic flux density of the field between the poles of this magnet.

26.4 Lenz's law

We use Faraday's law to calculate the magnitude of an induced e.m.f. Now, we can go on to think about the direction of the induced e.m.f. – in other words, which end of a wire or coil moving in a magnetic field becomes positive, and which becomes negative.

Fleming's right-hand rule gives the direction of a current caused by induced e.m.f. This is a particular case of a more general law, Lenz's law, which will be explained in this topic. First, we will see how the motor effect and the generator effect are related to each other.

The origin of electromagnetic induction

So far, we have not given an explanation of electromagnetic induction. You have seen, from the experiments at the beginning of this chapter, that it does occur, and you know the factors that affect it. But what is the origin of the current?

Figure 26.20 gives an explanation. A straight metal wire XY is being pushed downwards through a horizontal magnetic field of flux density *B*. Now, think about the free electrons in the wire. They are moving downwards, so they are, in effect, an electric current. Of course, because electrons are negatively charged, the conventional current is flowing upwards.

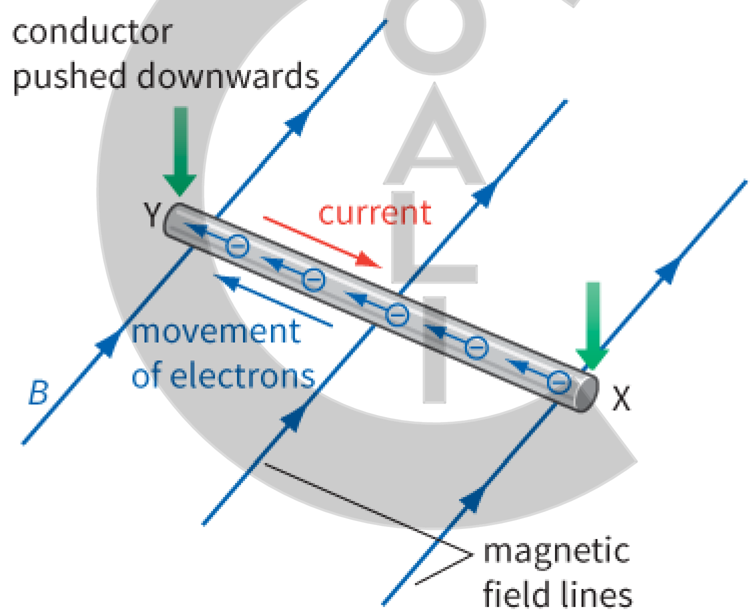


Figure 26.20: Showing the direction of the current caused by the induced e.m.f.

We now have a current flowing across a magnetic field, and the motor effect will, therefore, come into play. Each electron experiences a force of magnitude *Bev*. Using Fleming's left-hand rule, we can find the direction of the force on the electrons. The diagram shows that the electrons will be pushed in the direction from X to Y. So a current has been induced to flow in the wire; the direction of the conventional current is from Y to X.

Now, we can check that Fleming's right-hand rule gives the correct directions for motion, field and current, which indeed it does.

So, to summarise, there is a current caused by the induced e.m.f. current because the electrons are pushed by the motor effect. Electromagnetic induction is simply a consequence of the motor effect.

In Figure 26.20, electrons are found to accumulate at Y. This end of the wire is thus the negative end of the e.m.f. and X is positive. If the wire was connected to an external circuit, electrons would flow out of Y, round the circuit, and back into X. Figure 26.21 shows how the moving wire is equivalent to a cell (or any other source of e.m.f.).

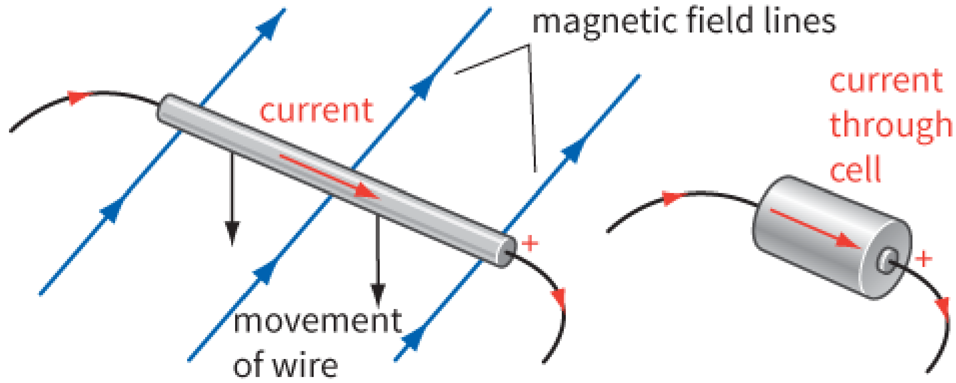


Figure 26.21: A moving wire in a magnetic field is a source of e.m.f. – equivalent to a cell.

Forces and movement

Electromagnetic induction is how we generate most of our electricity. We turn a coil in a magnetic field, and the mechanical energy we put in is transferred to electrical energy. By thinking about these energy transfers, we can deduce the direction of the current.

Figure 26.22 shows one of the experiments from earlier in this chapter. The north pole of a magnet is being pushed towards a coil of wire. There is a current in the coil, but what is its direction? The diagram shows the two possibilities.

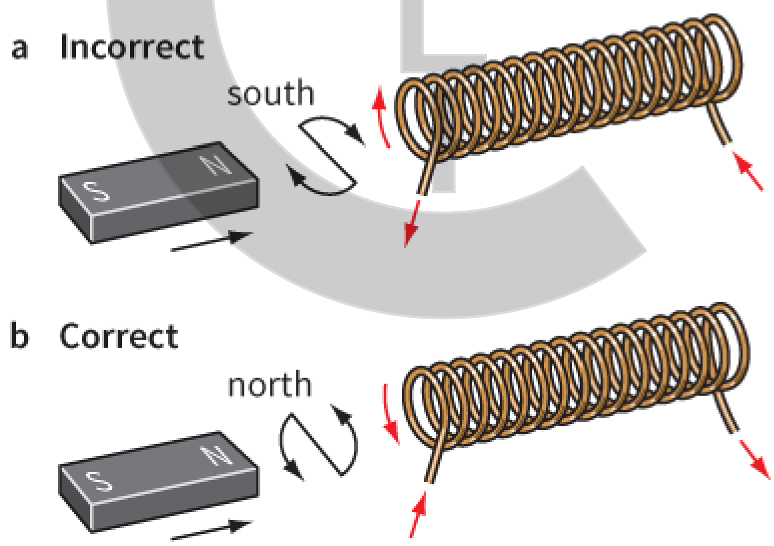


Figure 26.22: Moving a magnet towards a coil: the direction of the current caused by the induced e.m.f. is as shown in **b**, not **a**.

The current in the coil makes it into an electromagnet. One end becomes the north pole, the other the south pole. In Figure 26.22a, if the current is in this direction, the coil end nearest the approaching north pole of the magnet would be a south pole. These poles will attract one another, and you could gently let go of the magnet and it would be dragged into the coil. The magnet would accelerate into the coil, the current caused by induced e.m.f. would increase further, and the force of attraction between the two would also increase.

In this situation, we would be putting no (or very little at the start) energy into the system, but the magnet would be gaining kinetic energy, and the current would be gaining electrical energy. A nice trick if you could do it, but this would violate the principle of conservation of energy!

Figure 26.22b shows the correct situation. As the north pole of the magnet is pushed towards the coil, the current caused by the induced e.m.f. makes the end of the coil nearest the magnet become a north pole. The two poles repel one another, and you have to do work to push the magnet into the coil. The energy transferred by your work is transferred to electrical energy of the current. The principle of energy conservation is not violated in this situation.

Figure 26.23 shows how we can apply the same reasoning to a straight wire being moved in a downward direction through a magnetic field. There will be a current caused by induced e.m.f. in the wire, but in which direction? Since this is a case of a current across a magnetic field, a force will act on it (the motor effect), and we can use Fleming's left-hand rule to deduce its direction.

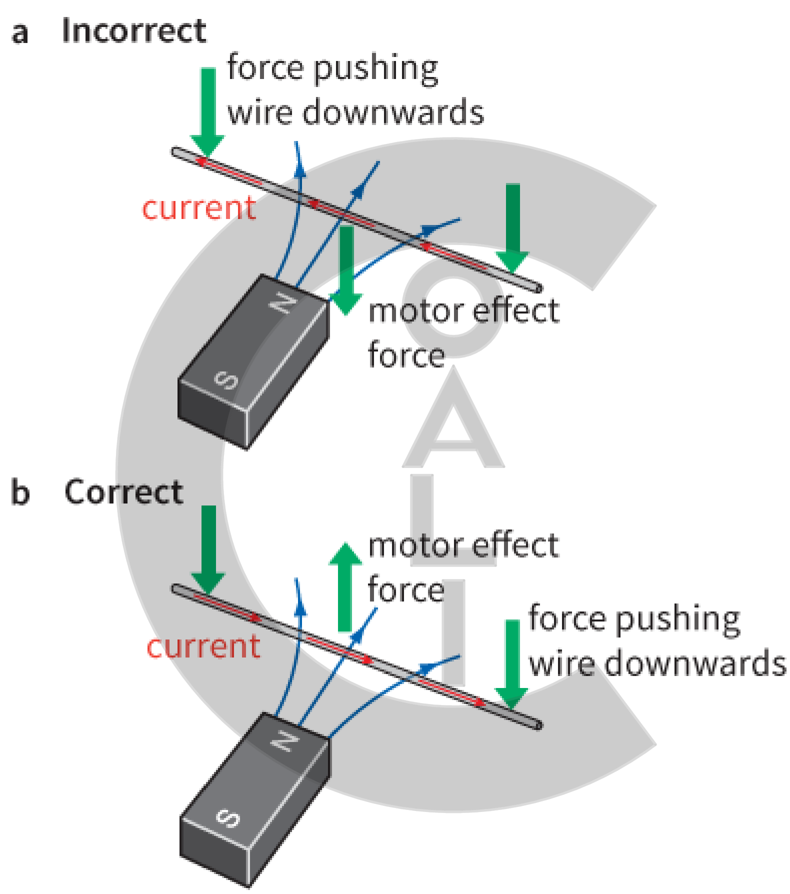


Figure 26.23: Moving a wire through a magnetic field: the direction of the current is as shown in **b**, not **a**.

First, we will consider what happens if the current caused by the induced e.m.f. is in the wrong direction. This is shown in Figure 26.23a. The left-hand rule shows that the force that results would be downward–in the direction in which we are trying to move the wire. The wire would thus be accelerated, the current would increase and again we would be getting both kinetic and electrical energy for no energy input.

The current must be as shown in Figure 26.23b. The force that acts on it due to the motor effect pushes against you as you try to move the wire through the field. You have to do work to move the wire, and hence to generate electrical energy. Once again, the principle of energy conservation is not violated.

Questions

13 Use the ideas in the previous topic to explain what happens if **a** you stop pushing the magnet towards the coil shown in Figure 26.22, and **b** you pull the magnet away from the coil.

14 Draw a diagram to show the directions of the current caused by induced e.m.f. and of the opposing force if you now try to move the wire shown in Figure 26.23 upwards through the magnetic field.

A general law for induced e.m.f.

Lenz's law summarises this general principle of energy conservation. The direction of a current caused by induced e.m.f. or e.m.f is such that it always produces a force that opposes the motion that is being used to produce it. If the direction of the e.m.f were opposite to this, we would be getting energy for nothing.

Here is a statement of Lenz's law:

Any induced e.m.f. will be established in a direction so as to produce effects that oppose the change that is producing it.

This law can be shown to be correct in any experimental situation. For example, in Figure 26.3, a sensitive ammeter connected in the circuit shows the direction of the current as the magnet is moved in and out. If a battery is later connected to the coil to make a larger and constant current in the same direction, a compass will show what the poles are at the end of the solenoid. If a north pole is moved into the solenoid, then the solenoid itself will have a north pole at that end. If a north pole is moved out of the solenoid, then the solenoid will have a south pole at that end.

Faraday's law of electromagnetic induction, and Lenz's law, may be summarised using the equation:

$$E = -\frac{\Delta(N\Phi)}{\Delta t}$$

where E is the magnitude of the induced e.m.f. and the minus sign indicates that this induced e.m.f. causes effects to oppose the change producing it.

The minus sign is there because of Lenz's law – it is necessary to emphasise the principle of conservation of energy.

KEY EQUATION

$$E = -\frac{\Delta(N\Phi)}{\Delta t}$$

Induced electromagnetic force.

Questions

15 A bar magnet is dropped vertically downwards through a long solenoid, which is connected to an oscilloscope (Figure 26.24). The oscilloscope trace shows how the e.m.f. induced in the coil varies with time as the magnet accelerates downwards.

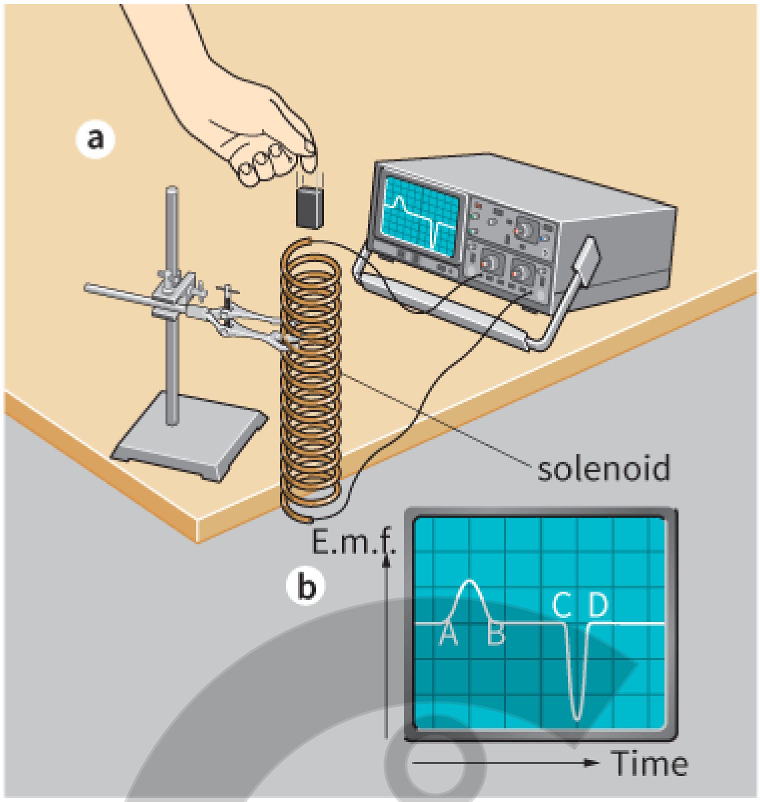


Figure 26.24: a A bar magnet falls through a long solenoid. **b** The oscilloscope trace shows how the induced e.m.f. varies with time.

a Explain why an e.m.f. is induced in the coil as the magnet enters it (section AB of the trace).

b Explain why no e.m.f. is induced while the magnet is entirely inside the coil (section BC).

c Explain why section CD shows a negative trace, why the peak e.m.f. is greater over this section, and why CD represents a shorter time interval than AB.

16 You can turn a bicycle dynamo by hand and cause the lamps to light up. Use the idea of Lenz's law to explain why it is easier to turn the dynamo when the lamps are switched off than when they are on.

26.5 Everyday examples of electromagnetic induction

An induced e.m.f. can be generated in a variety of ways, but can be explained in terms of Faraday's and Lenz's laws. An e.m.f. will be induced whenever there is a rate of change of magnetic flux linkage for a circuit or device. In this topic, we will examine the physics behind two devices – a generator and a transformer.

Generators

We can generate electricity by spinning a coil in a magnetic field. This is equivalent to using an electric motor backwards. Figure 26.25 shows such a coil in three different orientations as it spins.

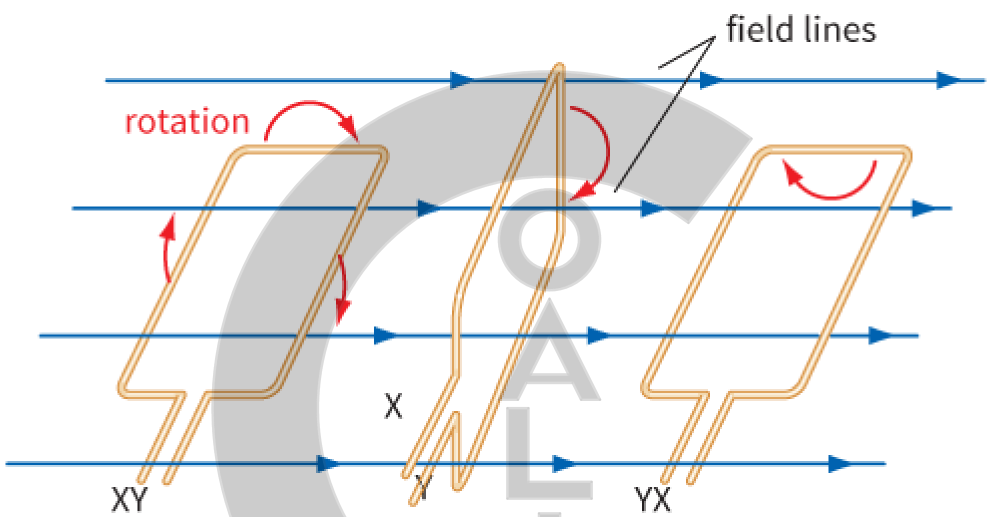


Figure 26.25: A coil rotating in a magnetic field.

Notice that the rate of change of magnetic flux linkage is maximum when the coil is moving through the **horizontal** position. In this position, we get a large induced e.m.f. As the coil moves through the **vertical** position, the rate of change of magnetic flux is zero and the induced e.m.f. is zero.

Figure 26.26 shows how the magnetic flux linkage varies with time for a rotating coil.

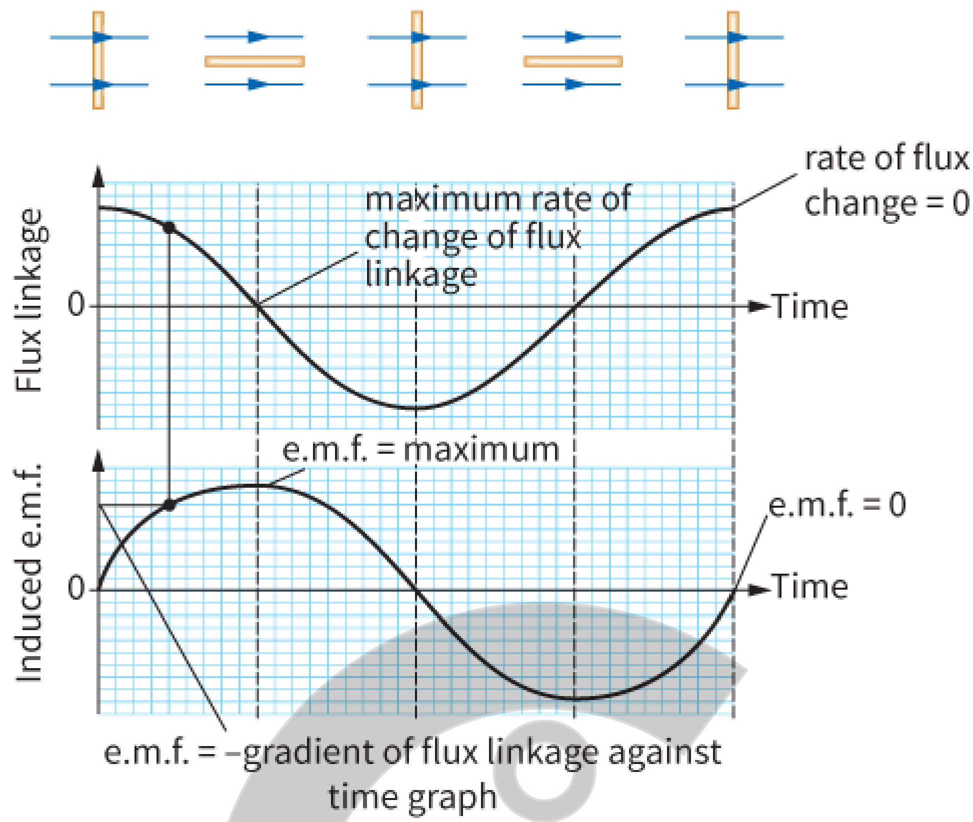


Figure 26.26: The magnetic flux linking a rotating coil as it changes. This gives rise to an alternating e.m.f. The orientation of the coil is shown above the graphs.

According to Faraday's and Lenz's laws, the induced e.m.f. is equal to minus the **gradient** of the flux linkage against time graph:

$$E = -\frac{\Delta(N\Phi)}{\Delta t}$$

When the flux linking the coil is:

- maximum, the rate of change of flux linkage is zero and hence the induced e.m.f. is zero
- zero, the rate of change of flux linkage is maximum (the graph is steepest) and hence the induced e.m.f. is also maximum.

Hence, for a coil like this, we get a varying e.m.f. – this is how alternating current is generated. In practice, it is simpler to keep the large coil fixed and spin an electromagnet inside it (Figure 26.27). A bicycle generator is similar, but in this case a permanent magnet is made to spin inside a fixed coil.

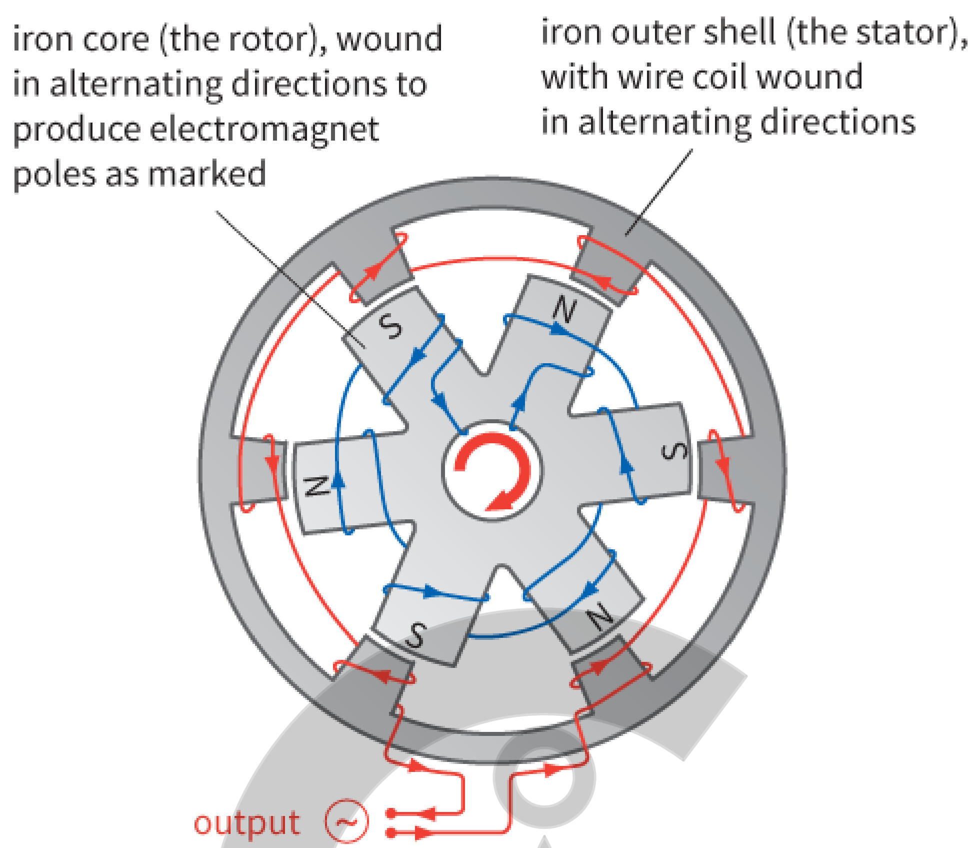


Figure 26.27: In a generator, an electromagnet rotates inside a coil.

Transformers

You may have studied transformers before your study of this course.

A simple transformer has a primary coil and a secondary coil, both wrapped around a soft iron core (ring). An alternating current is supplied to the primary coil. This produces a varying magnetic flux in the soft iron core (see Figure 26.28). The secondary coil is linked by the same changing magnetic flux in the soft iron core, so an e.m.f. is induced at the ends of this coil. According to Faraday's law, you can increase the induced e.m.f. at the secondary coil by increasing the number of turns of the secondary coil. Having fewer turns on the secondary will have the reverse effect.

Transformers are used to transport electrical energy using overhead cables.

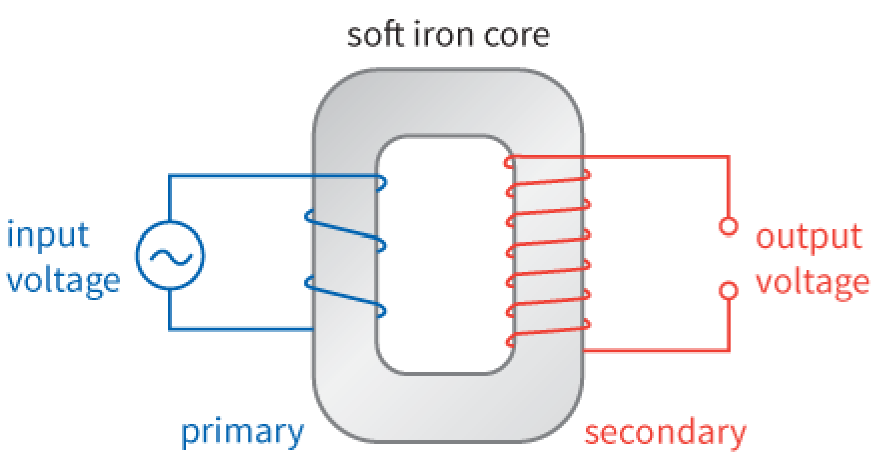


Figure 26.28: Faraday's law can be used to explain the output from a transformer.

Questions

17 Figure 26.29 represents a coil of wire ABCD being rotated in a uniform horizontal magnetic field. Copy and complete the diagram to show the direction of the current caused by induced e.m.f. in the coil, and the directions of the forces on sides AB and CD that oppose the rotation of the coil.

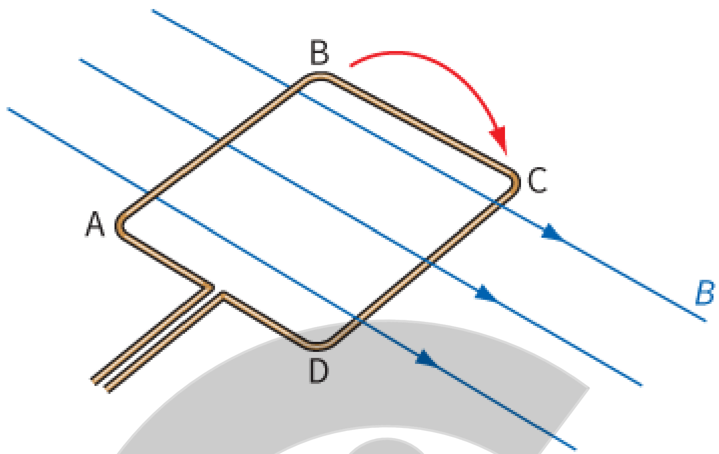


Figure 26.29: A coil rotating in a magnetic field.

18 Does a bicycle generator (Figure 26.7) generate alternating or direct current? Justify your answer.

19 The peak e.m.f. induced in a rotating coil in a magnetic field depends on four factors: magnetic flux density *B*, area of the coil *A*, number of turns *N* and frequency *f* of rotation. Use Faraday's law to explain why the magnitude of the induced e.m.f. must be proportional to each of these quantities.

20 Explain why, if a transformer is connected to a steady (d.c.) supply, no e.m.f. is induced across the secondary coil.

REFLECTION

Without looking at your textbook, summarise the factors that affect the e.m.f. induced in a circuit. Compare your summary with a fellow learner.

Make a deck of cards with all the physical quantities in this chapter. Do the same for the units for each quantity. Ask a fellow learner to match the quantities with their units.

How can you better support and encourage your classmates on future activities and questions?

SUMMARY

In a magnetic field of magnetic flux density B, the magnetic flux Φ passing through a cross-sectional area A is given by:

$$\Phi = BA$$

Magnetic flux linkage = N × magnetic flux = $N\Phi$

Magnetic flux and magnetic flux linkage are both measured in webers (Wb). 1 Wb = 1 T m^2.

An e.m.f. is induced in a circuit whenever there is a change in the magnetic flux linkage.

Faraday's law states that the magnitude of the induced e.m.f. is equal to the rate of change of magnetic flux linkage:

$$E = \frac{\Delta(N\Phi)}{\Delta t}$$

Lenz's law states that the induced current or e.m.f. is in a direction so as to produce effects that oppose the change that is producing it.

The equation for both Faraday's and Lenz's laws may be written as:

$$E = -\frac{\Delta(N\Phi)}{\Delta t}$$

EXAM-STYLE QUESTIONS

1 Which of the following units is **not** correct for magnetic flux? **[1]**

A $kg\ m^2\ s^{-2}\ A^{-1}$

B T

C $T\ m^2$

D Wb

2 A student thinks that electrical current passes through the core in a transformer to the secondary coil. Describe how you might demonstrate that this is not true and explain how an electrical current is actually induced in the secondary coil. Use Faraday's law in your explanation. **[3]**

3 A square coil of 100 turns of wire has sides of length 5.0 cm. It is placed in a magnetic field of flux density 20 mT, so that the flux is perpendicular to the plane of the coil.

a Calculate the flux through the coil. **[2]**

b The coil is now pulled from the magnetic field in a time of 0.10 s. Calculate the average e.m.f. induced in it. **[3]**

[Total: 5]

4 An aircraft of wingspan 40 m flies horizontally at a speed of 300 ± 10 m s^{-1} in a region where the vertical component of the Earth's magnetic field is 5.0×10^{-5} T.

Calculate the magnitude of the e.m.f. induced between the aircraft's wingtips; in your answer, include the absolute uncertainty. **[5]**

5 Figure 28.26 shows the magnetic flux linkage and induced e.m.f. as a coil rotates. Explain why the induced e.m.f. is a maximum when there is no flux linkage and the induced e.m.f. is zero when the flux linkage is a maximum. **[4]**

6 **a** Explain what is meant by a magnetic flux linkage of 1 Wb. **[2]**

b This is a graph of magnetic flux density through a 240 turn coil with a cross-sectional area 1.2×10^{-4} m^2 against time.

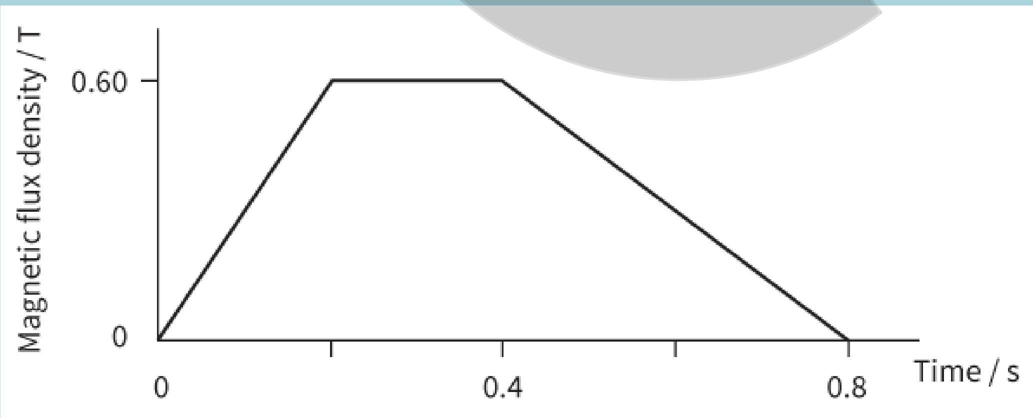


Figure 26.30

i Determine the maximum rate of change of flux in the coil. **[2]**

ii Determine the maximum magnitude of the induced e.m.f. in the coil. **[2]**

iii Sketch a diagram to show the induced e.m.f. varies with time. Mark values on both the e.m.f. and time axes. **[2]**

[Total: 8]

7 This diagram shows a square coil about to enter a region of uniform magnetic field of magnetic flux density 0.30 T. The magnetic field is at right angles to the plane of the coil. The coil has 150 turns and each side is 2.0 cm in length. The coil moves at a constant speed of 0.50 m s^{-1}.

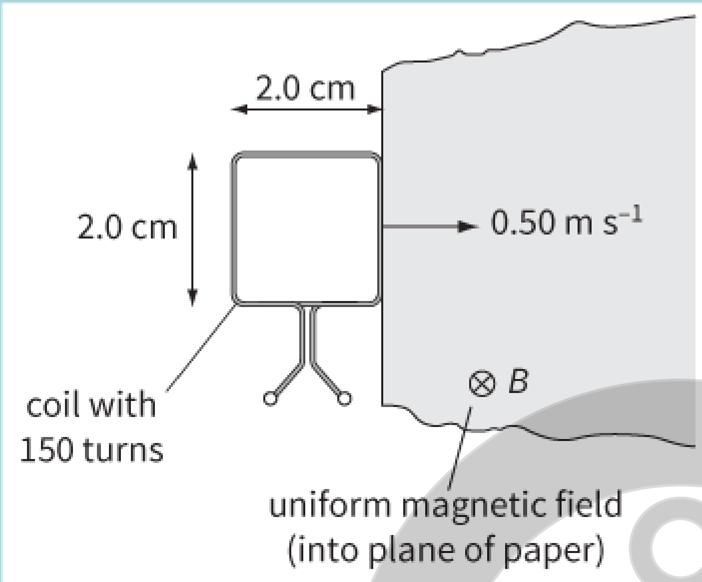


Figure 26.31

a i Calculate the time taken for the coil to completely enter the region of magnetic field. **[1]**

ii Determine the magnetic flux linkage through the coil when it is all within the region of magnetic field. **[2]**

b Explain why the magnitude of the induced e.m.f. is constant while the coil is entering the magnetic field. **[1]**

c Use your answer to part a to determine the induced e.m.f. across the ends of the coil. **[4]**

d Explain the induced e.m.f. across the ends of the coil when it is completely within the magnetic field. **[2]**

e Sketch a graph to show the variation of the induced e.m.f. with time from the instant that the coil enters the magnetic field. Your time axis should go from 0 to 0.08 s. **[2]**

[Total: 12]

8 a State Faraday's law of electromagnetic induction. **[2]**

b A circular coil of diameter 200 mm has 600 turns is shown. It is placed with its plane perpendicular to a horizontal magnetic field of uniform flux density 50 mT. The coil is then rotated through 90° about a vertical axis in a time of 120 ms.

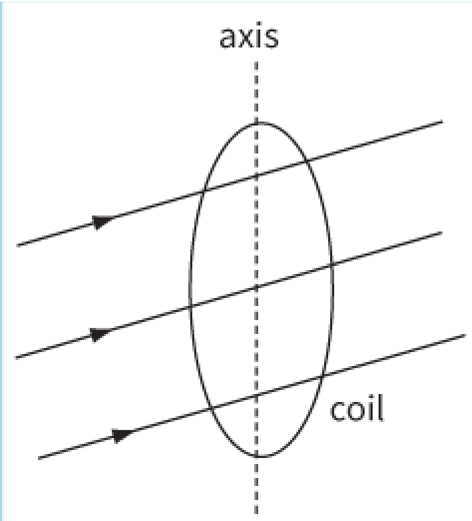


Figure 26.32

Calculate:

i the magnetic flux passing through the coil before the rotation **[2]**

ii the change of magnetic flux linkage produced by the rotation **[2]**

iii the average magnitude of the induced e.m.f. in the coil during the rotation. **[2]**

[Total: 8]

9 A bicycle wheel is mounted vertically on a metal axle in a horizontal magnetic field, as shown in the diagram. Sliding connections are made to the metal edge of the wheel and to the metal axle.

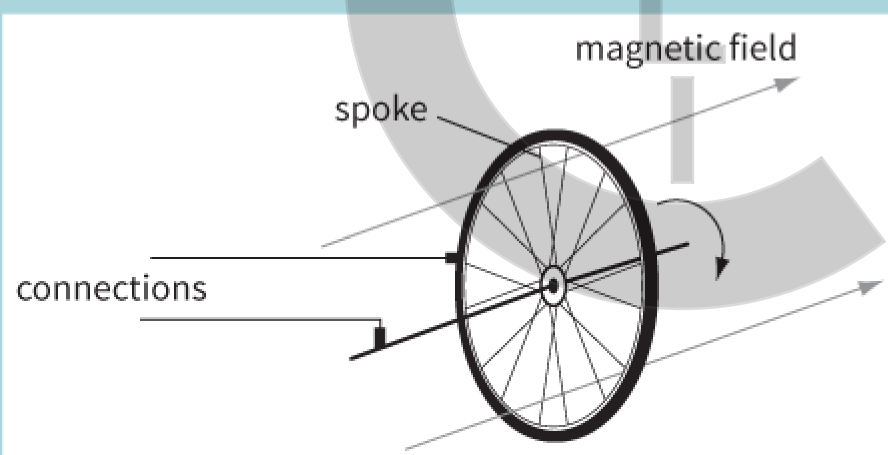


Figure 26.33

a **i** Explain why an e.m.f. is induced when the wheel rotates. **[2]**

ii State and explain two ways in which this e.m.f. can be increased. **[2]**

b The wheel rotates five times per second and has a radius of 15 cm. The magnetic flux density may be assumed to be uniform and of value 5.0×10^{-3} T.

Calculate:

i the area swept out each second by one spoke **[2]**

ii the induced e.m.f. between the contacts. **[2]**

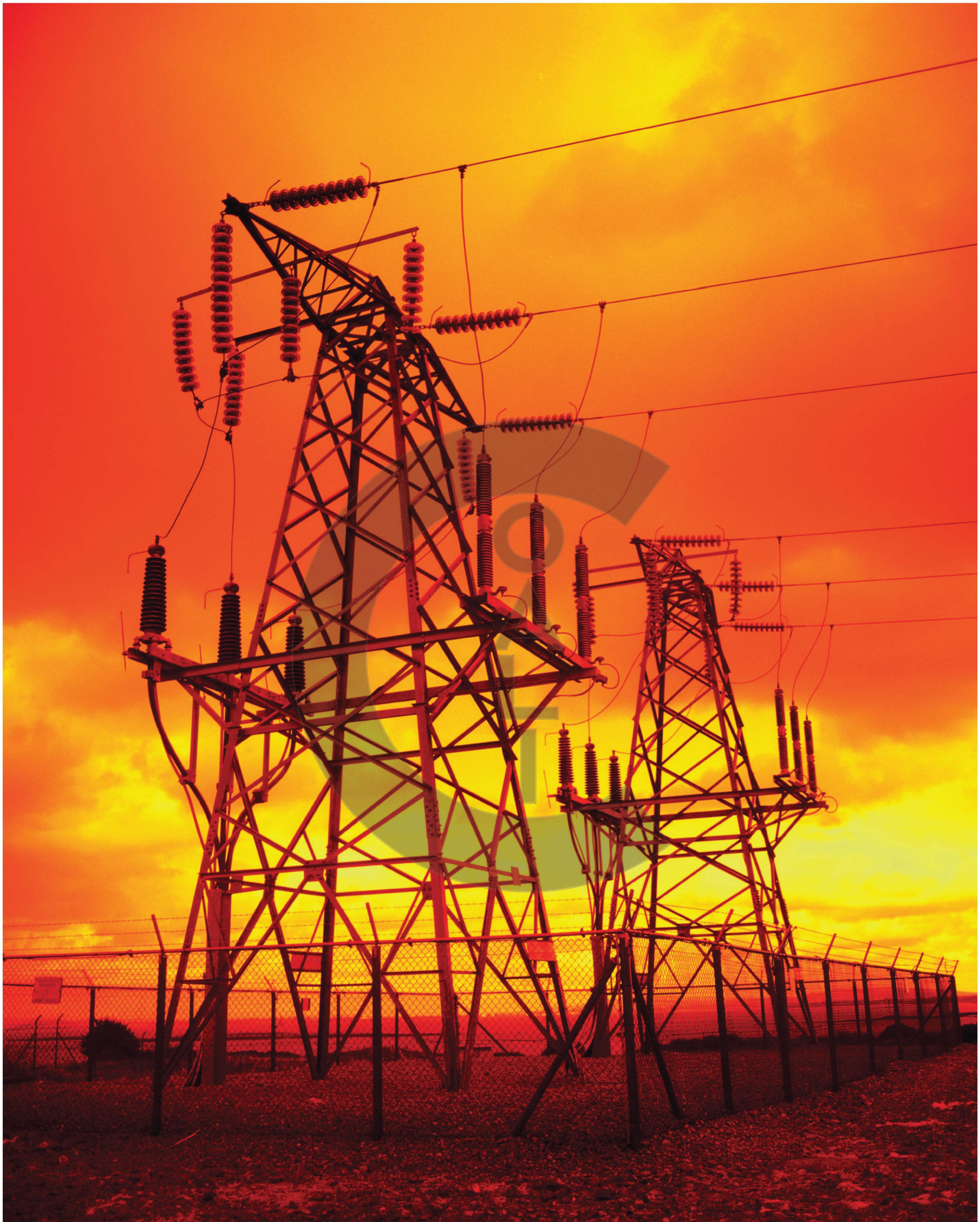
[Total: 8]

COALI

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
define magnetic flux Φ	26.2			
recall and use: $\Phi = BA$	26.2			
understand and use the concept of magnetic flux linkage	26.2			
understand and explain experiments that produce an e.m.f. induced in circuits	26.2			
recall and use Faraday's and Lenz's laws of electromagnetic induction.	26.4, 26.5			



〉Chapter 27

Alternating currents

LEARNING INTENTIONS

In this chapter you will learn how to:

- understand and use the terms period, frequency and peak value as applied to an alternating current or voltage
- use equations of the form $x = x_0 \sin \omega t$ representing a sinusoidally alternating current or voltage
- recall and use the fact that the mean power in a resistive load is half the maximum power for a sinusoidal alternating current
- distinguish between root-mean-square (r.m.s.) and peak values and recall and use $I_{r.m.s.} = \frac{I_0}{\sqrt{2}}$ and $V_{r.m.s.} = \frac{V_0}{\sqrt{2}}$ for a sinusoidal alternating current
- distinguish graphically between half-wave and full-wave rectification
- explain the use of a single diode for the half-wave rectification of an alternating current
- explain the use of four diodes (bridge rectifier) for the full-wave rectification of an alternating current
- analyse the effect of a single capacitor in smoothing, including the effect of the value of capacitance and the load resistance.

BEFORE YOU START

- In pairs, try to recall and explain the relationship for power dissipation in terms of current, potential difference and resistance from Chapter 8.
- The physics of alternating currents has similarities with simple harmonic motion (see Chapter 18). Discuss what you remember about period, frequency and angular frequency.
- Write down what you know about the behaviour of diodes in circuits. What's the most important property of a diode?
- Discuss the discharge of a capacitor through a resistor. Can you remember the factors that affect the time constant of a circuit?

DESCRIBING ALTERNATING CURRENT

In many countries, mains electricity is a supply of alternating current (a.c.). The first mains electricity supplies were developed towards the end of the 19th century; at that time, a great number of different voltages and frequencies were used in different places. In some places, the supply was direct current (d.c.). Nowadays, this has been standardised across much of the world, with standard voltages of 110 V or 230 V (or similar), and frequencies of 50 Hz or 60 Hz.

Mains electricity is transported along many kilometres of high-voltage power lines (cables). Transformers are used for stepping-up and stepping-down alternating voltages between the power stations and the consumers (Figure 27.1). From your prior knowledge of transformers and transmission of electrical energy, can you remember why it is necessary for power lines to use high voltage?

Figure 27.1: This engineer is working on a transformer used for increasing (stepping-up) the size of the alternating voltage to help with the transportation of electrical energy.

27.1 Sinusoidal current

An alternating current can be represented by a graph such as that shown in Figure 27.2. This shows that the current varies regularly. During half of the cycle, the current is positive, and in the other half it is negative. This means that the direction of the current reverses every half cycle. Whenever you use a mains appliance, the charges (free electrons) within the wire and appliance flow backwards and forwards. At any instant in time, the current has a particular magnitude and direction given by the graph.

The graph has the same shape as the graphs used to represent simple harmonic motion (s.h.m.) (see Chapter 18), and it can be interpreted in the same way. In a wire with a.c., the free electrons within the wire move back and forth with s.h.m. The variation of the current with time is a sine curve, so it is described as **sinusoidal**. (In principle, any current whose direction changes between positive and negative can be described as **alternating**, but we will only be concerned with those that have a regular, sinusoidal pattern.)

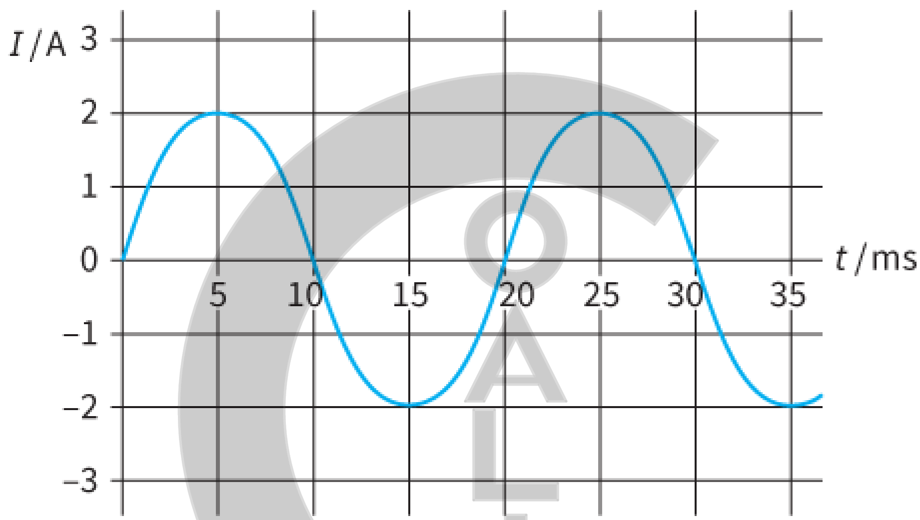


Figure 27.2: A graph to represent a sinusoidal alternating current.

An equation for a.c.

As well as drawing a graph, we can write an equation to represent alternating current. This equation gives us the value of the current *I* at any time *t*:

$$I = I_0 \sin \omega t$$

where *I* is the current at time *t*, I_0 is the **peak value** of the alternating current and ω is the angular frequency of the supply, measured in rad s^{-1} (radians per second). The peak value is the maximum magnitude of the current. It's very much like the 'amplitude' of the alternating current, except the unit is that of current.

This is related to the frequency *f* in the same way as for s.h.m.:

$$\omega = 2\pi f$$

and the frequency and period are related by:

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

KEY EQUATION

Alternating current:

$$I = I_0 \sin \omega t$$

Remember that your calculator must be in the radian mode when using this equation.

Questions

1 The following questions relate to the graph in Figure 27.2.

 a State the value of the current I and its direction when time t = 5 ms.

 b Determine the time the current next has the same value, but negative.

 c State the time T for one complete cycle (the period of the a.c).

 d Determine the frequency of this alternating current.

2 The following questions relate to the graph in Figure 27.2.

 a Determine the values of I_0 and ω.

 b Write an equation to represent this alternating current.

3 An alternating current, measured in amperes (A), is represented by the equation: $I = 5.0 \sin (120\pi t)$

 a Determine the values of I_0, ω, f and T.

 b Sketch a graph to represent the current.

27.2 Alternating voltages

Alternating current is produced in power stations by large generators like those shown in Figure 27.3.

Figure 27.3: Generators in the generating hall of a large power station.

As you have already seen in Chapter 26, a generator consists of a coil rotating in a magnetic field. An e.m.f. is induced in the coil according to Faraday's and Lenz's laws of electromagnetic induction.

This e.m.f. V varies sinusoidally, and so we can write an equation to represent it that has the same form as the equation for alternating current:

$$V = V_0 \sin \omega t$$

where V_0 is the **peak value** of the voltage. We can also represent this graphically, as shown in Figure 27.4.

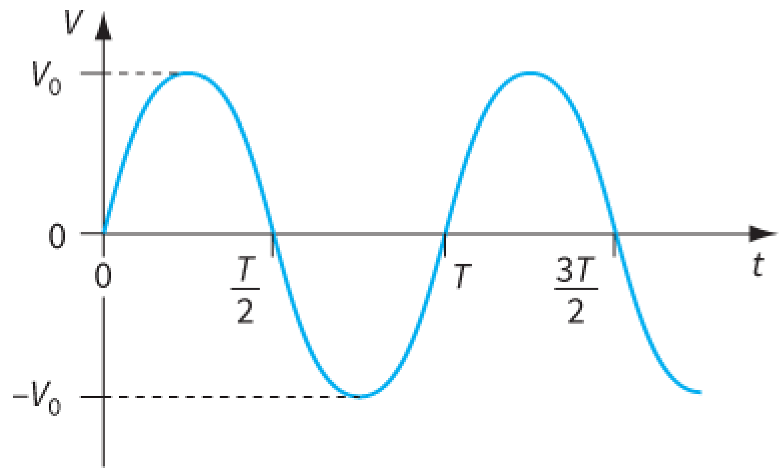


Figure 27.4: An alternating voltage.

Question

4 An alternating voltage V, in volt (V), is represented by the equation:

$V = 300 \sin(100\pi t)$

- **a** Determine the values of V_0, ω and f for this alternating voltage.
- **b** Calculate V when $t = 0.002$ s. (Remember that $100\pi t$ is in radians when you calculate this.)
- **c** Sketch a graph to show **two** complete cycles of this voltage.

Measuring frequency and voltage

An oscilloscope can be used to measure the frequency and voltage of an alternating current. Practical Activity 27.1 explains how to do this. There are two types of oscilloscope. The traditional cathode-ray oscilloscope (CRO) uses an electron beam. The alternative is a digital oscilloscope, which is likely to be much more compact and which can store data and display the traces later.

PRACTICAL ACTIVITY 27.1 MEASUREMENTS USING AN OSCILLOSCOPE

A CRO is an electron beam tube, as shown in Figure 25.4, but with an extra set of parallel plates to produce a horizontal electric field at right angles to the beam (Figure 27.5).

The principles of a cathode-ray oscilloscope (CRO)

The signal into the CRO is a repetitively varying voltage. This is applied to the *y*-input, which deflects the beam up and down using the parallel plates Y_1 and Y_2 shown in Figure 27.5. The time-base produces a p.d. across the other set of parallel plates X_1 and X_2 to move the beam from left to right across the screen.

When the beam hits the screen of the CRO, it produces a small spot of light. If you look at the screen and slow the movement down, you can see the spot move from left to right, while the applied signal moves the spot up and down. When the spot reaches the right side of the screen, it flies back very quickly and waits for the next cycle of the signal to start before moving to the right once again. In this way, the signal is displayed as a stationary trace on the screen. There may be many controls on a CRO, even more than those shown on the CRO illustrated in Figure 27.6.

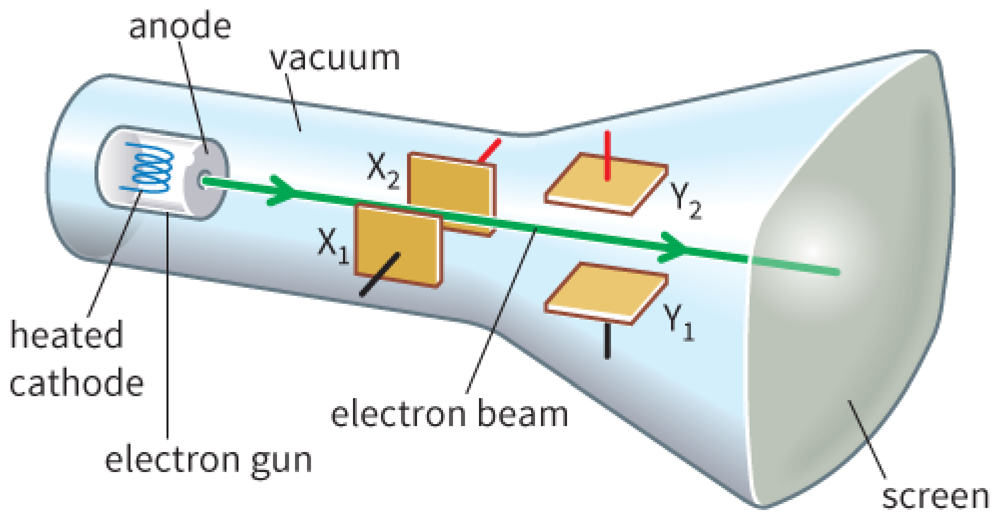


Figure 27.5: The construction of a cathode-ray oscilloscope. Cathode rays (beams of electrons) are produced in the electron gun and then deflected by electric fields before they strike the screen.

The controls

The X-shift and the Y-shift controls move the whole trace in the *x*-direction and the *y*-direction, respectively. The two controls that you must know about are the time-base and the Y-gain, or Y-sensitivity.

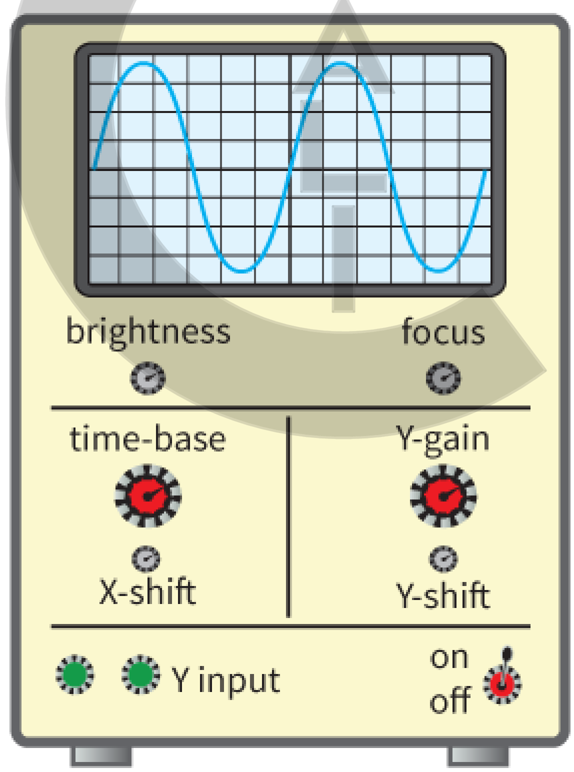


Figure 27.6: The controls of a typical CRO.

You can see in Figure 27.6 that the time-base control has units marked alongside. Let us suppose that this reads 5 ms/cm, although it might be 5 ms/division. This shows that 1 cm (or 1 division) on the *x*-axis represents 5 ms. Varying the time-base control alters the speed with which the spot moves across the screen.

If the time-base is changed to 1 ms/cm, then the spot moves faster and each centimetre represents a smaller time.

The Y-gain control has a unit marked in volts/cm, or sometimes volts/division. If the actual marking is 5 V/cm, then each centimetre on the *y*-axis represents 5 V in the applied signal.

It is important to remember that on the CRO screen, the *x*-axis represents **time** and the *y*-axis represents **voltage**.

Determining frequency and amplitude (peak value of voltage)

If you look at the CRO trace shown in Figure 27.7, you can see that the amplitude of the waveform, or the peak value of the voltage, is equivalent to 2 cm and the period of the trace is equivalent to 4 cm.

If the Y-gain or Y-sensitivity setting is 2 V/cm, then the peak voltage is 2 × 2 = 4 V. If the time-base setting is 5 ms/cm, then the period is 4 × 5 = 20 ms.

In the example:

$$\text{frequency} = \frac{1}{\text{period}} = \frac{1}{0.02} = 50 \text{ Hz}$$

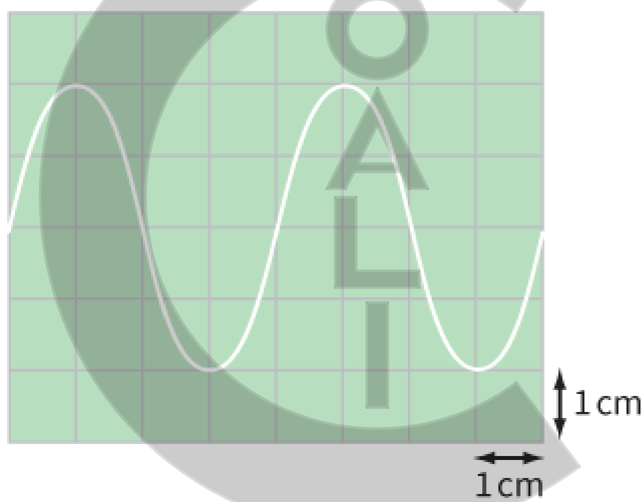


Figure 27.7: A typical trace on the screen of a CRO.

Questions

5 The Y-sensitivity and time-base settings are 5 V/cm and 10 ms/cm. The trace seen on the CRO screen is the one shown in Figure 27.7.

Determine the amplitude, period and frequency of the signal applied to the Y-input of the CRO.

6 Sketch the CRO trace for a sinusoidal voltage of frequency 100 Hz and amplitude 10 V, when the time-base is 10 ms/cm and the Y-sensitivity is 10 V/cm.

27.3 Power and alternating current

We use mains electricity to supply us with energy. If the current and voltage are varying all the time, does this mean that the power is varying all the time too? The answer to this is yes. You may have noticed that some fluorescent lamps flicker continuously, especially if you observe them out of the corner of your eye or when you move your head quickly from one side to the other. A tungsten filament lamp would flicker too, but the frequency of the mains has been chosen so that the filament does not have time to cool down noticeably between peaks in the supply.

Root-mean-square (r.m.s.) values

There is a mathematical relationship between the peak value V_0 of the alternating voltage and a direct voltage that delivers the same average electrical power. The direct voltage is about 70% of V_0. (You might have expected it to be about half, but it is more than this, because of the shape of the sine graph.) This steady direct voltage is known as the **root-mean-square (r.m.s.) value** of the alternating voltage. In the same way, we can think of the root-mean-square value of an alternating current, $I_{r.m.s.}$

The r.m.s. value of an alternating current is that steady current that delivers the same average power as the a.c. to a resistive load.

The lamps in Practical Activity 27.2 are the 'resistive loads'. A full analysis, which we will come to shortly, shows that $I_{r.m.s.}$ is related to I_0 by:

$$I_{r.m.s.} = \frac{I_0}{\sqrt{2}} \approx 0.707 \times I_0$$

This is where the factor of 70% comes from. Note that this factor only applies to sinusoidal alternating currents.

We also have r.m.s. voltage $V_{r.m.s.}$ across the resistive load. $V_{r.m.s}$ is related to the peak voltage V_0 by:

$$V_{r.m.s.} = \frac{V_0}{\sqrt{2}}$$

KEY EQUATIONS

Root-mean-square value:

$$I_{r.m.s.} = \frac{I_0}{\sqrt{2}} \approx 0.707 \times I_0$$

where I_0 is the peak (maximum) current.

$$V_{r.m.s.} = \frac{V_0}{\sqrt{2}} \approx 0.707 \times V_0$$

where V_0 is the peak (maximum) voltage.

PRACTICAL ACTIVITY 27.2

Comparing alternating current (a.c.) and direct current (d.c.)

Because the power supplied by an alternating current is varying all the time, we need to have some way of describing the **average power** that is being supplied. To do this, we compare an alternating current with a direct current, and try to find the direct current that supplies the same average power as the alternating current.

Figure 27.8 shows how this can be done in practice. Two filament lamps (our resistive loads) are placed side by side; one is connected to an a.c. supply (on the right) and the other to a d.c. supply (the batteries on the left). The a.c. supply is adjusted so that the two lamps are equally bright, indicating that the two supplies are providing energy at the same average rate. The output voltages are then compared on the double-beam oscilloscope.

A typical trace is shown in Figure 27.9. This shows that the a.c. trace sometimes rises above the steady d.c. trace, and sometimes falls below it. This makes sense: sometimes the a.c. is delivering more power than the d.c., and sometimes less, but the average power is the same for both.

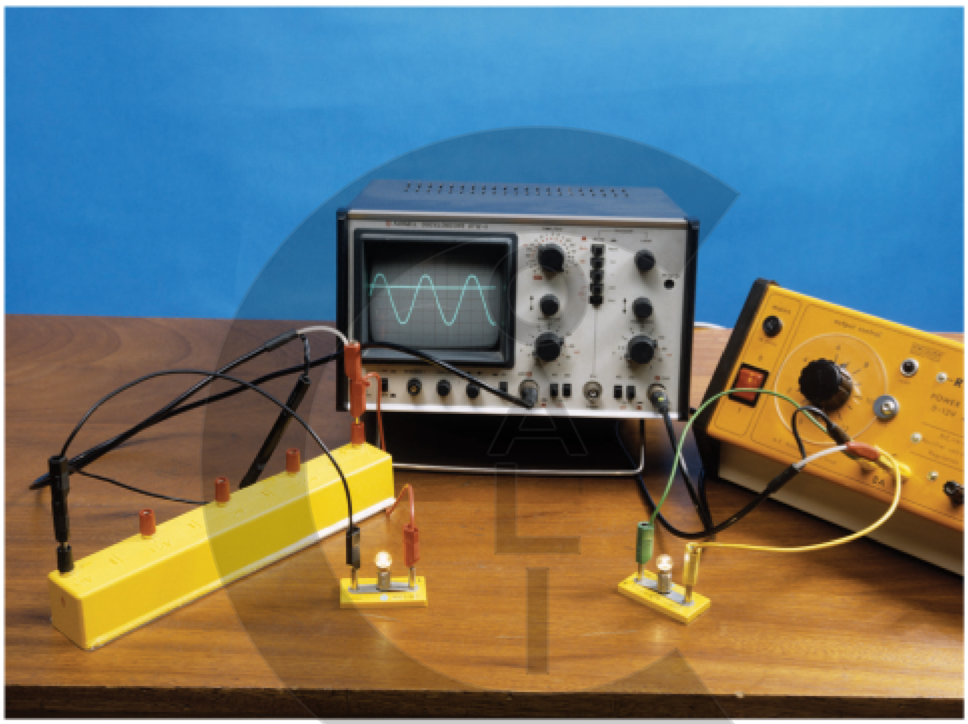


Figure 27.8: Comparing direct and alternating currents that supply the same power. The lamps are equally bright.

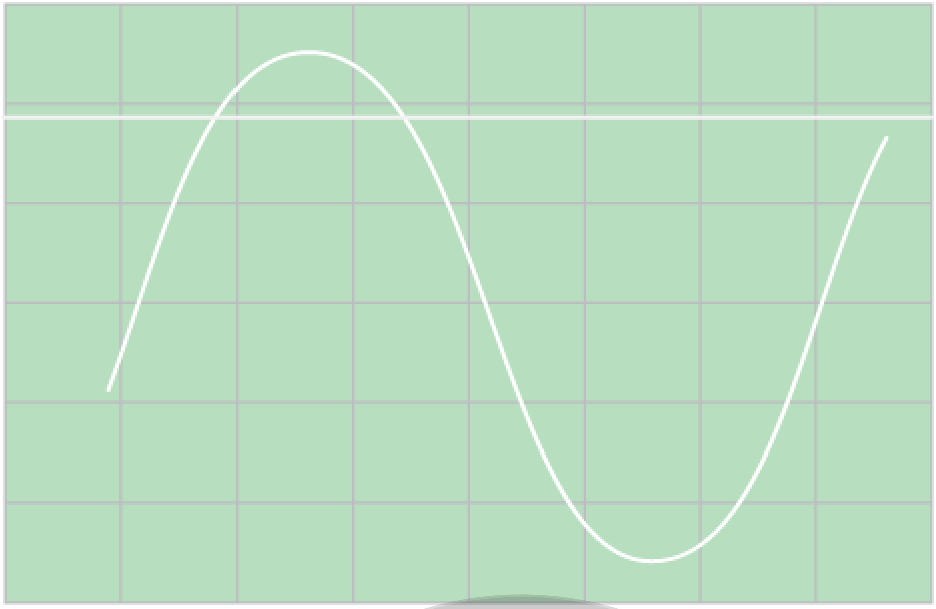


Figure 27.9: The oscilloscope trace from the experiment shown in Figure 27.8.

Questions

7 The alternating current (in ampere, A) in a resistor is represented by the equation: $I = 2.5 \sin(100\pi t)$

Calculate the r.m.s. value for this alternating current.

8 The mains supply to domestic consumers in many European countries has an r.m.s. value of 230 V for the alternating voltage. (Note that it is the r.m.s. value that is generally quoted, not the peak value.)

Calculate the peak value of the alternating voltage.

Calculating power

The importance of r.m.s. values is that they allow us to apply equations from our study of direct current to situations where the current is alternating. So, to calculate the average power dissipated in a resistor, we can use the usual formulae for power:

$$P = I^2R = IV = \frac{V^2}{R}$$

Remember that it is essential to use the r.m.s. values of I and V, as in Worked example 1. If you use peak values, your answer will be too great by a factor of 2.

Where does this factor of 2 come from? Recall that r.m.s. and peak values are related by:

$$I_0 = \sqrt{2} I_{r.m.s}$$

So, if you calculate I^2R using I_0 instead of $I_{r.m.s.}$, you will introduce a factor of $(\sqrt{2})^2$ or 2. The same is true if you calculate power using V_0 instead of $V_{r.m.s.}$. It follows that, for a sinusoidal alternating current, peak power is twice average power.

WORKED EXAMPLE

1 A 20 Ω resistor is connected to an alternating supply. The voltage across the resistor has peak value 25 V.

Calculate the average power dissipated in the resistor.

Step 1 Calculate the r.m.s. value of the voltage.

$$V_{r.m.s} = \frac{V_0}{\sqrt{2}} = \frac{25}{\sqrt{2}} = 17.7\ \text{V}$$

Step 2 Now calculate the average power dissipated. (Remember you must use the r.m.s. value, and not the peak value.)

$$P = \frac{V^2}{R} = \frac{17.7^2}{20} = 15.6\ \text{W}$$

Note that, if we had used V_0 rather than $V_{r.m.s.}$, we would have found:

$$P = \frac{25^2}{20} = 31.3\ \text{W}$$

which is double the correct answer.

Questions

9 Calculate the average power dissipated in a resistor of resistance 100 Ω when a sinusoidal alternating current has a peak value of 3.0 A.

10 The sinusoidal voltage across a 1.0 kΩ resistor has a peak value 325 V.

- **a** Calculate the r.m.s. value of the alternating voltage.
- **b** Use $V = IR$ to calculate the r.m.s. current in the resistor.
- **c** Calculate the average power dissipated in the resistor.
- **d** Calculate the **peak** power dissipated in the resistor.

Explaining root-mean-square

We will now briefly consider the origin of the term root-mean-square and show how the factor of $\sqrt{2}$ in the equation $I_0 = \sqrt{2} I_{r.m.s}$ comes about.

The equation $P = I^2R$ shows us that the power P is directly proportional to the square of the current I. Figure 27.10 shows how we can calculate I^2 for an alternating current. The current I varies sinusoidally, and during half of each cycle it is negative. However, I^2 is always positive (because the square of a negative number is positive). Notice that I^2 varies up and down, and that it has twice the frequency of the current.

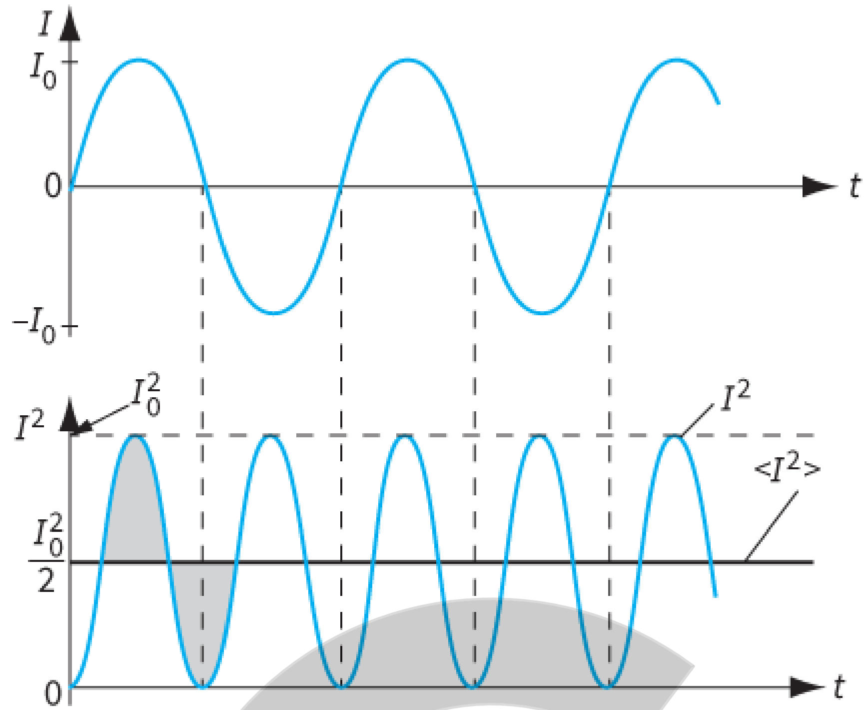


Figure 27.10: An alternating current I is alternately positive and negative, while I^2 is always positive.

Now, if we consider $<I^2>$, the average (mean) value of I^2, we find that its value is half of the square of the peak current (because the graph is symmetrical). That is:

$$< I^2 > = \frac{1}{2} I_0^2$$

To find the r.m.s. value of I, we now take the square root of $<I^2>$.

This gives:

$$I_{r.m.s} = \sqrt{< I^2 >} = \sqrt{\frac{1}{2} I_0^2}$$

Or

$$I_0 = \sqrt{2} I_{r.m.s}$$

Summarising this process: to find the r.m.s. value of the current, we find the **r**oot of the **m**ean of the **s**quare of the current – hence **r.m.s**.

27.4 Rectification

Many electrical appliances work with alternating current. Some, like electrical heaters, will work equally well with d.c. or a.c. However, there are many appliances, such as electronic equipment, which require d.c. For these, the alternating mains voltage must be converted to direct voltage by the process of **rectification**.

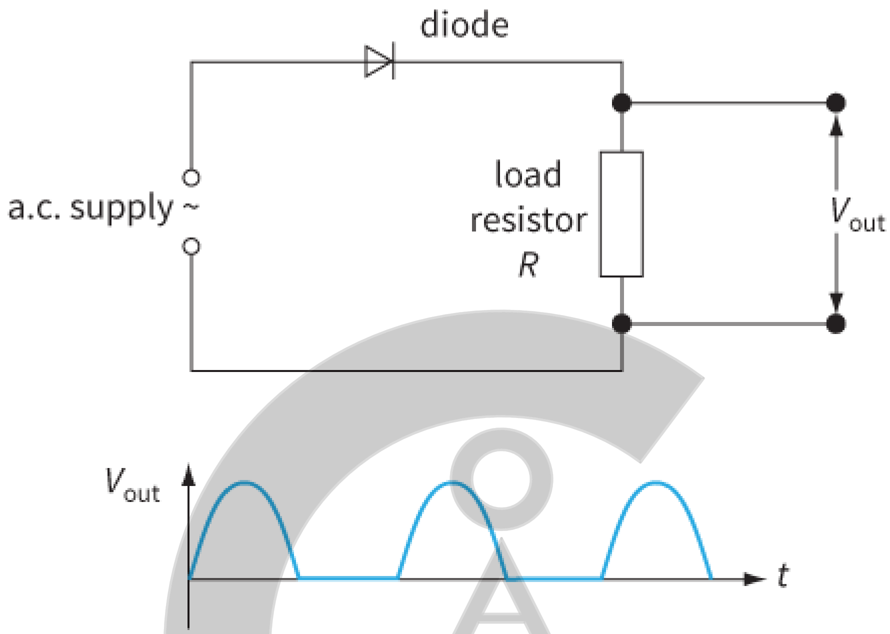


Figure 27.11: Half-wave rectification of a.c. requires a single diode.

A simple way to do this is to use a diode, which is a component that will only allow current in only one direction. (You have already met diodes in Chapter 10.) Figure 27.11 shows a circuit for doing this. An alternating input voltage is applied to a circuit with a diode and a resistor in series. The diode will only conduct during the positive cycles of the input voltage. Hence, there will be a current in the load resistor only during these positive cycles. The output voltage V_{out} across the resistor will fluctuate as shown in the V_{out} against time t graph. This graph is identical to the input alternating voltage, except the negative cycles have been 'chopped-off'.

This type of rectification is known as half-wave rectification. For one-half of the time the voltage is zero, and this means that the power available from a half-wave rectified supply is reduced.

The bridge rectifier

To overcome this problem of reduced power, a bridge rectifier circuit is used. This consists of four diodes connected across the input alternating voltage, as shown in Figure 27.12. The output voltage V_{out} is taken across the load resistor R. The resulting output voltage across the load resistor R is full-wave rectified.

The way in which this works is shown in Figure 27.13.

- During the **positive** cycles of the input voltage, A is positive and B is negative. The diodes 2 and 3 conduct because they are both in forward bias. The diodes 1 and 4 are in reverse bias, and therefore do not conduct. The current in the load resistor R will be downwards. Figure 27.13a shows the direction of the current.
- During the **negative** cycles of the input voltage, B is positive and A is negative. The diodes 4 and 1 conduct because they are now both in forward bias. The diodes 2 and 3 are in reverse bias, and therefore do not conduct. The current in the load resistor R will still be downwards. Figure 27.13b shows the direction of the current.

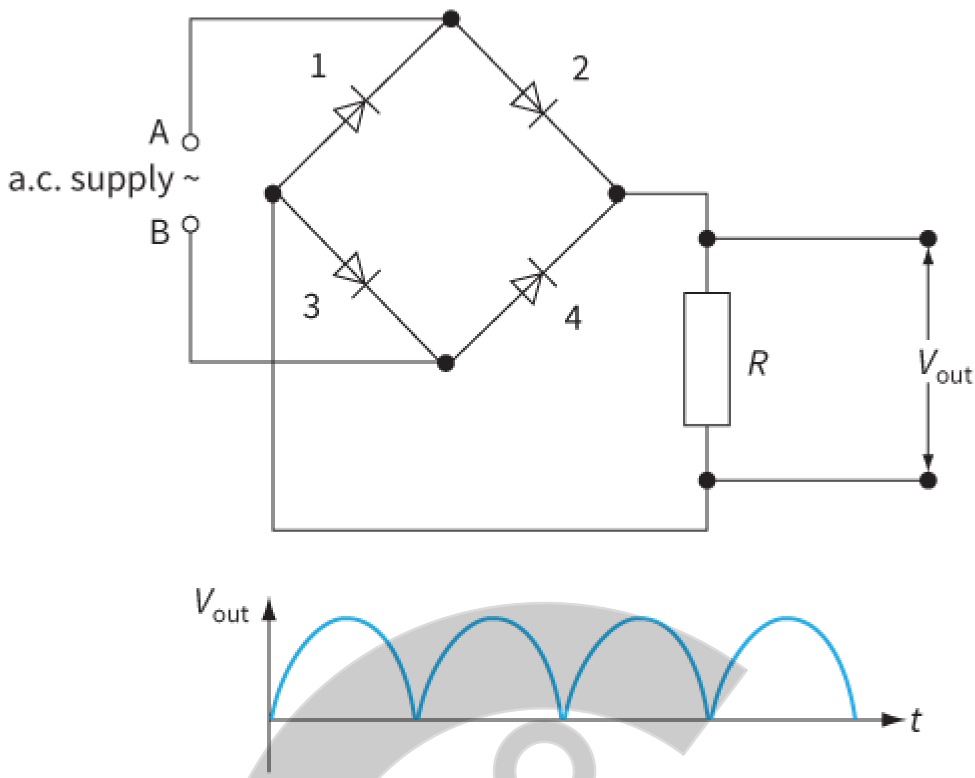


Figure 27.12: Full-wave rectification of a.c. using a diode bridge.

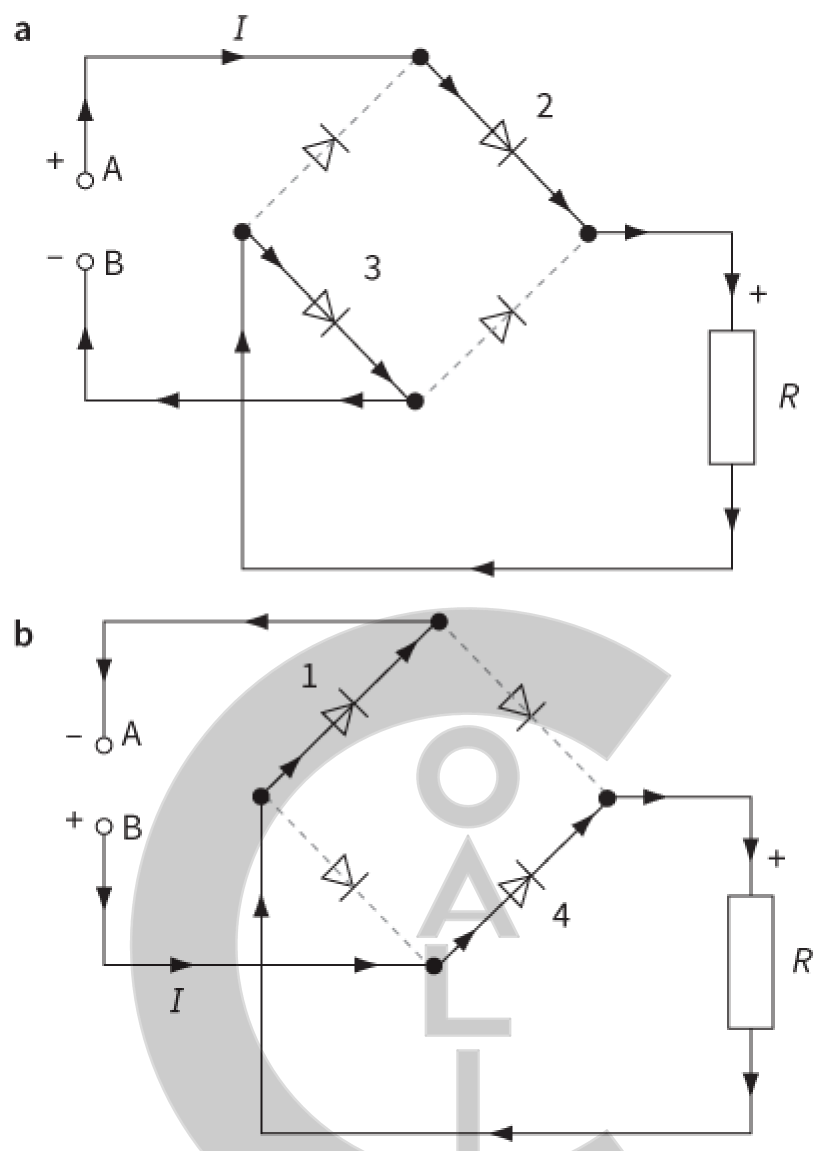


Figure 27.13: Direction of current during full-wave rectification **a** for positive cycles and **b** for negative cycles.

Note that in both positive and negative cycles, the current direction in the load resistor *R* is always the same (downwards). This means that the top end of *R* must always be positive.

You can construct a bridge rectifier using light-emitting diodes (LEDs) that light up when current flows through them. By connecting this bridge to a slow a.c. supply (for instance, 1 Hz from a signal generator), you can see the sequence in which the diodes conduct during rectification.

Question

11 Explain why, when terminal B in Figure 27.13 is positive (during the negative cycles), only diodes 1 and 4 conduct.

Smoothing

In order to produce steady d.c. from the 'bumpy' d.c. that results from rectification, a smoothing capacitor is necessary in the circuit. This capacitor, of capacitance *C*, is in parallel with the load resistor of resistance *R*. This is shown in Figure 27.14. The idea is that the capacitor charges up and maintains the voltage at a high level. It discharges gradually when the rectified voltage drops, but the voltage soon rises again and the capacitor charges up again. The result is an output voltage with 'ripple'.

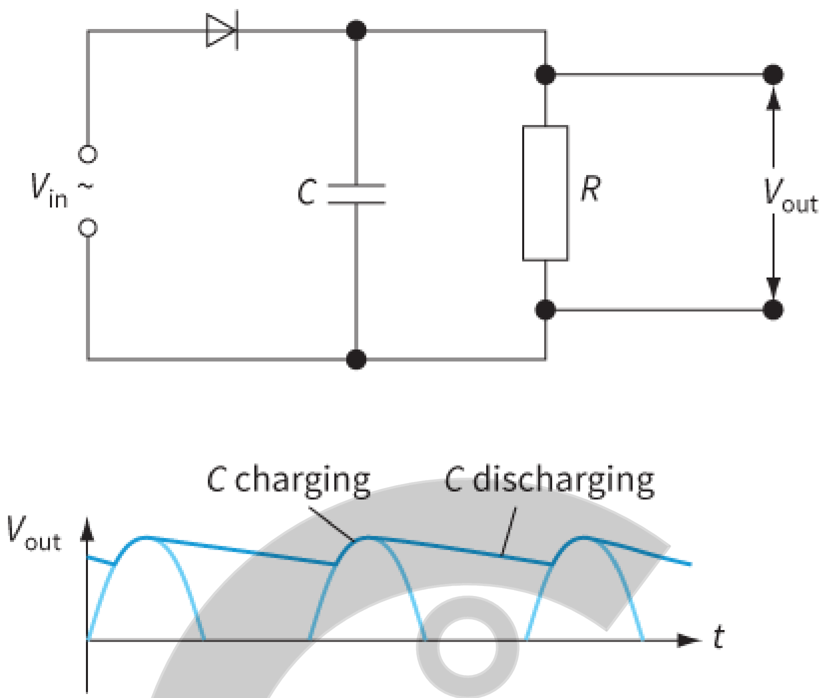


Figure 27.14: A smoothing capacitor is connected across (in parallel with) the load resistor.

The amount of ripple can be controlled by carefully choosing the capacitance *C* of the capacitor and the resistance *R* of the load resistor. A capacitor with a large capacitance value discharges more slowly than a capacitor with a small capacitance value, so will give a smaller ripple. Similarly, if the resistance *R* of the resistor is increased, then this too leads to a slower discharge of the capacitor. You may have already met the physics of discharging capacitors in Chapter 23. So, the size of the ripple can be reduced by increasing the time constant *CR* of the capacitor–resistor circuit. Ideally, though this is definitely not a general rule, *CR* must be much greater than the time interval between the adjacent peaks of the output signal – you want the capacitor to be still discharging between the 'gaps' between the positive cycles. This is illustrated in Worked example 2.

Note that, in Figures 27.11 to 27.14, we have represented the load on the supply by a resistor. This represents any components that are connected to the supply. For example, a rectifier circuit can be used to charge the battery of a mobile phone or provide a direct voltage supply for small radio.

WORKED EXAMPLE

2 Figure 27.15 shows the output voltage from a half-wave rectifier. The load resistor has resistance 1.2 kΩ. A student wishes to smooth the output voltage by placing a capacitor across the load resistor.

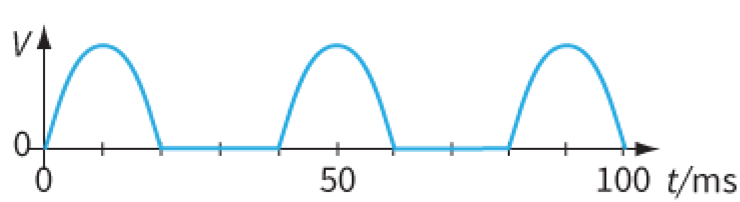


Figure 27.15: Output from a half-wave rectifier.

With the help of a calculation, suggest if a 10 pF capacitor or a 500 μF capacitor would be suitable for this task.

Step 1 Calculate the time constant with the 10 pF capacitor.

time constant = $CR = 10 \times 10^{-9} \times 1.2 \times 10^{3} = 1.2 \times 10^{-5}$ s (= 0.012 ms)

Step 2 Compare the time constant with the time interval between the adjacent peaks of the output signal.

The time constant of 0.012 ms is very small compared with time interval of 40 ms between the adjacent peaks of the output. If this capacitor were to be used, it would discharge far too quickly. There would be no smoothing of the output voltages – the 10 pF capacitor is not suitable.

Step 3 Repeat the steps for the 500 μF capacitor.

time constant = $CR = 500 \times 10^{-6} \times 1.2 \times 10^{3} = 0.60$ s (= 600 ms)

Now, the time constant of 600 ms is much larger than 40 ms. This capacitor will not discharge completely between the positive cycles of the half-wave rectified signal. The 500 μF capacitor would be adequate for the smoothing task.

Questions

12 Sketch the following voltage patterns:

- **a** a sinusoidal alternating voltage
- **b** the same voltage as part **a**, but half-wave rectified
- **c** the same voltage as part **b**, but smoothed
- **d** the same voltage as part **a**, but full-wave rectified
- **e** the same voltage as part **d**, but smoothed.

13 A student wires a bridge rectifier incorrectly as shown in Figure 27.16. Explain what you would expect to observe when an oscilloscope is connected across the load resistor *R*.

14 A bridge rectifier circuit is used to rectify an alternating current through a resistor. A smoothing capacitor is connected across the resistor. Figure 27.17 shows how the current varies. Use sketches to show the changes you would expect:

- **a** if the resistance *R* of the resistor is increased
- **b** if the capacitance *C* of the capacitor is decreased.

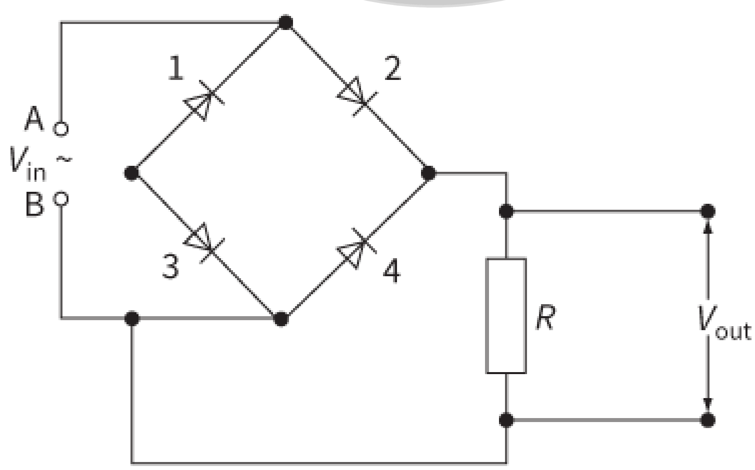


Figure 27.16: A bridge rectifier circuit that is wired incorrectly. For Question 13.

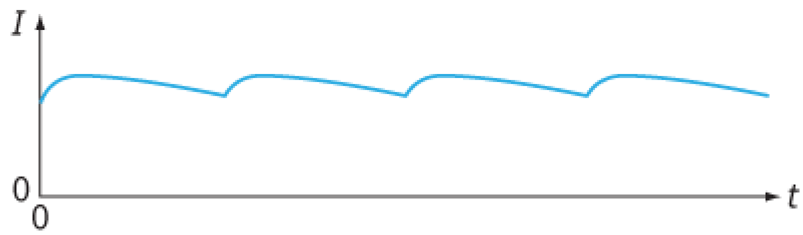


Figure 27.17: A smoothed, rectified current. For Question 14.

REFLECTION

Without looking at your textbook, summarise all the key equations from this chapter.

Make a list of mains operated devices in your laboratory. For each device, determine the power, r.m.s. current and r.m.s. voltage.

Give yourself and a classmate one minute to draw a circuit diagram for a full-wave rectifier circuit. Compare your circuit diagrams. Which diagram was more accurate? How would you make this diagram more accurate if you were to draw it in the future?

SUMMARY

A sinusoidal alternating current can be represented by $I = I_0 \sin \omega t$, where I_0 is the peak value of the current.

The root-mean-square (r.m.s.) value of an alternating current is that steady current that delivers the same average power as the a.c. to a resistive load; for a sinusoidal a.c.:

$$I_{r.m.s.} = \frac{I_0}{\sqrt{2}} \approx 0.707 \times I_0$$

The relationship between root-mean-square (r.m.s.) voltage $V_{r.m.s}$ and peak voltage V_0 is:

$$V_{r.m.s.} = \frac{V_0}{\sqrt{2}}$$

The power P dissipated in a resistor can be calculated using the equations:

$$P = VI,\ P = I^2R \text{ and } P = \frac{V^2}{R}$$

where V and I are the r.m.s. values of the voltage and current, respectively.

A single diode is used for the half-wave rectification of an alternating current. Four diodes (bridge rectifier) are used for the full-wave rectification of an alternating current.

A capacitor placed in parallel with a resistive load will smooth the rectified alternating voltage. The greater the time constant CR of the capacitor-resistor network, the smaller is the size of the ripple.

EXAM-STYLE QUESTIONS

1 The **maximum** power dissipated in a resistor carrying an alternating current is 10 W.

What is the mean power dissipated in the resistor? **[1]**

A 5.0 W

B 7.1 W

C 10 W

D 14 W

2 The alternating current I in ampere (A) in a filament lamp is represented by the equation:

$I = 1.5 \sin(40t)$.

Which of the following is correct? **[1]**

A The angular frequency of the alternating current is 40 rad s^{-1}.

B The frequency of alternating current is 40 Hz.

C The maximum current is 3.0 A.

D The peak voltage is 1.5 V.

3 Write down a general expression for the sinusoidal variation with time t of:

a an alternating voltage V **[1]**

b an alternating current I (you may assume that I and V are in phase) **[1]**

c the power P dissipated due to this current and voltage. **[1]**

[Total: 3]

4 The alternating current I in ampere (A) in a circuit is represented by the equation:

$I = 2.0 \sin(50\pi t)$.

a State the peak value of the current. **[1]**

b Calculate the frequency of the alternating current. **[2]**

c Sketch a graph to show **two** cycles of the variation of current with time. Mark the axes with suitable values. **[2]**

d Calculate $I_{r.m.s.}$, the r.m.s. value of current, and mark this on your graph in part c. **[1]**

e Determine **two** values of time t at which the current $I = I_{r.m.s.}$. **[3]**

[Total: 9]

5 A heater of resistance 6.0 Ω is connected to an alternating current supply. The output voltage from the supply is 20 V r.m.s.

Calculate:

a the average power dissipated in the heater **[2]**

b the maximum power dissipated in the heater **[1]**

c the energy dissipated by the heater in 5.0 minutes. **[2]**

[Total: 5]

6 An oscilloscope is used to display the variation of voltage across a 200 Ω resistor with time. The trace is shown. The time-base of the oscilloscope is set at 5 ms div^{-1} and the Y-gain at 0.5 V div^{-1}.

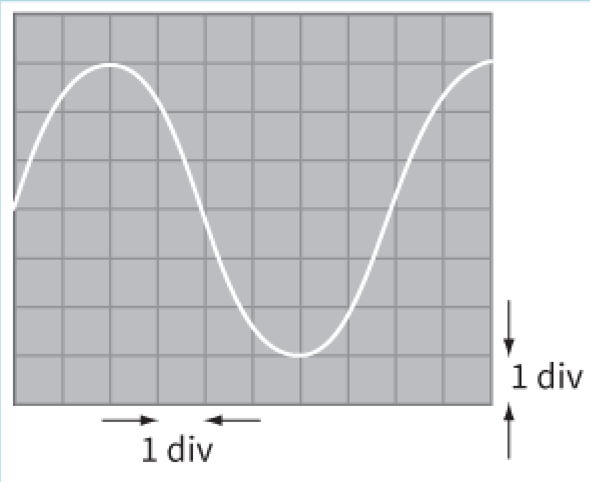


Figure 27.18

Determine:

a the period and hence the frequency of the alternating voltage **[2]**

b the peak voltage and hence the r.m.s. voltage **[2]**

c the r.m.s. current in the resistor **[1]**

d the mean power dissipated in the resistor. **[2]**

[Total: 7]

7 **a** State the relationship between the peak current I_0 and the r.m.s. current I_{rms} for a sinusoidally varying current. **[1]**

b The current in a resistor connected to a steady d.c. supply is 2.0 A. When the same resistor is connected to an a.c. supply, the current in it has a peak value of 2.0 A. The heating effects of the two currents in the resistor are different.

i Explain why the heating effects are different and state which heating effect is the greater. **[2]**

ii Calculate the ratio of the power dissipated in the resistor by the d.c. current to the power dissipated in the resistor by the a.c. current. **[2]**

[Total: 5]

8 A sinusoidal voltage of 6.0 V r.m.s. and frequency 50 Hz is connected to a diode and a resistor R of resistance 400 Ω as shown in the diagram.

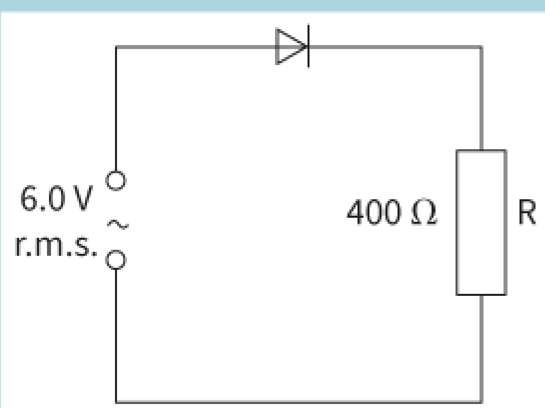


Figure 27.19

a Sketch a graph showing the variation with time of both the supply waveform (use a dotted line) and the voltage across R (use a solid line). Put numerical scales on both the voltage and time axes. **[4]**

b An uncharged capacitor C is connected across R. When the 6.0 V r.m.s. supply is switched on, the capacitor charges fully during the first quarter of a cycle. You may assume that the p.d. across the diode is zero when it conducts. For the next three-quarters of the first cycle, the diode stops conducting and the p.d. across R falls to one-half of the peak value. During this time the mean p.d. across R is 5.7 V.

For the last three-quarters of the first cycle, calculate:

i the time taken **[1]**

ii the mean current in R **[2]**

iii the charge flowing through R **[2]**

iv the capacitance of C. **[2]**

c Explain why the diode stops conducting during part of each cycle in part **b**. **[2]**

[Total: 13]

9 The rectified output from a circuit is connected to a resistor R of resistance 1000 Ω. Graph A shows the variation with time t of the p.d. V across the resistor. Graph B shows the variation of V when a capacitor is placed across R to smooth the output.

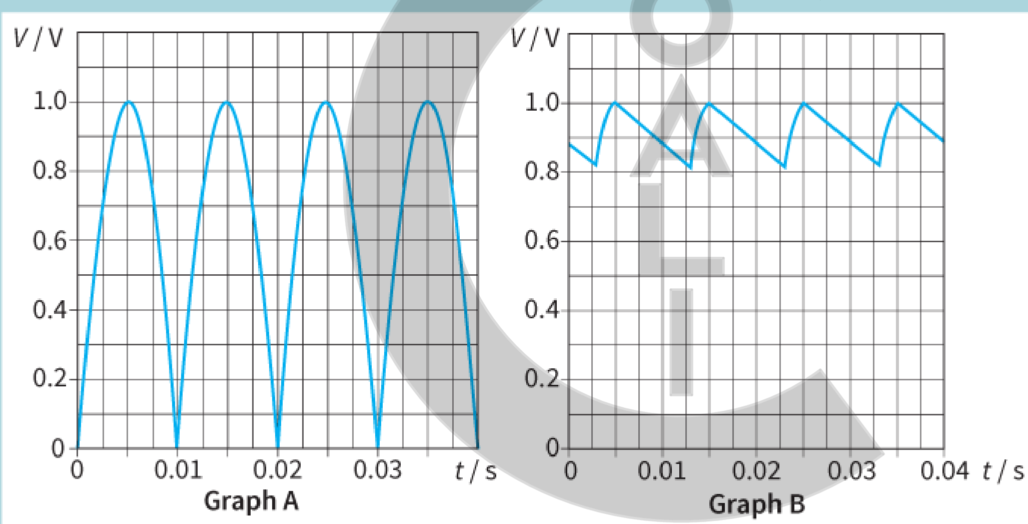


Figure 27.20

Explain how the rectification is achieved. Draw a circuit diagram to show the components involved. **[6]**

b Explain the action of the capacitor in smoothing the output. **[3]**

c Using graph B between t = 0.005 and t = 0.015 s, determine:

i the time during which the capacitor is charging **[1]**

ii the mean value of the p.d. across R **[1]**

iii the average power dissipated in R. **[2]**

[Total: 13]

10 Electrical energy is supplied by a high-voltage power line that has a total resistance of 4.0 Ω. At the input to the line, the root-mean-square (r.m.s.) voltage has a value of 400 kV and the input power is 500 MW.

a **i** Explain what is meant by **root-mean-square voltage**. **[2]**

ii Calculate the minimum voltage that the insulators that support the line must withstand without breakdown. [2]

b i Calculate the value of the r.m.s. current in the power line. [2]

ii Calculate the power loss on the line. [2]

iii Suggest why it is an advantage to transmit the power at a high voltage. [2]

[Total: 10]

11 A student has designed a full-wave rectifier circuit.

The output voltage for this circuit is taken across a resistor of resistance 120 Ω. The variation of the output voltage with time is shown.

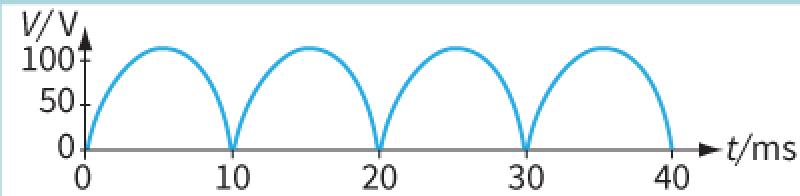


Figure 27.21

A capacitor is now connected across the resistor. The graph shows the new variation of the output voltage with time.

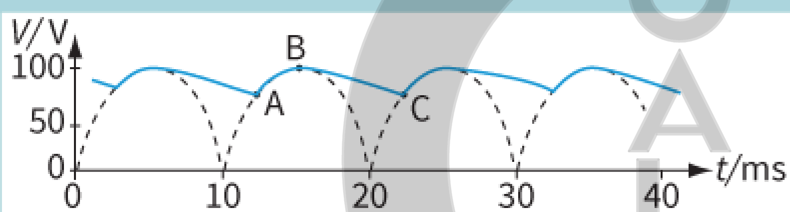


Figure 27.22

a Explain the variation of the output variation between points:

i AB [1]

ii BC. [1]

b Use the second graph to determine the value of the capacitance *C*. [3]

(You may use the equation $V = V_0 e^{-\frac{t}{CR}}$ from Chapter 23.)

[Total: 5]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic...	Needs more work	Almost there	Ready to move on
understand the terms period, frequency and peak value as applied to an alternating current or voltage	26.1, 26.2			
use the equations $I = I_0 \sin \omega t$ and $V = V_0 \sin \omega t$ for sinusoidally alternating current and voltage, respectively	26.1, 26.2			
understand that the mean power in a resistive load is half the maximum power for a sinusoidal alternating current	27.3			
understand root-mean-square (r.m.s.) and peak values	27.3			
recall and use: $I_{r.m.s.} = \frac{I_0}{\sqrt{2}}$ and $V_{r.m.s} = \frac{V_0}{\sqrt{2}}$	27.3			
understand half-wave and full-wave rectification	27.4			
explain how a single diode produces half-wave rectification	27.4			
explain how four diodes (bridge rectifier) produce full-wave rectification	27.4			
understand smoothing capacitors, and understand how smoothing effects are governed by capacitance of the smoothing capacitor and the resistance of the load resistor.	27.4			


C
O
A
L
I

〉Chapter 28

Quantum physics

LEARNING INTENTIONS

In this chapter you will learn how to:

- understand that electromagnetic radiation has a particulate nature
- understand that a photon is a quantum of electromagnetic energy
- recall and use $E = hf$
- use the electronvolt (eV) as a unit of energy
- understand that a photon has momentum and that the momentum is given by $p = \frac{E}{c}$
- understand that photoelectrons may be emitted from a metal surface when it is illuminated by electromagnetic radiation
- understand and use the terms threshold frequency and threshold wavelength
- explain photoelectric emission in terms of photon energy and work function energy
- recall and use $hf = \Phi + \frac{1}{2}mv_{max}{}^2$
- explain why the maximum kinetic energy of photoelectrons is independent of intensity, whereas the photoelectric current is proportional to intensity
- understand that the photoelectric effect provides evidence for a particulate nature of electromagnetic radiation, while phenomena such as interference and diffraction provide evidence for a wave nature
- describe and interpret qualitatively the evidence provided by electron diffraction for the wave nature of particles
- understand the de Broglie wavelength as the wavelength associated with a moving particle
- recall and use $\lambda = \frac{h}{p}$
- understand that there are discrete electron energy levels in isolated atoms (such as atomic hydrogen)
- understand the appearance and formation of emission and absorption line spectra
- recall and use the relation $hf = E_1 - E_2$.

BEFORE YOU START

- The principle of conservation of energy is an important idea in physics. Write down some examples of this from several topics in physics. Share your list with a partner.
- In pairs, discuss the concepts of momentum and kinetic energy.

WHAT IS LIGHT?

When the first laser was made in 1960, it seemed like a clever idea, but it was a long time before it found any useful application. Today, lasers are everywhere – in CD and DVD machines, computer disc drives, supermarket barcode scanners–there are probably more lasers than people. Figure 28.1 shows a patient undergoing laser eye surgery.

The invention of the laser was only possible when scientists had cracked the mystery of the nature of light.

You already know that light is a wave. What experimental evidence is there for the wave-like behaviour of such waves? You will see in this chapter that electromagnetic waves have a dual nature – they interact with matter as ‘particles’ and propagate through space as a wave.

Figure 28.1: This patient is undergoing laser eye surgery, which improves the focusing of the eye by modifying the shape of the surface of the eyeball.

28.1 Modelling with particles and waves

In this chapter, we will study two very powerful scientific models – particles and waves – to see how they can help us to understand more about both light and matter. First, we will take a closer look at each of these models in turn.

Particle models

In order to explain the properties of matter, we often think about the particles of which it is made and the ways in which they behave. We imagine particles as being objects that are hard, have mass and move about according to the laws of Newtonian mechanics (Figure 28.2). When two particles collide, we can predict how they will move after the collision, based on knowledge of their masses and velocities before the collision. If you have played snooker or pool, you will have a pretty good idea of how particles behave.

Particles are a macroscopic model. Our ideas of particles come from what we observe on a macroscopic scale– when we are walking down the street, or observing the motion of stars and planets, or working with trolleys and balls in the laboratory. But what else can we explain using a particle model?

The importance of particle models is that we can apply them to the microscopic world, and explain more phenomena.

We can picture gas molecules as small, hard particles, rushing around and bouncing haphazardly off one another and the walls of their container. This is the kinetic model of a gas that we studied in depth in Chapter 20. We can explain the macroscopic (larger scale) phenomena of pressure and temperature in terms of the masses and speeds of the microscopic particles. This is a very powerful model, which has been refined to explain many other aspects of the behaviour of gases.

Table 28.1 shows how, in particular topics of science, we can use a particle model to interpret and make predictions about macroscopic phenomena.

Figure 28.2: Pool balls provide a good model for the behaviour of particles on a much smaller scale.

Topic	Model	Macroscopic phenomena
electricity	flow of electrons	current
gases	kinetic theory	pressure, temperature and volume of a gas
solids	crystalline materials	mechanical properties
radioactivity	nuclear model of the atom	radioactive decay, fission and fusion reactions
chemistry	atomic structure	chemical reactions

Table 28.1: Particle models in science.

Wave models

Waves are something that we see on the sea. There are tidal waves, and little ripples. Some waves have foamy tops, others are breaking on the beach.

Physicists have an idealised picture of a wave – it is shaped like a sine graph. You will not see any waves quite this shape on the sea. However, it is a useful picture, because it can be used to represent some simple phenomena. More complicated waves can be made up of several simple waves, and physicists can cope with the mathematics of sine waves. (This is the principle of superposition, which we looked at in detail in Chapter 13.)

Waves are a way in which energy is transferred from one place to another. In any wave, something is changing in a regular way, while energy is travelling along. In water waves, the surface of the water moves up and down periodically and energy is transferred horizontally.

Table 28.2 shows some phenomena that we explain in terms of waves.

Phenomenon	Varying quantity
sound	pressure (or density)
light (and other electromagnetic waves)	electric field strength and magnetic flux density
waves on strings	displacement

Table 28.2: Wave models in science.

The characteristic properties of waves are that they all show reflection, refraction, diffraction and interference. Waves themselves do not have mass or charge. Since particle models can also explain reflection and refraction, it is **diffraction** and **interference** that we regard as the defining characteristics of waves. If we can show diffraction and interference, we know that we are dealing with waves (Figure 28.3).

Figure 28.3: A diffraction grating splits up light into its component colours and can produce dramatic effects in photographs.

Waves or particles?

Wave models and particle models are both very useful. They can explain a great many different observations. But which should we use in a particular situation? And what if both models seem to work when we are trying to explain something?

This is just the problem that physicists struggled with for over a century, in connection with light. Does light travel as a wave or as particles?

For a long time, Newton's view prevailed–light travels as particles. This was set out in 1704 in his famous book *Opticks*. He could use this model to explain both reflection and refraction. His model suggested that light travels **faster** in water than in air. In 1801, Thomas Young, an English physicist, demonstrated that light showed diffraction and interference effects. Physicists were still very reluctant to abandon Newton's particle model of light. The ultimate blow to Newton's model came from the work carried out by the French physicist Léon Foucault in 1853. His experiments on the speed of light showed that light travelled more **slowly** in water than in air. Newton's model had at last been tested and it was in direct contradiction with experimental results. Most scientists had to accept that light travelled through space as a wave.

28.2 Particulate nature of light

We expect light to behave as waves, but can light also behave as particles? The answer is yes, and you are probably already familiar with some of the evidence.

If you place a Geiger counter next to a source of gamma radiation, you will hear an irregular series of clicks. The counter is detecting γ-rays (gamma-rays). But γ-rays are part of the electromagnetic spectrum. They belong to the same family of waves as visible light, radio waves, X-rays and so on.

So, here are waves giving individual or discrete clicks, which are indistinguishable from the clicks given by γ-particles (alpha-particles) and γ-particles (beta-particles). We can conclude that γ-rays behave like particles when they interact with the gas particles within a Geiger counter.

This effect is most obvious with γ-rays, because they are at the most energetic end of the electromagnetic spectrum. It is harder to show the same effect for visible light.

Photons

The **photoelectric effect**, and Einstein's explanation of it, convinced physicists that light could behave as a stream of particles. Before we go on to look at this in detail, we need to see how to calculate the energy of photons.

Newton used the word **corpuscle** for the particles that he thought made up light. Nowadays, we call them **photons** and we believe that all electromagnetic radiation consists of photons. A photon is a 'packet of energy' or a quantum of electromagnetic energy. Gamma-photons (γ-photons) are the most energetic. According to Albert Einstein, who based his ideas on the work of another German physicist, Max Planck, the energy E of a photon in joules (J) is related to the frequency f in hertz (Hz) of the electromagnetic radiation of which it is part, by the equation:

$$E = hf$$

The constant h has an experimental value equal to 6.63×10^{-34} J s.

This constant h is called the **Planck constant**. It has units of joule seconds (J s), but you may prefer to think of this as 'joules per hertz'. The energy of a photon is directly proportional to the frequency of the electromagnetic waves, that is:

$$E \propto f$$

Hence, high-frequency radiation means high-energy photons.

Notice that the equation $E = hf$ shows us the relationship between a particle-like property (the photon energy E) and a wave-like property (the frequency f). It is called the **Einstein relation** and applies to all electromagnetic waves.

The frequency f and wavelength λ of an electromagnetic wave are related to the wave speed c by the wave equation $c = f\lambda$, so we can also write this equation as:

$$E = \frac{hc}{\lambda}$$

where h is the Planck constant, f is frequency and λ is wavelength.

KEY EQUATION

Einstein relation:

$$E = hf \text{ and } E = \frac{hc}{\lambda}$$

It is worth noting that the energy of the photon is inversely proportional to the wavelength. Hence the short-wavelength X-ray photon is far more energetic than the long-wavelength photon of light.

Now, we can work out the energy of a γ-photon. Gamma-rays typically have frequencies greater than 10^{20} Hz. The energy of a γ-photon is therefore greater than $(6.63 \times 10^{-34} \times 10^{20}) \approx 10^{-13}$ J. This is a very small amount of energy on the human scale, so we don't notice the effects of individual γ-photons. However, some astronauts have reported seeing flashes of light as individual cosmic rays, high-energy γ-photons, passed through their eyeballs.

The energy of individual photons can be quite small, but the rate at which photons emitted by a source can be enormous. This is illustrated in Worked example 1 for a light-emitting diode.

WORKED EXAMPLE

1 A light-emitting diode (LED) emits light of wavelength 670 nm. The radiant power of the light from the LED is 50 mW.

Calculate the rate at which photons are emitted from this LED.

Step 1 Calculate the energy E of a single photon.

$$E = \frac{hc}{\lambda} = \frac{6.63\times10^{-34}\times3.00\times10^{8}}{670\times10^{-9}} = 2.97\times10^{-19}\text{ J}$$

(Note: 1 nm = 10^{-9} m)

Step 2 Calculate the rate of photons emitted.

The rate at which the photons are emitted is the equivalent to the number of photons emitted per second.

$$\text{rate of photon emission} = \frac{\text{power}}{\text{energy of a single photon}} = \frac{50\times10^{-3}}{2.97\times10^{-19}} = 1.68\times10^{17}\text{ s}^{-1}$$

(Note: 1 mW = 10^{-3} W)

In one second, there are about 1.7×10^{17} photons emitted by the LED.

Questions

To answer questions 1 to 7 you will need these values:

speed of light in a vacuum $c = 3.00 \times 10^8$ m s^{-1}

Planck constant $h = 6.63 \times 10^{-34}$ J s

1 Calculate the energy of a high-energy γ-photon, of frequency 1.0×10^{26} Hz.

2 Visible light has wavelengths in the range 400 nm (violet) to 700 nm (red). Calculate the energy of a photon of red light and a photon of violet light.

3 Determine the wavelength of the electromagnetic waves for each photon, a to e. Then use Figure 28.4 to identify the region of the electromagnetic spectrum to which each belongs.

The photon energy is:

a 10^{-12} J

b 10^{-15} J

c 10^{-18} J

d 10^{-20} J

e 10^{-25} J

4 A 1.0 mW laser produces red light of wavelength 6.48×10^{-7} m. Calculate how many photons the laser produces per second.

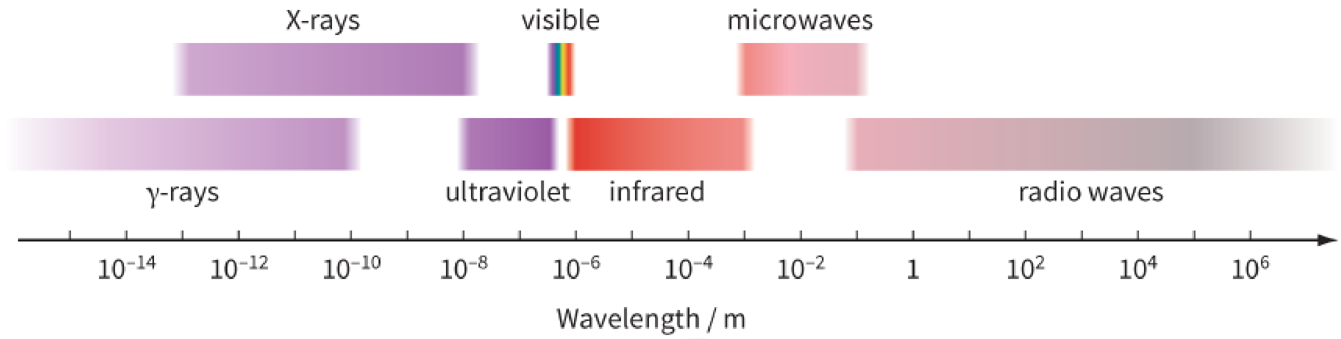


Figure 28.4: Wavelengths of the electromagnetic spectrum. The boundaries between some regions are fuzzy.

The electronvolt (eV)

The energy of a photon is extremely small and far less than a joule. Hence, the joule is not a very convenient unit for measuring photon energies. You may remember from Chapter 15 that we use another energy unit, the **electronvolt (eV)**, when considering amounts of energy much smaller than a joule.

To recap from Chapter 15: when an electron travels through a potential difference, energy is transferred. If an electron, which has a charge of magnitude 1.60×10^{-19} C, travels through a potential difference of 1 V, its energy change W is given by:

$$W = QV = 1.60 \times 10^{-19} \times 1 = 1.60 \times 10^{-19} \text{ J}$$

We can use this as the electronvolt:

One electronvolt (1 eV) is the energy gained by an electron travelling through a potential difference of one volt.

Therefore:

$$1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$$

So when an electron moves through 1 V, 1 eV of energy is gained or transferred to the electron. When one electron moves through 2 V, 2 eV of energy is gained. When five electrons move through 10 V, a total of 50 eV is transferred and so on.

- To convert from eV to J, multiply by 1.60×10^{-19}.
- To convert from J to eV, divide by 1.60×10^{-19}.

Question

5 An electron travels through a cell of e.m.f. 1.2 V.

Calculate the energy is transferred to the electron. Give your answer in both eV and J.

6 Calculate the energy in eV of an X-ray photon of frequency 3.0×10^{18} Hz.

7 With the help of a calculation, identify the region of the electromagnetic spectrum (Figure 28.4) a photon of energy 10 eV belongs.

When a charged particle is accelerated through a potential difference V, its kinetic energy increases. For an electron (charge e), accelerated from rest, we can write:

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

We need to be careful when using this equation. It does not apply when a charged particle is accelerated through a large voltage to speeds approaching the speed of light c. For this, we would have to take account of relativistic effects. (The mass of a particle increases as its speed gets closer to 3.00×10^8 m s^{-1}.)

Rearranging the equation gives the electron's speed:

$$v = \sqrt{\frac{2eV}{m}}$$

This equation applies to any type of charged particle, including protons (charge $+e$) and ions.

Question

8 A proton, initially at rest, is accelerated through a potential difference of 1500 V. A proton has charge $+1.60 \times 10^{-19}$ C and mass 1.67×10^{-27} kg.

Calculate:

a its final kinetic energy in joules (J)

b its final speed.

PRACTICAL ACTIVITY 28.1

Estimating the Planck constant *h*

You can obtain an estimate of the value of the Planck constant h by means of a simple experiment. It makes use of light-emitting diodes (LEDs) of different colours (Figure 28.5). You may recall from Chapter 10 that an LED conducts in one direction only (the forward direction) and that it requires a minimum voltage, the **threshold voltage**, to be applied in this direction before it allows a current. This experiment makes use of the fact that LEDs of different colours require different threshold voltages before they conduct and emit light.

- An LED giving light of red colour emits photons that are of low energy. It requires a low threshold voltage to make it conduct.
- An LED giving light of blue colour emits higher-energy photons. It requires a higher threshold voltage to make it conduct.

What is happening to produce photons of light when an LED conducts? The simplest way to think of this is to say that the electrical energy of a single electron passing through the diode is transferred to the energy of a single photon.

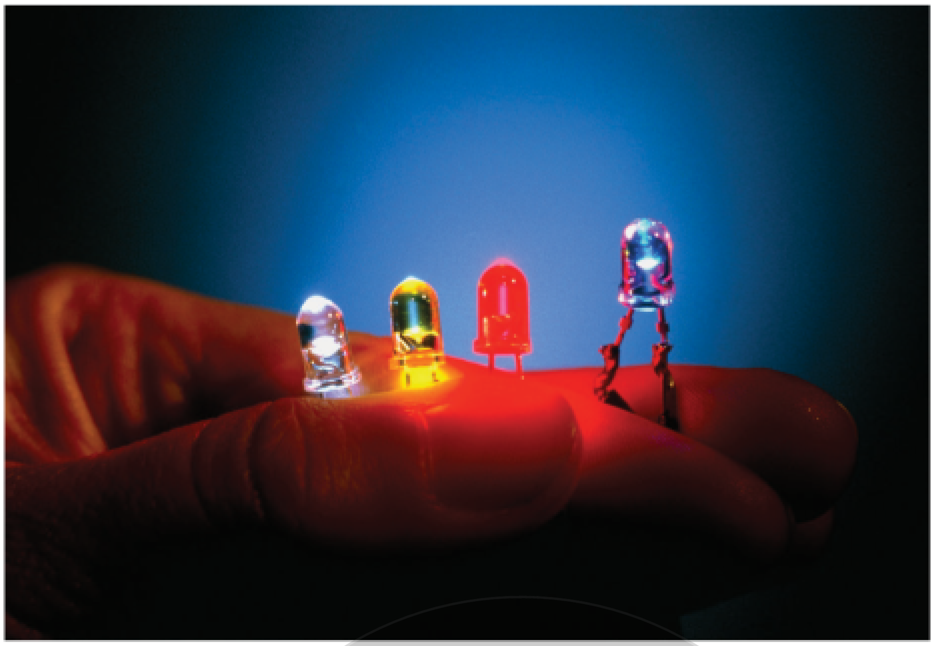


Figure 28.5: Light-emitting diodes (LEDs) come in different colours.

Hence, we can write:

energy transferred by electron = energy of photon

$$eV = \frac{hc}{\lambda}$$

where *V* is the threshold voltage for the LED. The values of *e* and *c* are known. Measurements of *V* and λ will allow you to calculate *h*. So the measurements required are:

- *V* – the voltage across the LED when it begins to conduct (its threshold voltage). It is found using a circuit like the one shown in Figure 28.6a.
- λ – the wavelength of the light emitted by the LED. This is found by measurements using a diffraction grating or from the wavelength quoted by the manufacturer of the LED.

If several LEDs of different colours are available, *V* and λ can be determined for each and a graph of *V* against $\frac{1}{\lambda}$ drawn (see Figure 28.6b). The graph passes through the origin and has gradient $\frac{hc}{e}$ and, hence, *h* can be estimated.

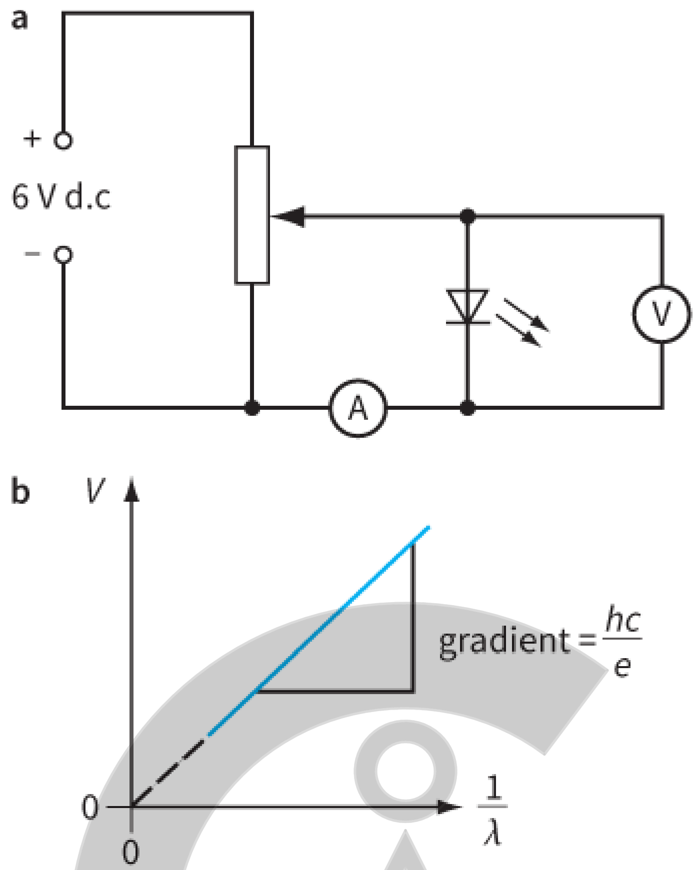


Figure 28.6: a A circuit to determine the threshold voltage required to make an LED conduct. An ammeter helps to show when this occurs. **b** The graph used to determine *h* from this experiment.

Question

9 In an experiment to determine the Planck constant *h*, LEDs of different colours were used. The p.d. required to make each conduct was determined, and the wavelength of their light was taken from the manufacturer's catalogue. The results are shown in Table 28.3. For each LED, calculate the experimental value for *h* and, hence, determine an average value for the Planck constant.

Colour of LED	Wavelength / 10^{-9} m	Threshold voltage / V
infrared	910	1.35
red	670	1.70
amber	610	2.00
green	560	2.30

Table 28.3: Results from an experiment to determine *h*.

28.3 The photoelectric effect

In the photoelectric effect, light shines on a metal surface and electrons are released from it. The Greek word for light is photo, hence, the word ‘photoelectric’. The electrons removed from the metal plate in this manner are often known as photoelectrons.

The apparatus used to observe the photoelectric effect is shown in Practical Activity 28.2. Light from a lamp is shone onto a negatively charged metal plate and some of the electrons in the metal are emitted. A simple explanation is that light is a wave that carries energy and this energy releases electrons from the metal. However, detailed observations of the effect at first proved difficult to explain, in particular, that there is a minimum threshold frequency of light below which no effect is observed.

PRACTICAL ACTIVITY 28.2

Observing the photoelectric effect

You can observe the photoelectric effect yourself by fixing a clean zinc plate to the top of a gold-leaf electroscope (Figure 28.7). Give the electroscope a negative charge and the leaf deflects. Now, shine electromagnetic radiation from a mercury discharge lamp on the zinc and the leaf gradually falls. (A mercury lamp strongly emits ultraviolet radiation.) Charging the electroscope gives it an excess of electrons. Somehow, the electromagnetic radiation from the mercury lamp helps electrons to escape from the surface of the metal.

Placing the mercury lamp closer causes the leaf to fall more rapidly. This is not very surprising. However, if you insert a sheet of glass between the lamp and the zinc, the radiation from the lamp is no longer effective. The gold leaf does not fall. Glass absorbs ultraviolet radiation and it is this component of the radiation from the lamp that is effective.

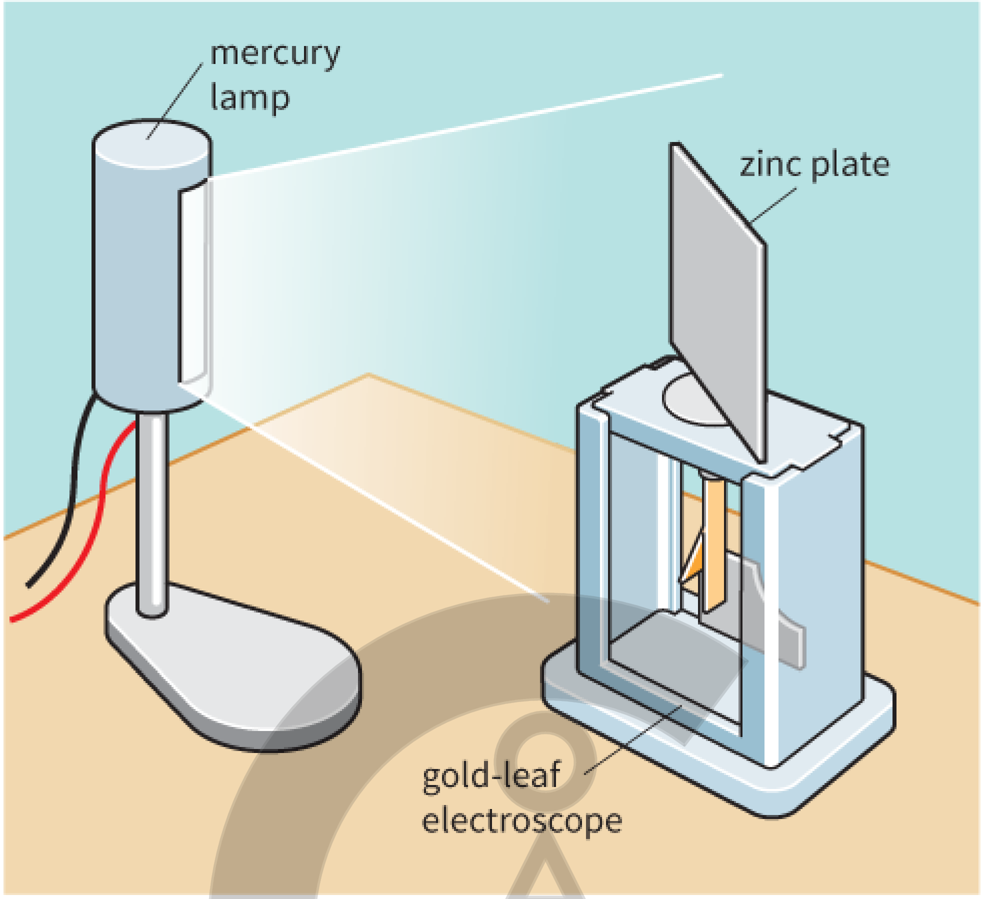


Figure 28.7: A simple experiment to observe the photoelectric effect.

28.4 Threshold frequency and wavelength

If you try the experiment described in Practical Activity 28.2 with a bright filament lamp, you will find it has no effect. A filament lamp does not produce ultraviolet radiation. There is a minimum frequency that the incident radiation must have in order to release electrons from the metal. This is called the **threshold frequency**. The threshold frequency is a property of the metal plate being exposed to electromagnetic radiation.

The threshold frequency is defined as the minimum frequency required to release electrons from the surface of a metal.

Since $c = f\lambda$, this implies that the threshold frequency has an equivalent longest wavelength for the liberation of electrons from the surface of a metal. This is called the **threshold wavelength**.

Threshold wavelength is the longest wavelength of the incident electromagnetic radiation that would eject electrons from the surface of a metal.

Physicists found it hard to explain why weak ultraviolet radiation could have an immediate effect on the electrons in the metal, but very bright light of lower frequency had no effect. They imagined light waves arriving at the metal, spread out over its surface and they could not see how weak ultraviolet waves could be more effective than the intense visible waves. In 1905, Albert Einstein came up with an explanation based on the idea of photons.

Metals (such as zinc) have electrons that are not very tightly held within the metal. These are the conduction electrons, and they are free to move about within the metal. When photons of electromagnetic radiation strike the metal, some electrons break free from the surface of the metal (Figure 28.8). They only need a small amount of energy (about 10^{-19} J) to escape from the metal surface.

We can picture the electrons as being trapped in an energy 'well' (Figure 28.9). A single electron requires a minimum energy Φ (Greek letter phi) to escape the surface of the metal. The **work function energy**, or simply **work function**, of a metal is the minimum amount of energy required by an electron to escape its surface. Energy is needed to release the surface electrons because they are attracted by the electrostatic forces due to the positive metal ions.

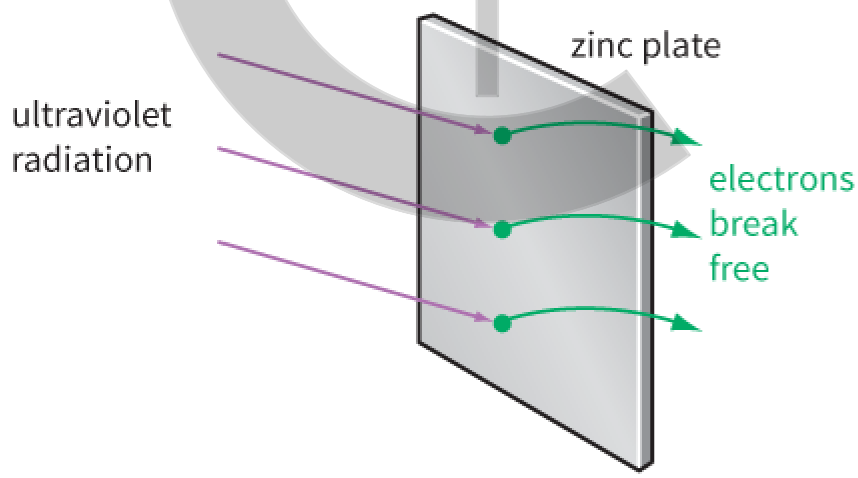


Figure 28.8: The photoelectric effect. When a photon of ultraviolet radiation strikes the metal plate, its energy may be sufficient to release an electron.

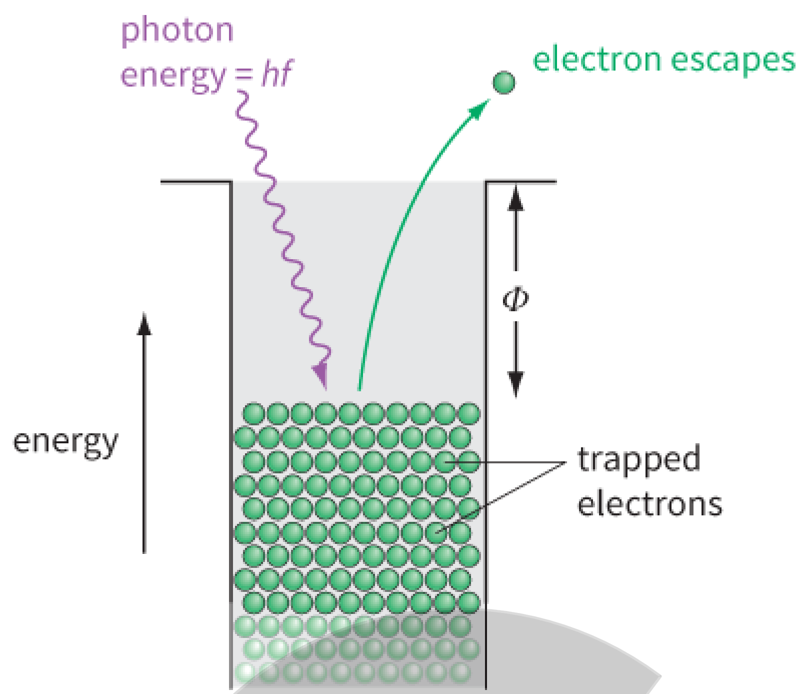


Figure 28.9: A single photon may interact with a single electron to release it.

Einstein did not picture electromagnetic waves interacting with all of the electrons in the metal. Instead, he suggested that a single photon could provide the energy needed by an individual electron to escape. The photon energy would need to be at least as great as Φ. By this means, Einstein could explain the threshold frequency. A photon of visible light has energy less than Φ, so it cannot release an electron from the surface of zinc.

When a photon arrives at the metal plate, it may be captured by an electron. The electron gains all of the photon's energy and the photon no longer exists. Some of the energy is needed for the electron to escape from the energy well; the rest is the electron's kinetic energy.

Now we can see that the photon model works because it models electromagnetic waves as concentrated 'packets' of energy, each one able to release an electron from the metal.

Here are some rules for the photoelectric effect:

- Electrons from the surface of the metal are removed.
- A single photon can only interact, and hence exchange its energy, with a single electron (one-to-one interaction).
- A surface electron is removed **instantaneously** from the metal surface when the energy of the incident photon is greater than, or equal to, the work function Φ of the metal. (The frequency of the incident radiation is greater than, or equal to, the threshold frequency of the metal. Alternatively, the wavelength of the incident radiation is less than, or equal to, the threshold wavelength of the metal.)
- Energy must be conserved when a photon interacts with an electron.
- Increasing the intensity of the incident radiation does not release a single electron when its frequency is less than the threshold frequency. The intensity of the incident radiation is directly proportional to the **rate** at which photons arrive at the plate. Each photon still has energy that is less than the work function.

Photoelectric experiments showed that the electrons released had a range of kinetic energies up to some maximum value, $k.e._{max}$. These fastest-moving electrons are the ones that were least tightly held in the metal.

Imagine a single photon interacting with a single surface electron and freeing it. According to Einstein:

$$\text{energy of photon} = \text{work function} + \text{maximum kinetic energy of electron}$$
$$hf = \Phi + k.e._{max}$$
$$hf = \Phi + \frac{1}{2}mv_{max}^2$$

where *hf* is the energy of the photon, Φ is the work function of the metal and $\frac{1}{2}mv_{max}^2$ is the maximum kinetic energy of the emitted photoelectron.

This equation, known as **Einstein's photoelectric equation**. It can also be written as:

$$\frac{hc}{\lambda} = \Phi + \frac{1}{2}mv_{max}^2$$

KEY EQUATION

Einstein's photoelectric equation:

$$hf = \Phi + \frac{1}{2}mv_{max}^2 \text{ or } \frac{hc}{\lambda} = \Phi + \frac{1}{2}mv_{max}^2$$

The photoelectric equation can be understood as follows:

- We start with a photon of energy *hf*.
- It is absorbed by an electron.
- Some of the energy (Φ) is used in escaping from the metal. The rest remains as kinetic energy of the electron.
- If the photon is absorbed by an electron that is lower in the energy well, the escaping electron will have less kinetic energy than $k.e._{max}$ (Figure 28.10).

What happens when the incident radiation has a frequency equal to the threshold frequency f_0 of the metal?

The kinetic energy of an electron is zero. Hence, according to Einstein's photoelectric equation:

$$hf_0 = \Phi$$

Hence, the threshold frequency f_0 is given by the expression:

$$f_0 = \frac{\Phi}{h}$$

and, the threshold wavelength γ_0 is given by the expression:

$$\lambda_0 = \frac{hc}{\Phi}$$

What happens when the incident radiation has frequency less than the threshold frequency? A single photon can still give up its energy to a single electron, but this electron cannot escape from the attractive forces of the positive metal ions. The energy absorbed from the photons appears as kinetic energy of the electrons. These electrons lose their kinetic energy to the metal ions when they collide with them. This warms up the metal. This is why a metal plate placed close to a table lamp gets hot.

Different metals have different threshold frequencies, and hence different work functions. For example, alkali metals such as sodium, potassium and rubidium have threshold frequencies in the visible region of the electromagnetic spectrum. The conduction electrons in zinc are more tightly bound within the metal and so its threshold frequency is in the ultraviolet region of the spectrum.

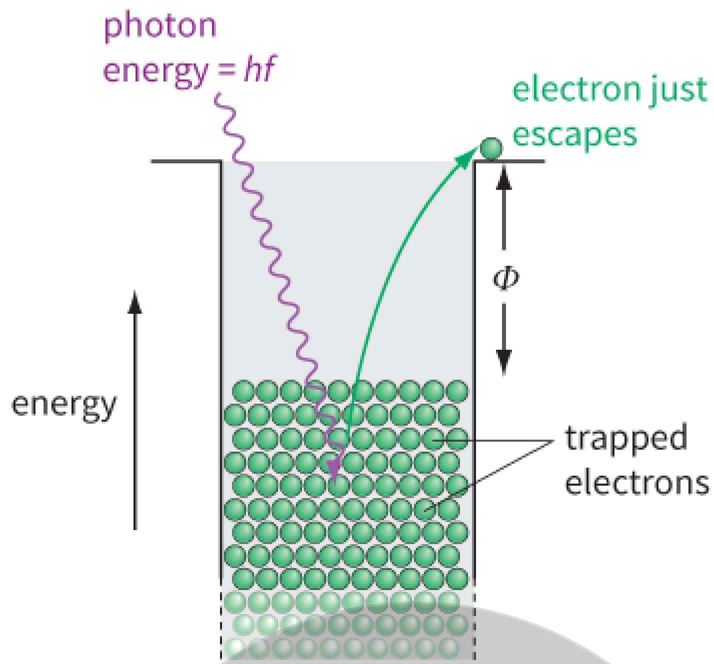


Figure 28.10: A more tightly bound electron needs more energy to release it from the metal.

Table 28.4 summarises the observations of the photoelectric effect.

Observation	**Comments**
Emission of electrons happens as soon as the electromagnetic radiation is incident on the metal.	A single photon interacts with a single electron. If the energy of the incident photon is equal to, or greater than, the work function of the metal, the electrons will be ejected instantaneously.
Even weak (low-intensity) electromagnetic radiation is effective.	Low-intensity means smaller rate of photons incident on the metal surface. The energy of each photon depends on the frequency or wavelength – not the intensity. As long as each photon has energy equal to, or greater than, the work function of the metal, the electrons will be ejected. Low intensity would imply smaller rate of emission of electrons.
Increasing intensity of electromagnetic radiation increases rate at which electrons leave metal.	Greater intensity means greater rate of photons incident on the metal surface. If the electrons are collected as part of an external circuit, then the photoelectric current would be directly proportional to the intensity of the incident radiation – this is provided the threshold frequency of the metal has been exceeded.

Observation	Comments
Increasing intensity has no effect on kinetic energies of electrons.	Greater intensity does not mean more energetic photons, so electrons cannot have more kinetic energy. The maximum kinetic energy of the electrons is given by $k.e._{max} = hf - \Phi$; it is independent of intensity.
A minimum threshold frequency is needed for the emission of electrons.	Electrons will be emitted from the metal surface when the incident radiation has frequency equal to or greater than the threshold frequency.
Increasing frequency of electromagnetic radiation increases maximum kinetic energy of electrons.	Higher frequency means more energetic photons; so electrons gain more kinetic energy and can move faster. Once again, you can use $k.e._{max} = hf - \Phi$ to explain the observation.

Table 28.4: The success of the photon model in explaining the photoelectric effect.

Questions

You will need these values to answer questions 10 to 13:

speed of light in a vacuum $c = 3.00 \times 108$ m s^{-1}

Planck constant $h = 6.63 \times 10^{-34}$ J s

mass of electron $m_e = 9.11 \times 10^{-31}$ kg

elementary charge $e = 1.60 \times 10^{-19}$ C

10 Photons of energies 1.0 eV, 2.0 eV and 3.0 eV strike a metal surface whose work function is 1.8 eV.

- **a** State which of these photons could cause the release of an electron from the metal.
- **b** Calculate the maximum kinetic energies of the electrons released in each case. Give your answers in eV and in J.

11 Table 28.5 shows the work functions of several different metals.

- **a** State which metal requires the highest frequency of electromagnetic waves to release electrons.
- **b** State which metal will release electrons when the lowest frequency of electromagnetic waves is incident on it?
- **c** Calculate the threshold frequency for zinc.
- **d** Calculate the threshold wavelength for potassium.

Metal	Work function Φ / J	Work function Φ / eV
caesium	3.0×10^{-19}	1.9
calcium	4.3×10^{-19}	2.7
gold	7.8×10^{-19}	4.9
potassium	3.2×10^{-19}	2.0
zinc	6.9×10^{-19}	4.3

Table 28.5: Work functions of several different metals.

12 Electromagnetic waves of wavelength 2.4×10^{-7} m are incident on the surface of a metal whose work function is 2.8×10^{-19} J.

a Calculate the energy of a single photon.

b Calculate the maximum kinetic energy of electrons released from the metal.

c Determine the maximum speed of the emitted photoelectrons.

13 When electromagnetic radiation of wavelength 2000 nm is incident on a metal surface, the maximum kinetic energy of the electrons released is found to be 4.0×10^{-20} J.

Calculate the work function of the metal in joules (J).

28.5 Photons have momentum too

Figure 28.11: Comet Hyakutake. The tail of a comet is evidence that photons of sunlight have momentum.

The photoelectric effect provides evidence for the particle-like behaviour of photons. Is there any other evidence for this type of behaviour of electromagnetic radiation? In 1619, German mathematician and astronomer Johann Kepler suggested that the long tail of a comet points away from the Sun because sunlight exerts pressure on this tail. Figure 28.11 shows the tail of the Comet Hyakutake in the night sky.

Kepler was almost correct. In 1905, Albert Einstein, as part of his Special Theory of Relativity, showed that a photon travelling in a vacuum has momentum, even though it has no mass. The steady stream of momentum-carrying photons in sunlight is responsible for exerting pressure (or force) on objects in space. Satellites orbiting the Earth – or space probes sent to explore the planets in our Solar System – have to take account of tiny pressures exerting by colliding photons. A satellite orbiting the Earth would experience a pressure of about 9 μN m^{-2} from sunlight.

Einstein showed that the momentum *p* of a photon is related to its energy *E* by the equation:

$$p = \frac{E}{c}$$

where *c* is the speed of light in a vacuum.

The energy *E* of a photon can be written either as:

$$E = hf \text{ or } E = \frac{hc}{\lambda}$$

Worked example 2 shows how you can **estimate** the pressure exerted by photons hitting a metal plate.

WORKED EXAMPLES

2 A 2.0 mW laser beam is incident normally on a fixed metal plate. The cross-sectional area of the beam is 4.0×10^{-6} m^2. The light from the laser has frequency 4.7×10^{14} Hz.

Calculate the momentum of the photon, and the pressure exerted by the laser beam on the metal plate. You may assume that the photons are all absorbed by the plate.

Step 1 Calculate the momentum of each photon.

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{6.63\times10^{-34}\times4.7\times10^{14}}{3.0\times10^{8}} = 1.04\times10^{-27}\ \text{Ns}$$

(Note: the units can either be written as kg m s^{-1} or N s.)

Step 2 Calculate the number of photons incident on the plate per second.

$$\text{number of photons per second} = \frac{\text{power}}{\text{energy of each proton}} = \frac{0.002}{hf} = \frac{0.002}{6.63\times10^{-34}\times4.7\times10^{14}} = 6.42\times10^{15}\ \text{s}^{-1}$$

Step 3 Calculate the force exerted on the plate by assuming we can use Newton's second law. Consider a time interval of 1.0 s.

$$\begin{aligned}\text{force} &= \text{rate of change of momentum of the photons}\\ &= \text{number of photons per second} \times \text{momentum of each photon}\\ &= 6.42\times10^{15}\times1.04\times10^{-27}\\ &= 6.68\times10^{-12}\ \text{N}\end{aligned}$$

Step 4 Calculate the pressure.

$$\text{pressure} = \frac{\text{force}}{\text{area}} = \frac{6.68\times10^{-12}}{4.0\times10^{-6}} = 1.7\times10^{-6}\ \text{Pa}$$

This is a tiny pressure and would not be noticeable on the fixed metal plate. However, if this plate was in deep-space, it would, over a period of time, show some movement.

28.6 Line spectra

We will now look at another phenomenon that we can explain in terms of light as photons. We rely a great deal on light to inform us about our surroundings. Using our eyes we can identify many different colours. Scientists take this further by analysing light, by splitting it up into a spectrum. The technical term for the splitting of light into its components is **dispersion**. You will be familiar with the ways in which this can be done, using a prism or a diffraction grating (Figure 28.12).

The spectrum of white light shows that it consists of a range of wavelengths, from about 400 nm (violet) to about 700 nm (red), as in Figure 28.13. This is a **continuous spectrum**.

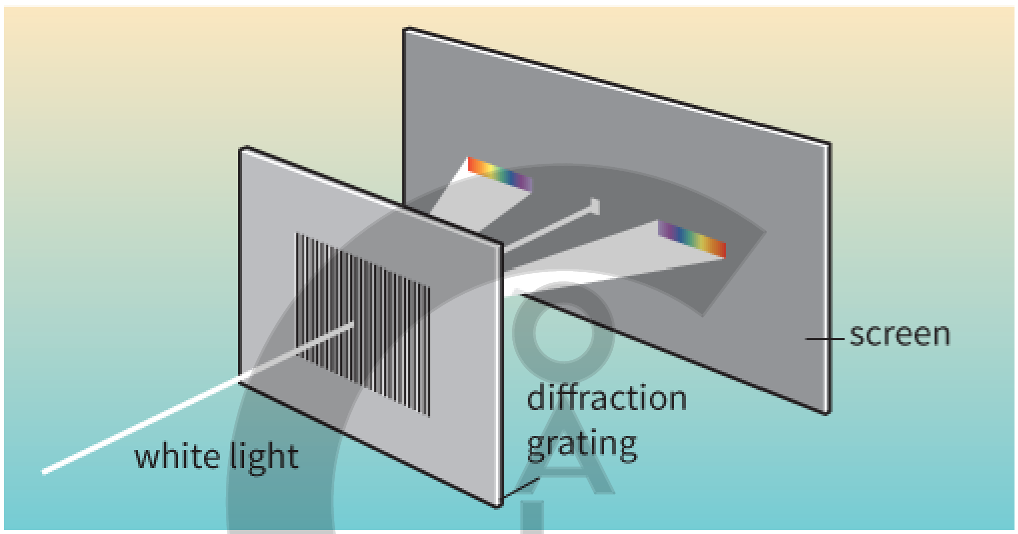


Figure 28.12: White light is split up into a continuous spectrum when it passes through a diffraction grating.

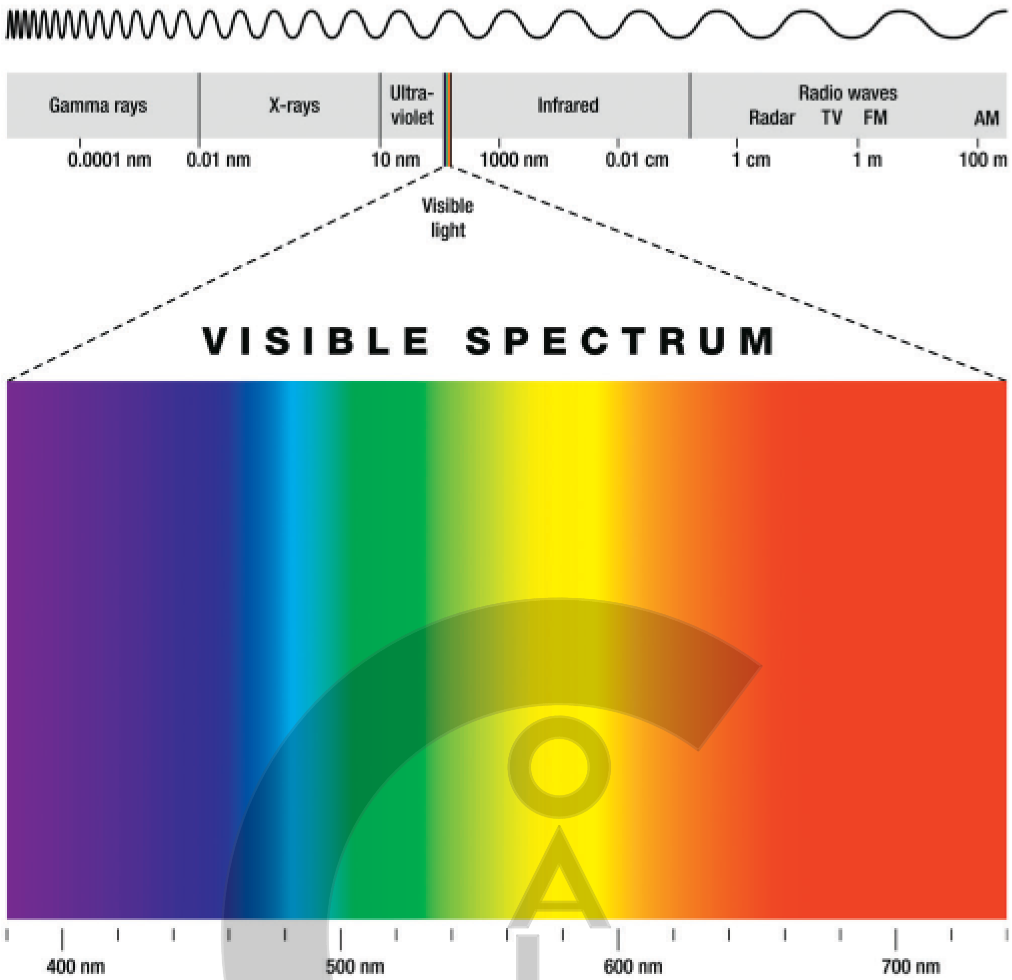


Figure 28.13: Spectra of white light.

It is more interesting to look at the spectrum from a hot gas. If you look at a lamp that contains a gas such as neon or sodium, you will see that only certain colours are present. Each colour has a unique wavelength. If the source is narrow and it is viewed through a diffraction grating, a **line spectrum** is seen.

Figure 28.14a–c show the line spectra of hot gases of the elements mercury, helium and cadmium. Each element has a spectrum with a unique collection of wavelengths. Line spectra can therefore be used to identify elements. This is exactly what the British astronomer William Huggins did when he deduced which elements are the most common in the stars.

Figure 28.14: Spectra of light from **a** mercury, **b** helium and **c** cadmium vapour.

These line spectra, which show the composition of light emitted by hot gases, are called **emission line spectra**.

There is another kind of spectra, called **absorption line spectra**, which are observed when white light is passed through cool gases. After the light has passed through a diffraction grating, the continuous white light spectrum is found to have black lines across it (Figure 28.15). Certain wavelengths have been absorbed as the white light passed through the cool gas.

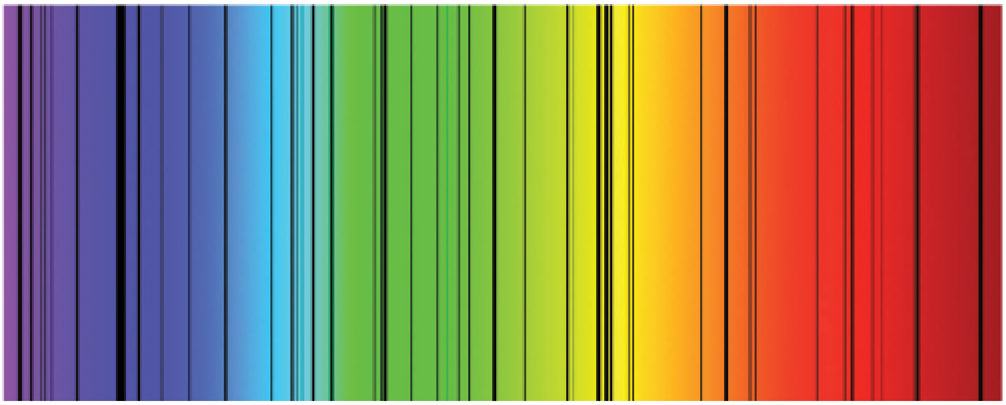


Figure 28.15: An absorption line spectrum formed when white light is passed through cool mercury vapour.

Absorption line spectra are found when the light from stars is analysed. The interior of the star is very hot and emits white light of all wavelengths in the visible range. However, this light has to pass through the **cooler** outer layers of the star. As a result, certain wavelengths are absorbed. Figure 28.16 shows the spectrum for the Sun.

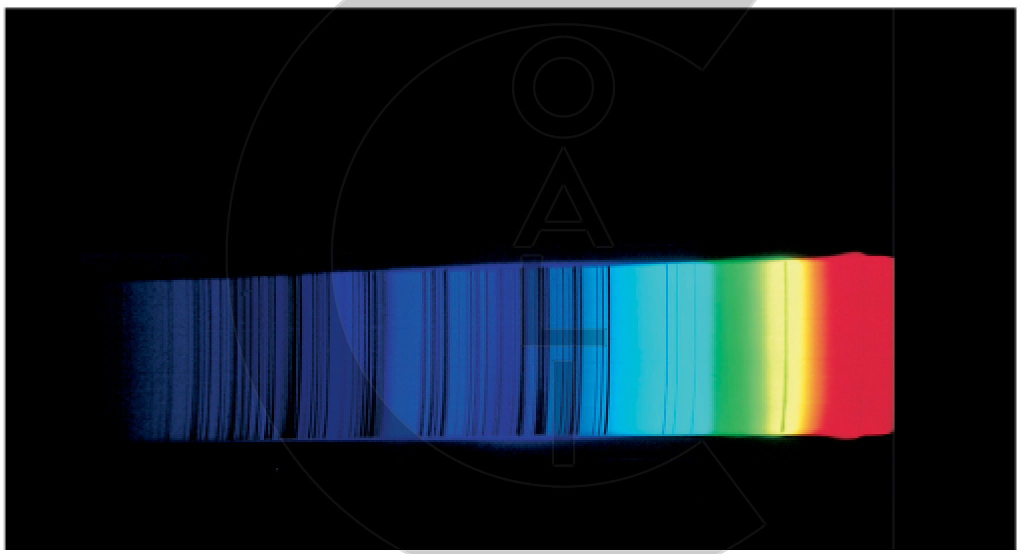


Figure 28.16: The Sun's spectrum shows dark lines. These dark lines arise when light of specific wavelengths coming from the Sun's hot interior is absorbed by its cooler atmosphere.

28.7 Explaining the origin of line spectra

From the description in the previous topic, we can see that the atoms of a given element (e.g., helium) can only emit or absorb light of certain wavelengths.

Different elements emit and absorb different wavelengths. How can this be? To understand this, we need to establish two points:

- First, as with the photoelectric effect, we are dealing with light (an electromagnetic wave) interacting with matter. Hence, we need to consider light as consisting of photons. For light of a single wavelength λ and frequency f, the energy E of each photon is given by the equation:

$$E = hf \text{ or } E = \frac{hc}{\lambda}$$

- Second, when light interacts with matter, it is the electrons that absorb the energy from the incoming photons. When the electrons lose energy, light is emitted by matter in the form of photons.

What does the appearance of the line spectra tell us about electrons in atoms? They can only absorb, or emit, photons of certain energies. From this we deduce that electrons in atoms can themselves only have certain fixed values of energy. This idea seemed very odd to scientists a hundred years ago. Figure 28.17 shows a diagram of the permitted **energy levels** (or **energy states**) of the electron of a hydrogen atom. An electron in a hydrogen atom can have only one of these values of energy. It cannot have an energy that is between these energy levels.

The energy levels of the electron are similar to the rungs of a ladder. The energy levels have **negative** values because external energy has to be supplied to remove an electron from the atom. The negative energy shows that the electron is trapped within the atom by the attractive forces of the atomic nucleus. An electron with zero energy is free from the atom.

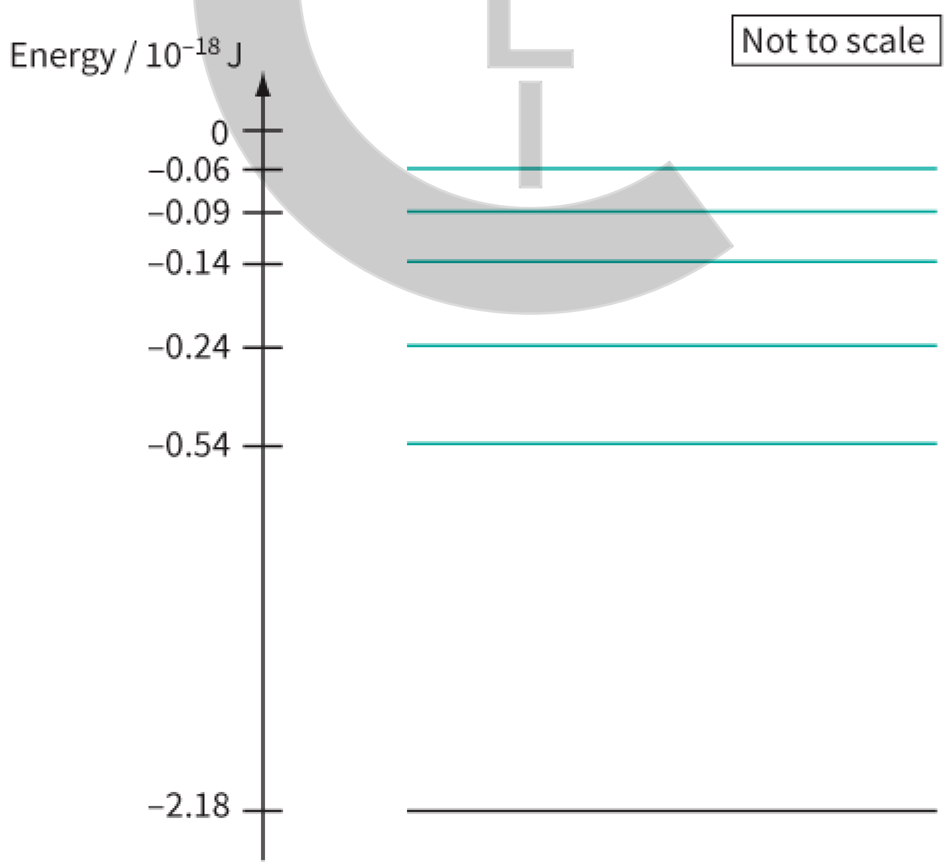


Figure 28.17: Some of the energy levels of the hydrogen atom.

The energy of the electron in the atom is said to be **quantised**. This is one of the most important statements of quantum physics.

Now we can explain what happens when an atom emits light. One of its electrons falls from a high energy level to a lower one (Figure 28.18a). The electron makes a **transition** to a lower energy level. It is important to note that, even on this microscopic scale, energy must be conserved. The loss of energy of the electron leads to the emission of a single photon of light. The one-to-one interaction rule of quantum physics is paramount – a single electron is responsible for producing a single photon. The energy of this photon is exactly equal to the energy difference between the two energy levels. If the electron makes a transition from a higher energy level, the energy loss of the electron is larger and this leads to the emission of a more energetic photon. The distinctive energy levels of an atom mean that the energy of the photons emitted, and hence the wavelengths emitted, will be unique to that atom. This explains why only certain wavelengths are present in the emission line spectrum of a hot gas.

Atoms of different elements have different line spectra because they have different spacings between their energy levels. It is not within the scope of this book to discuss why this is.

Similarly, we can explain the origin of absorption line spectra. White light consists of photons of many different energies. For a photon to be absorbed, it must have exactly the right energy to lift an electron from one energy level to another higher energy level (Figure 28.18b). This 'excited' electron, at the higher energy level, will eventually make a transition to a lower energy level – but this time, the photon will be re-emitted in any direction, and not necessarily in the original direction of the white light. This leads to lower intensity for photons of a specific wavelength. White light photons with energy not matching the difference between the energy levels will carry on moving in the original direction. The net result of all this is a dark absorption line seen against the background of a continuous spectrum.

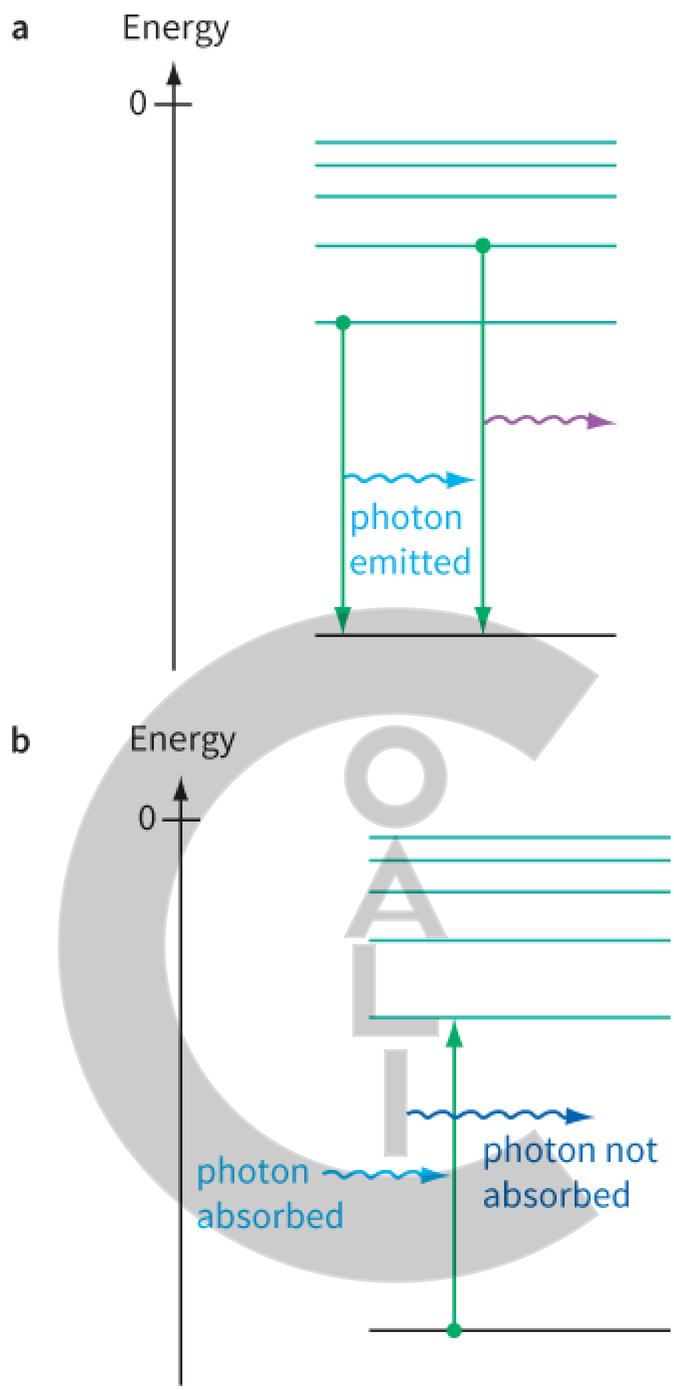


Figure 28.18: **a** When an electron drops to a lower energy level, it emits a single photon. **b** A photon must have just the right energy if it is to be absorbed by an electron.

28.8 Photon energies

When an electron changes its energy from one level E_1 to another E_2, it either emits or absorbs a **single** photon. The energy of the photon *hf* is simply equal to the **difference** in energies between the two levels:

$$\begin{aligned} \text{photon energy} &= \Delta E \\ hf &= E_1 - E_2 \end{aligned}$$

or

$$\frac{hc}{\lambda} = E_1 - E_2$$

Referring back to the energy level diagram for hydrogen (Figure 28.17), you can see that, if an electron falls from the second level to the lowest energy level (known as the **ground state**), it will emit a photon of energy:

$$\begin{aligned} \text{photon energy} &= \Delta E \\ hf &= ((-0.54) - (-2.18)) \times 10^{-18}\ \text{J} \\ &= 1.64 \times 10^{-18}\ \text{J} \end{aligned}$$

We can calculate the frequency *f* and wavelength λ of the emitted electromagnetic radiation.

The frequency is:

$$\begin{aligned} f &= \frac{E}{h} \\ &= \frac{1.64 \times 10^{-18}}{6.63 \times 10^{-34}} \\ &= 2.47 \times 10^{15}\ \text{Hz} \end{aligned}$$

The wavelength is:

$$\begin{aligned} \lambda &= \frac{c}{f} \\ &= \frac{3.00 \times 10^{8}}{2.47 \times 10^{15}} \\ &= 1.21 \times 10^{-7}\ \text{m} \approx 121\ \text{nm} \end{aligned}$$

This is a wavelength in the ultraviolet region of the electromagnetic spectrum.

KEY EQUATIONS

The energy of a photon, absorbed or emitted, as a result of an electron making a transition between two energy levels E_1 and E_2:

$$hf = E_1 - E_2$$

$$\frac{hc}{\lambda} = E_1 - E_2$$

Questions

14 Figure 28.19 shows part of the energy level diagram for the electrons in an imaginary atom. The arrows represent three transitions between the energy levels. For each of these transitions:

a calculate the energy of the photon

b calculate the frequency and wavelength of the electromagnetic radiation (emitted or absorbed)

c state whether the transition contributes to an emission line in the spectrum or an absorption line in the spectrum.

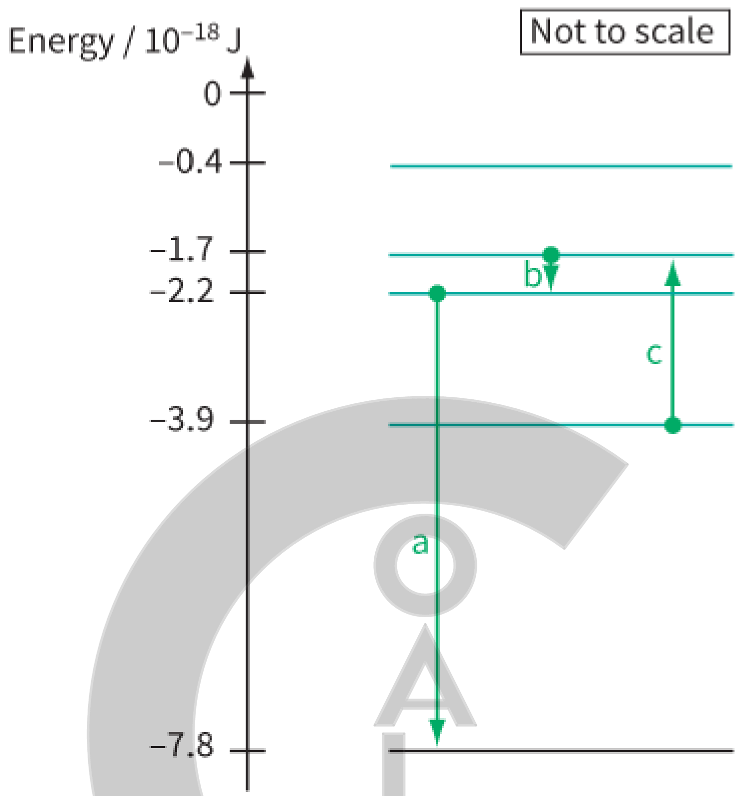


Figure 28.19: Electron energy level diagram, showing three electron transitions a, b and c. For Question 14.

15 Figure 28.20 shows another energy level diagram. In this case, energy is given in electronvolts (eV). The list shows the energies of some photons:

6.0 eV 9.0 eV 11 eV 20 eV 25 eV 34 eV 45 eV

State and explain which of these photons will be absorbed by the electrons.

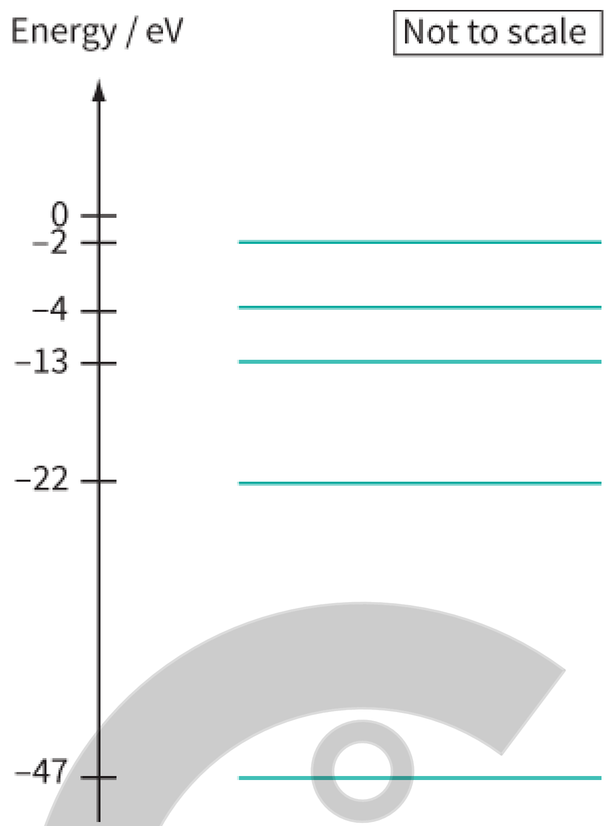


Figure 28.20: An energy level diagram. For Question 15.

16 The line spectrum for a particular type of atom is found to include the following wavelengths:

83 nm 50 nm 25 nm

a Calculate the corresponding photon energies in eV.

b Sketch the energy levels that could give rise to these photons. On the diagram, indicate the corresponding electron transitions responsible for these three spectral lines.

28.9 The nature of light: waves or particles?

It is clear that, in order to explain the photoelectric effect, we must use the idea of light (and all electromagnetic radiation) as particles. Similarly, photons explain the appearance of line spectra. However, to explain diffraction, interference and polarisation of light, we must use the wave model. How can we sort out this dilemma?

We have to conclude that sometimes light shows wave-like behaviour; at other times it behaves as particles (photons). In particular, when light is absorbed by a metal surface, it behaves as particles. Individual photons are absorbed by individual electrons in the metal. In a similar way, when a Geiger counter detects γ-radiation, we hear individual γ-photons being absorbed in the tube.

So what is light? Is it a wave or a particle? Physicists have come to terms with the dual nature of light. This duality is referred to as the **wave–particle duality** of light. In simple terms:

- Light interacts with matter (e.g., electrons) as a particle – the photon. The evidence for this is provided by the photoelectric effect.
- Light propagates through space as a wave. The evidence for this comes from the diffraction and interference of light using slits.

28.10 Electron waves

Light has a dual nature. (In fact, it is not only light, but all electromagnetic waves that have this dual nature.) Is it possible that particles such as electrons also have a dual nature? This interesting question was first considered by Louis de Broglie (pronounced 'de Broy') in 1924 (Figure 28.21).

Figure 28.21: Louis de Broglie provided an alternative view of how particles behave.

De Broglie imagined that electrons would travel through space as a wave. He proposed that the wave-like property of a particle like the electron can be represented by its wavelength λ, which is related to its momentum p of the particle by the equation:

$$\lambda = \frac{h}{p}$$

where h is the Planck constant. The wavelength λ is often referred to as the **de Broglie wavelength**. The waves associated with the electron are referred to as matter waves.

The momentum p of a particle is the product of its mass m and its velocity v. Therefore, the de Broglie equation may also be written as:

$$\lambda = \frac{h}{mv}$$

The Planck constant h is the same constant that appears in the equation $E = hf$ for the energy of a photon. It is fascinating how the Planck constant h is tangled with the behaviour of both matter as waves (e.g. electrons) and electromagnetic waves as 'particles' (photons).

The wave property of the electron was eventually confirmed in 1927 by researchers in America and in England. The Americans, Clinton Davisson and Edmund Germer, showed experimentally that electrons were diffracted by crystals of nickel. The diffraction of electrons confirmed their wave-like property. In England, George Thomson fired electrons into thin sheets of metal in a vacuum tube. He, too, provided evidence that electrons were diffracted by the metal atoms.

Louis de Broglie received the 1929 Nobel Prize in Physics. Clinton Davisson and George Thomson shared the Nobel Prize in Physics in 1937.

Electron diffraction

We can reproduce the same diffraction results in the laboratory using an electron diffraction tube (Figure 28.22).

In an electron diffraction tube, the electrons from the heated filament are accelerated to high speeds by the large potential difference between the negative heater (cathode) and the positive electrode (anode). A beam of electrons passes through a thin sample of polycrystalline graphite. It is made up of many tiny crystals, each of which consists of large numbers of carbon atoms arranged in uniform atomic layers. The electrons emerge from the graphite film and produce diffraction rings on the phosphor screen. The diffraction rings are similar to those produced by light (a wave) passing through a small circular hole. The rings cannot be explained if electrons behaved as particles. Diffraction is a property of waves. Hence, the rings can only be explained if the electrons travel through the graphite film as a wave. The electrons are diffracted by the individual carbon atoms and the spacing between the layers of carbon atoms. The atomic layers of carbon behave like a diffraction grating with many slits. The electrons show diffraction effects because their de Broglie wavelength λ is similar to the spacing between the atomic layers.

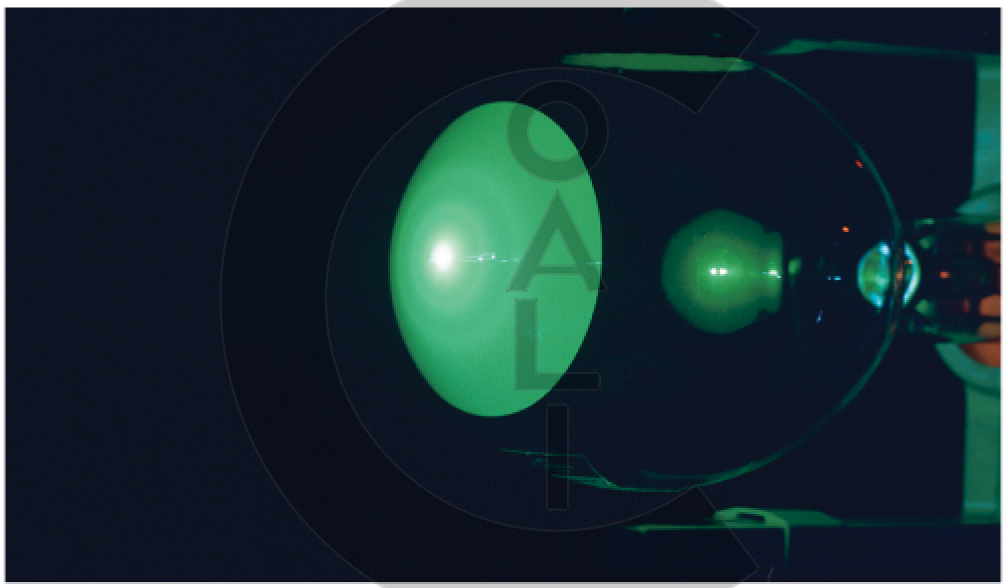


Figure 28.22: When a beam of electrons passes through a graphite film, as in this vacuum tube, a diffraction pattern is produced on the phosphor screen.

This experiment shows that electrons appear to travel as waves. If we look a little more closely at the results of the experiment, we find something even more surprising. The phosphor screen gives a flash of light for each electron that hits it. These flashes build up to give the diffraction pattern (Figure 28.23). But if we see flashes at particular points on the screen, are we not seeing individual electrons – in other words, are we not observing particles?

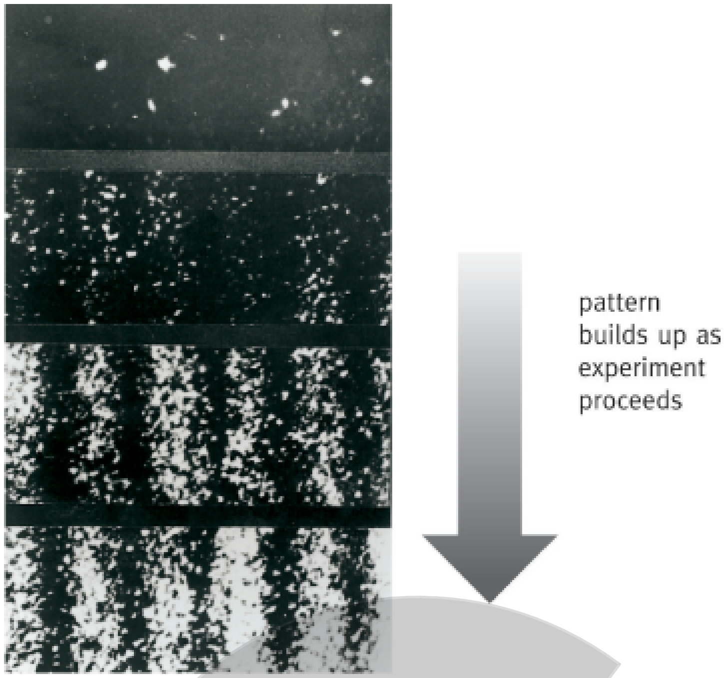


Figure 28.23: The speckled diffraction pattern shows that it arises from many individual electrons striking the screen.

PRACTICAL ACTIVITY 28.3

Investigating electron diffraction

If you have access to an electron diffraction tube (Figure 28.24), you can see for yourself how a beam of electrons is diffracted. The electron gun at one end of the tube produces a beam of electrons. By changing the voltage between the anode and the cathode, you can change the energy of the electrons, and hence their speed. The beam strikes a graphite target, and a diffraction pattern appears on the screen at the other end of the tube.

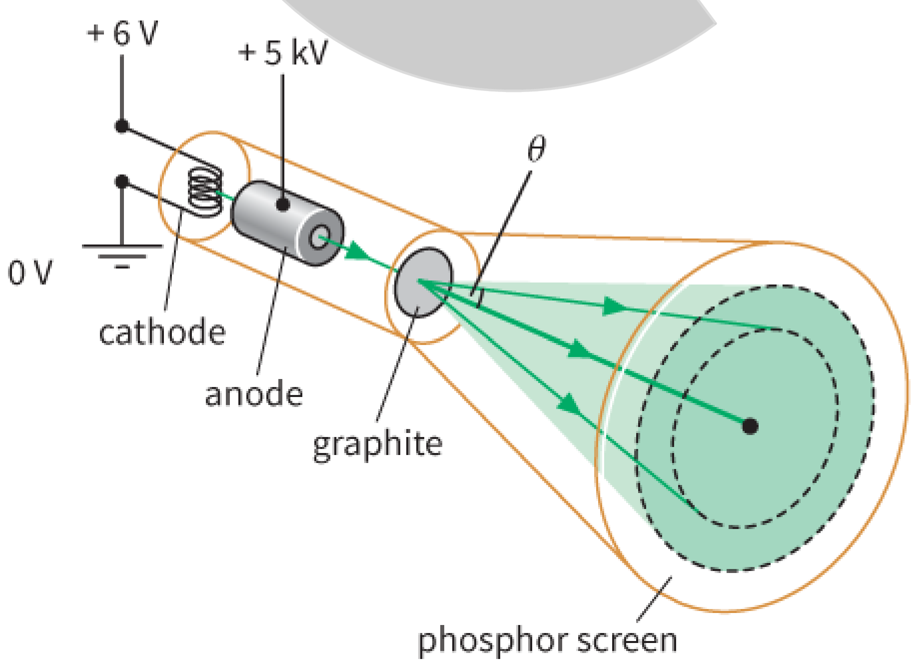


Figure 28.24: Electrons are accelerated from the cathode to the anode; they form a beam that is diffracted as it passes through the graphite film.

You can use an electron diffraction tube to investigate how the wavelength of the electrons depends on their speed. Qualitatively, you should find that increasing the anode–cathode voltage makes the pattern of diffraction rings shrink. The electrons have more kinetic energy (they are faster); the shrinking pattern shows that their wavelength has decreased. You can find the wavelength λ of the electrons by measuring the angle θ at which they are diffracted:

$$\lambda = 2d\sin\theta$$

where d is the spacing of the atomic layers of graphite.

You can find the speed of the electrons from the anode–cathode voltage V:

$$\tfrac{1}{2}mv^2 = eV$$

WORKED EXAMPLE

3 Calculate the de Broglie wavelength of an electron travelling through space at a speed of 1.0×10^7 m s^{-1}. State whether or not these electrons can be diffracted by solid materials. (Atomic spacing in solid materials ~ 10^{-10} m.)

Step 1 According to the de Broglie equation, we have:

$$\lambda = \frac{h}{mv}$$

Step 2 The mass of an electron is 9.11×10^{-31} kg. Hence:

$$\lambda = \frac{6.63\times10^{-34}}{9.11\times10^{-31}\times1.0\times10^{7}}$$

$$= 7.3\times10^{-11}\text{ m}$$

Electrons travelling at 10^7 m s^{-1} have a de Broglie wavelength of order of magnitude 10^{-10} m – this is comparable the atomic spacing. Hence, the electrons can be diffracted by matter.

Question

17 X-rays are used to find out about the spacings of atomic planes in crystalline materials.

a Describe how beams of electrons could be used for the same purpose.

b How might electron diffraction be used to identify a sample of a metal?

People waves

The de Broglie equation applies to all matter; anything that has mass. It can also be applied to objects like golf balls and people!

Imagine a 65 kg person running at a speed of 3.0 m s^{-1} through an opening of width 0.80 m. According to the de Broglie equation, the wavelength of this person is:

$$\lambda = \frac{h}{mv} = \frac{6.63\times10^{-34}}{65\times3.0} = 3.4\times10^{-36}\ \text{m}$$

This wavelength is very small indeed compared with the size of the gap, hence no diffraction effects would be observed. People cannot be diffracted through everyday gaps. The de Broglie wavelength of this person is much smaller than any gap the person is likely to try to squeeze through! For this reason, we do not use the wave model to describe the behaviour of people; we get much better results by regarding people as large particles.

Question

18 A beam of electrons is accelerated from rest through a p.d. of 1.0 kV.

- **a** What is the energy (in eV) of each electron in the beam?
- **b** Calculate the speed, and hence the momentum (*mv*), of each electron.
- **c** Calculate the de Broglie wavelength of each electron.
- **d** Would you expect the beam to be significantly diffracted by a metal film in which the atoms are separated by a spacing of 0.25×10^{-9} m?

Probing matter

All moving particles have a de Broglie wavelength. The structure of matter can be investigated using the diffraction of particles. Diffraction of slow-moving neutrons (known as thermal neutrons) from nuclear reactors is used to study the arrangements of atoms in metals and other materials. The wavelength of these neutrons is about 10^{-10} m, which is roughly the separation between the atoms.

Diffraction of slow-moving electrons is used to explore the arrangements of atoms in metals (Figure 28.25) and the structures of complex molecules such as DNA (Figure 28.26). It is possible to accelerate electrons to the right speed so that their wavelength is similar to the spacing between atoms, around 10^{-10} m.

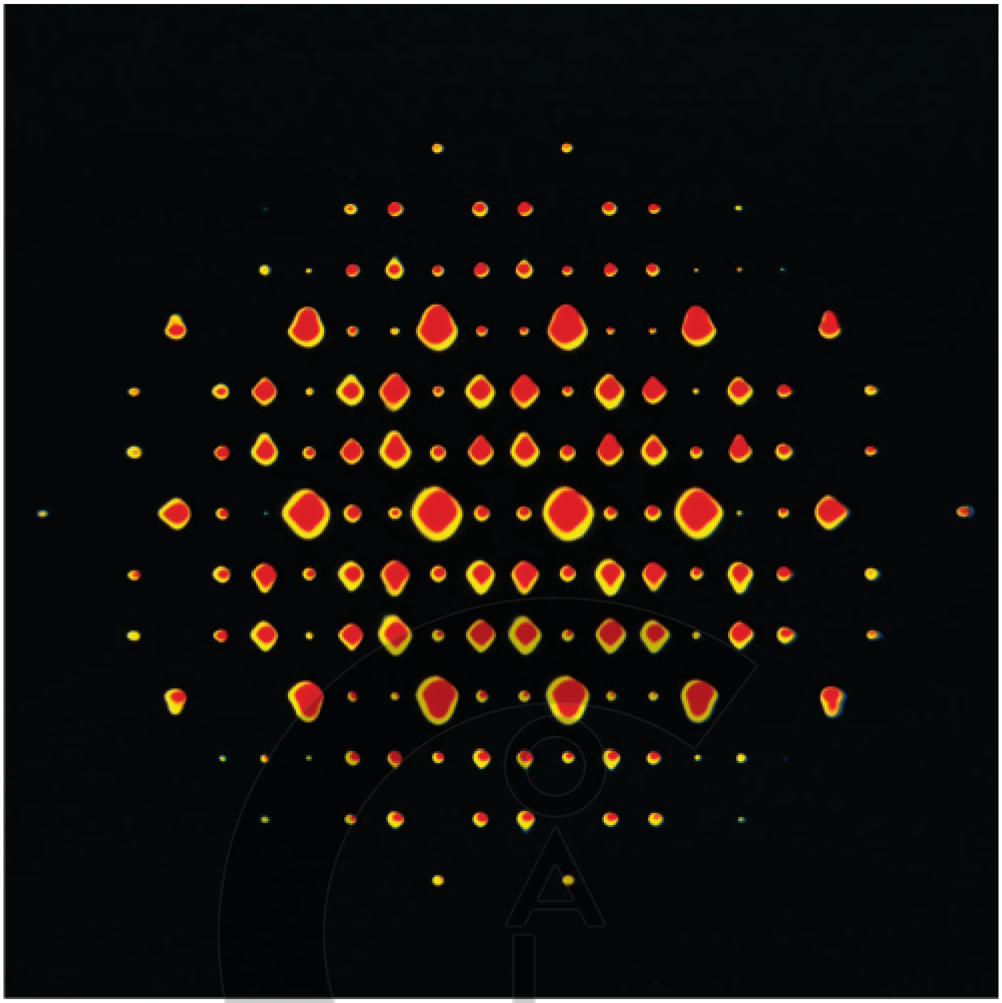


Figure 28.25: Electron diffraction pattern for an alloy of titanium and nickel. From this pattern, we can deduce the arrangement of the atoms and their separations.

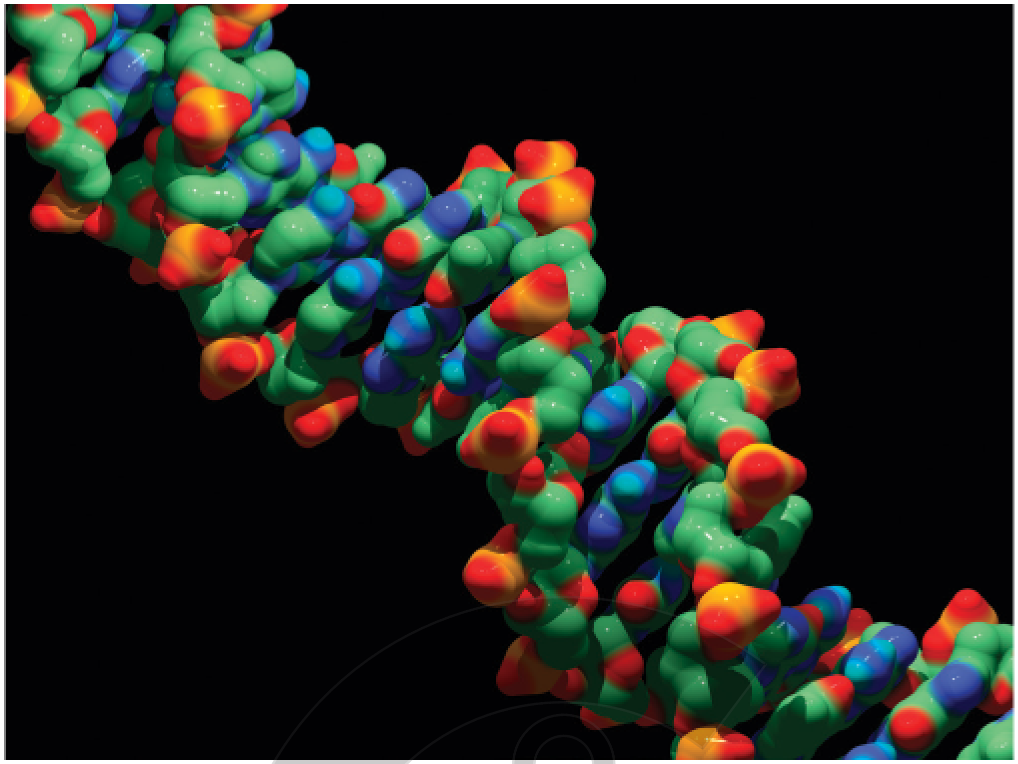


Figure 28.26: The structure of the giant molecule DNA, deduced from electron diffraction.

High-speed electrons from particle accelerators have been used to determine the diameter of atomic nuclei. This is possible because high-speed electrons have wavelengths of order of magnitude 10^{-15} m. This wavelength is similar to the size of atomic nuclei. Electrons travelling close to the speed of light are being used to investigate the internal structure of the nucleus. These electrons have to be accelerated by voltages up to 10^{9} V.

The nature of the electron: wave or particle?

The electron has a dual nature, just like electromagnetic waves. This duality is referred to as the **wave–particle duality** of the electron. In simple terms:

- An electron interacts with matter as a particle. The evidence for this is provided by Newtonian mechanics.
- An electron travels through space as a wave. The evidence for this comes from the diffraction of electrons.

28.11 Revisiting photons

It is worth finishing this topic on quantum physics by further examining the photon. A photon has momentum *p* and energy *E*. The two key equations for a photon are:

$$p = \frac{E}{c} \text{ and } E = \frac{hc}{\lambda}$$

Therefore,

$$\begin{aligned} p &= \frac{E}{c} \\ &= \frac{hc}{\lambda c} \\ &= \frac{h}{\lambda} \end{aligned}$$

This equation is identical to the de Broglie equation for momentum of particle and its wavelength. So, it appears that the equation can be used for the particle-like (photon) behaviour of electromagnetic radiation and the wave-like behaviour of particles. The de Broglie equation is an intriguing equation of quantum physics.

REFLECTION

Without looking at your textbook, summarise all the key equations containing the Planck constant h, and the key points of the photelectric effect.

Compare your summary with a fellow learner. Did you miss out any key ideas?

If you were the teacher, what comments would you make about your summary?

SUMMARY

For electromagnetic waves of frequency f and wavelength λ, each photon has energy E given by:

$$E = hf \quad \text{or} \quad E = \frac{hc}{\lambda}$$

where h is the Planck constant.

One electronvolt is the energy transferred when an electron travels through a potential difference of 1 V:

$1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$

A particle of charge e accelerated through a voltage V has kinetic energy given by:

$$eV = \tfrac{1}{2}mv^2$$

The photoelectric effect is an example of a phenomenon explained in terms of the particle-like (photon) behaviour of electromagnetic radiation.

Einstein's photoelectric equation is:

$$hf = \Phi + \tfrac{1}{2}mv_{max}{}^2 \quad \text{or} \quad \frac{hc}{\lambda} = \Phi + \tfrac{1}{2}mv_{max}{}^2$$

where Φ = work function of the metal.

The threshold frequency is the **minimum** frequency of the incident electromagnetic radiation that will release an electron from the metal surface.

The threshold wavelength is the **longest** wavelength of the incident electromagnetic radiation that will release an electron from the metal surface.

Electron diffraction is an example of a phenomenon explained in terms of the wave-like behaviour of matter.

The de Broglie wavelength λ of a particle is related to its momentum p by the de Broglie equation:

$$\lambda = \frac{h}{p}$$

where p = momentum of the particle = mv.

Both electromagnetic radiation (such as light) and matter (such as electrons) exhibit wave–particle duality; that is, they show both wave-like and particle-like behaviours, depending on the circumstances. In wave–particle duality:

- **interaction** is explained in terms of **particles**
- **propagation** through space is explained in terms of **waves**.

Photons have no mass, but they have momentum. The momentum p of a photon of energy E is given by the equation:

$$p = \frac{E}{c}$$

Line spectra arise for isolated atoms (the electrical forces between such atoms is negligible).

The energy of an electron in an isolated atom is quantised. The electron is allowed to exist in specific energy states known as energy levels.

An electron loses energy when it makes a transition from a higher energy level to a lower energy level. A photon of electromagnetic radiation is emitted because of this energy loss. The result is an emission line spectrum.

Absorption line spectra arise when a photon of electromagnetic radiation is absorbed by electrons in isolated atoms. An electron absorbs a photon of the correct energy to allow it to make a transition to a higher energy level.
The frequency *f* and the wavelength λ of the emitted or absorbed radiation are related to the energy levels E_1 and E_2 by the equations: $$hf = E_1 - E_2 \quad \text{and} \quad \frac{hc}{\lambda} = E_1 - E_2$$

EXAM-STYLE QUESTIONS

1 In which of the following can you use the term work function in your explanation? **[1]**

- **a** Diffraction of electrons by graphite
- **b** Interference of light from a diffraction grating
- **c** Photoelectric effect
- **d** Reflection of light

2 A researcher is carrying out an experiment on the photoelectric effect. Electromagnetic radiation of different frequencies is incident on a metal and the maximum kinetic energy of the emitted electrons is determined.

The researcher plots a straight-line graph of maximum kinetic energy of the electrons $k.e_{max}$ against the frequency f of the radiation.

Which row is correct? **[1]**

	Gradient of graph	y-intercept of graph
A	the Planck constant	work function of metal
B	threshold frequency	the Planck constant
C	threshold wavelength	threshold frequency
D	work function of metal	threshold wavelength

Table 28.6

3 Calculate the energy of a photon of frequency 4.0×10^{18} Hz. **[2]**

4 The microwave region of the electromagnetic spectrum is considered to have wavelengths ranging from 5 mm to 50 cm. Calculate the range of energy of microwave photons. **[3]**

5 In a microwave oven, the photons are used to warm food. Each photon has energy 1.02×10^{-5} eV.

- **a** Calculate the energy of each photon in joule (J). **[1]**
- **b** Calculate the frequency of the photons. **[1]**
- **c** Calculate the wavelength of the photons. **[1]**

[Total: 3]

6 **a** Alpha-particles of energy 5.0 MeV are emitted in the radioactive decay of radium. Express this energy in joules. **[1]**

- **b** Electrons in an cathode-ray tube are accelerated through a potential difference of 10 kV. Calculate their energy:
  - **i** in electronvolts **[1]**
  - **ii** in joules. **[1]**
- **c** In a nuclear reactor, neutrons are slowed to energies of 6×10^{-21} J. Calculates this in eV. **[1]**

[Total: 4]

7 A helium nucleus (charge = $+3.2 \times 10^{-19}$ C; mass = 6.8×10^{-27} kg) is accelerated through a potential difference of 7500 V.

Calculate:

a its kinetic energy in electronvolts **[2]**

b its kinetic energy in joules **[1]**

c its speed. **[2]**

[Total: 5]

8 Ultraviolet light with photons of energy 2.5×10^{-18} J is shone onto a zinc plate. The work function of zinc is 4.3 eV.

Calculate the maximum energy with which an electron can be emitted from the zinc plate. Give your answer in:

a eV **[3]**

b J. **[1]**

[Total: 4]

9 Calculate the minimum frequency of electromagnetic radiation that will cause the emission of photoelectrons from the surface of gold. **[2]**

(Work function for gold = 4.9 eV.)

10 The diagram shows five of the energy levels in a helium ion. The lowest energy level is known as the ground state.

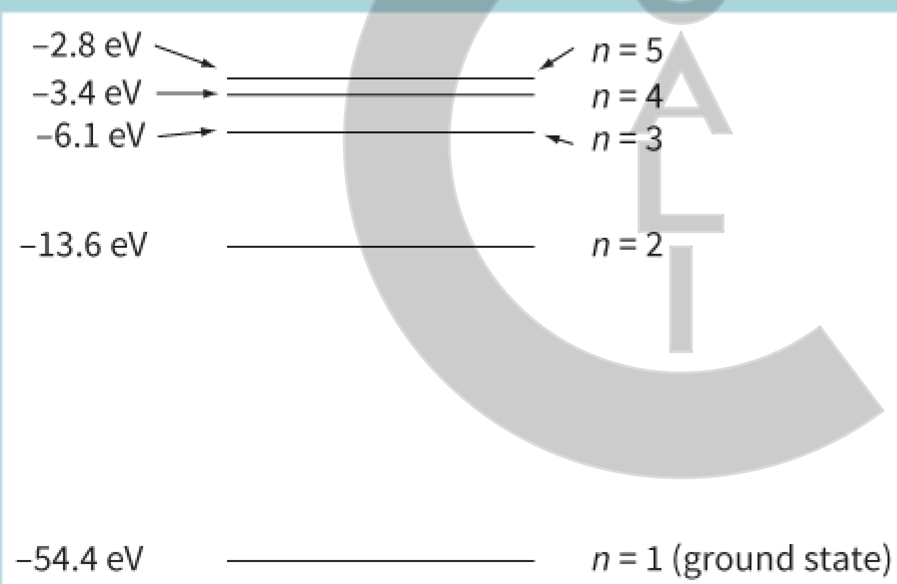


Figure 28.27

a Determine the energy, in joules, that is required to completely remove the remaining electron, originally in its ground state, from the helium ion. **[2]**

b Determine the frequency of the radiation that is emitted when the electron drops from the level $n = 3$ to $n = 2$. State the region of the electromagnetic spectrum in which this radiation lies. **[3]**

c Without further calculation, describe qualitatively how the frequency of the radiation emitted when the electron drops from the level $n = 2$ to $n = 1$ compares with the energy of the radiation emitted when it drops from $n = 3$ to $n = 2$. **[2]**

[Total: 7]

11 The spectrum of sunlight has dark lines. These dark lines are due to the absorption of certain wavelengths by the cooler gases in the atmosphere of the Sun.

a One particular dark spectral line has a wavelength of 590 nm. Calculate the energy of a photon with this wavelength. **[2]**

b The diagram shows some of the energy levels of an isolated atom of helium.

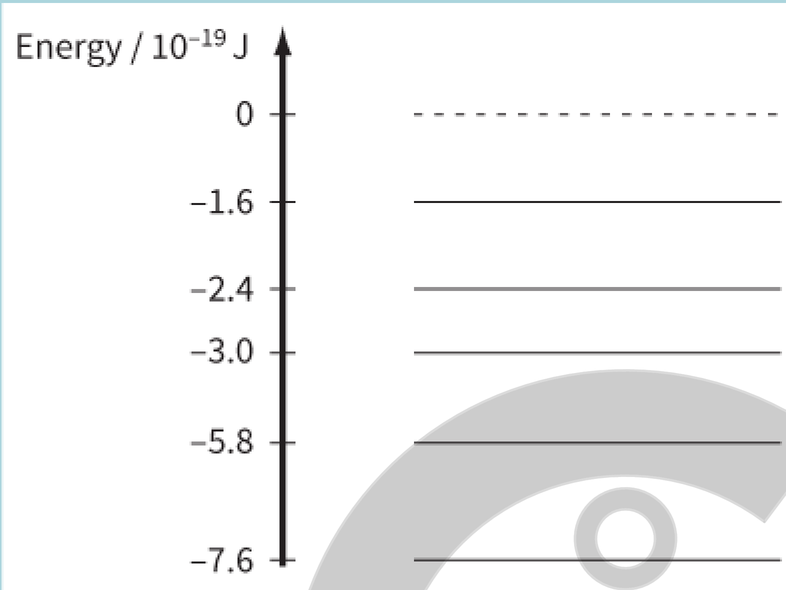


Figure 28.28

i Explain the significance of the energy levels having negative values. **[1]**

ii Explain, with reference to the energy level diagram, how a dark line in the spectrum may be due to the presence of helium in the atmosphere of the Sun. **[2]**

iii All the light absorbed by the atoms in the Sun's atmosphere is re-emitted. Suggest why a dark spectral line of wavelength of 590 nm is still observed from the Earth. **[1]**

[Total: 6]

12 The diagram shows three of the energy levels in an isolated hydrogen atom. The lowest energy level is known as the ground state.

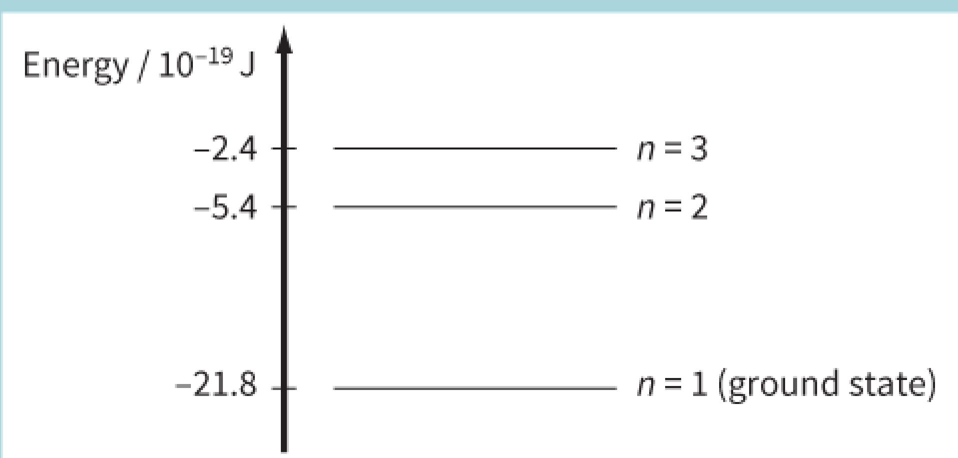


Figure 28.29

a Explain what happens to an electron in the ground state when it absorbs the energy from a photon energy 21.8×10^{-19} J. **[1]**

b **i** Explain why a photon is emitted when an electron makes a transition between energy levels $n = 3$ and $n = 2$. **[2]**

ii Calculate the wavelength of electromagnetic radiation emitted when an electron makes a jump between energy levels $n = 3$ and $n = 2$. **[3]**

iii In the diagram, each energy level is labelled with its 'principal quantum number' n. Use the energy level diagram to show that the energy E of an energy level is inversely proportional to n^2. **[4]**

[Total: 10]

13 a i Explain what is meant by the wave–particle duality of electromagnetic radiation. **[2]**

ii Explain how the photoelectric effect gives evidence for this phenomenon. **[2]**

The diagram shows the maximum kinetic energy E of the emitted photoelectrons as the frequency f of the incident radiation on a sodium plate is varied.

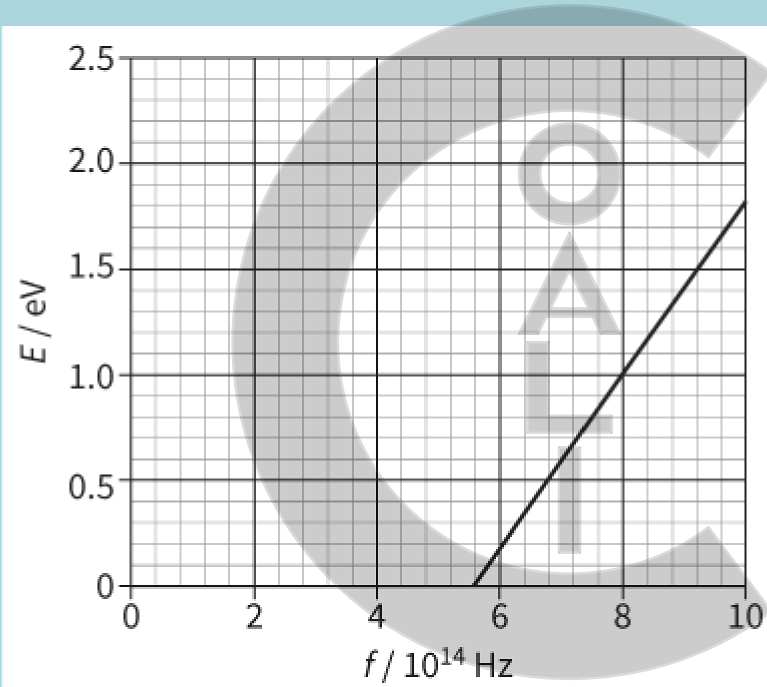


Figure 28.30

b Explain why there are no photoelectrons emitted when the frequency of the incident light is less than 5.6×10^{14} Hz. **[2]**

c Determine the work function for sodium. Explain your answer. **[3]**

d Use the graph to determine the value of the Planck constant. Explain your answer. **[3]**

[Total: 12]

14 a State what is meant by the **de Broglie wavelength** of an electron. **[2]**

b The diagram shows the principles of an electron tube used to demonstrate electron diffraction.

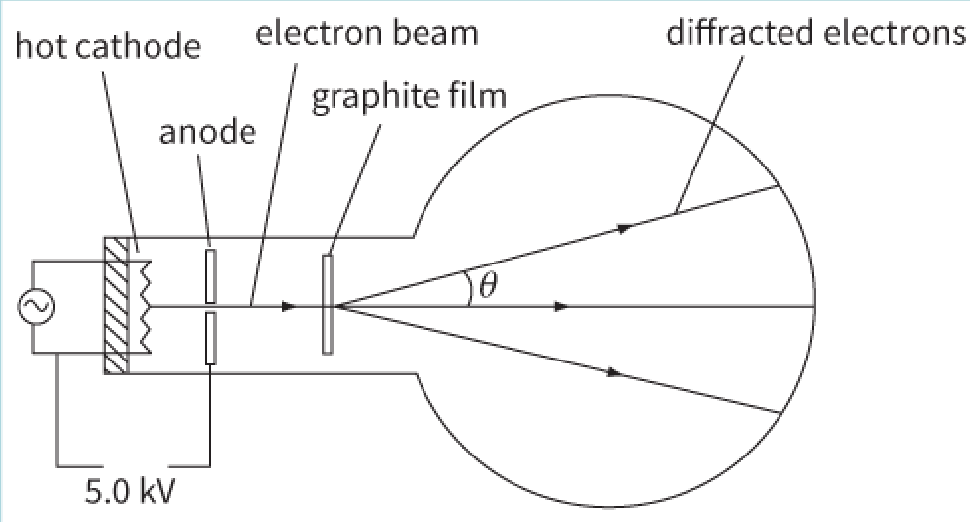


Figure 28.31

i Calculate the kinetic energy E (in joules) of the electrons incident on the graphite film. **[1]**

ii Show that the momentum of an electron is equal to $\sqrt{2Em_e}$ where m_e is the mass of an electron, and hence calculate the momentum of an electron. ($m_e = 9.11 \times 10^{-31}$ kg) **[3]**

iii Calculate the de Broglie wavelength of the electrons. **[2]**

c Explain how the wavelengths of neutrons and electrons moving with the same energy would compare. **[3]**

[Total: 11]

15 a Describe the importance of the Planck constant h in describing the behaviour of electromagnetic radiation and of electrons. **[2]**

b Light of wavelength 550 nm is incident normally on a metal plate. The intensity of the light is 800 W m^{-2}. All the incident light is absorbed by the metal plate. The plate has dimensions 5.0 cm × 5.0 cm.

i Explain how the light hitting the plate exerts force on the plate. **[3]**

ii Calculate the momentum of each photon of light. **[2]**

iii Calculate the force exerted on the plate due to the light. **[5]**

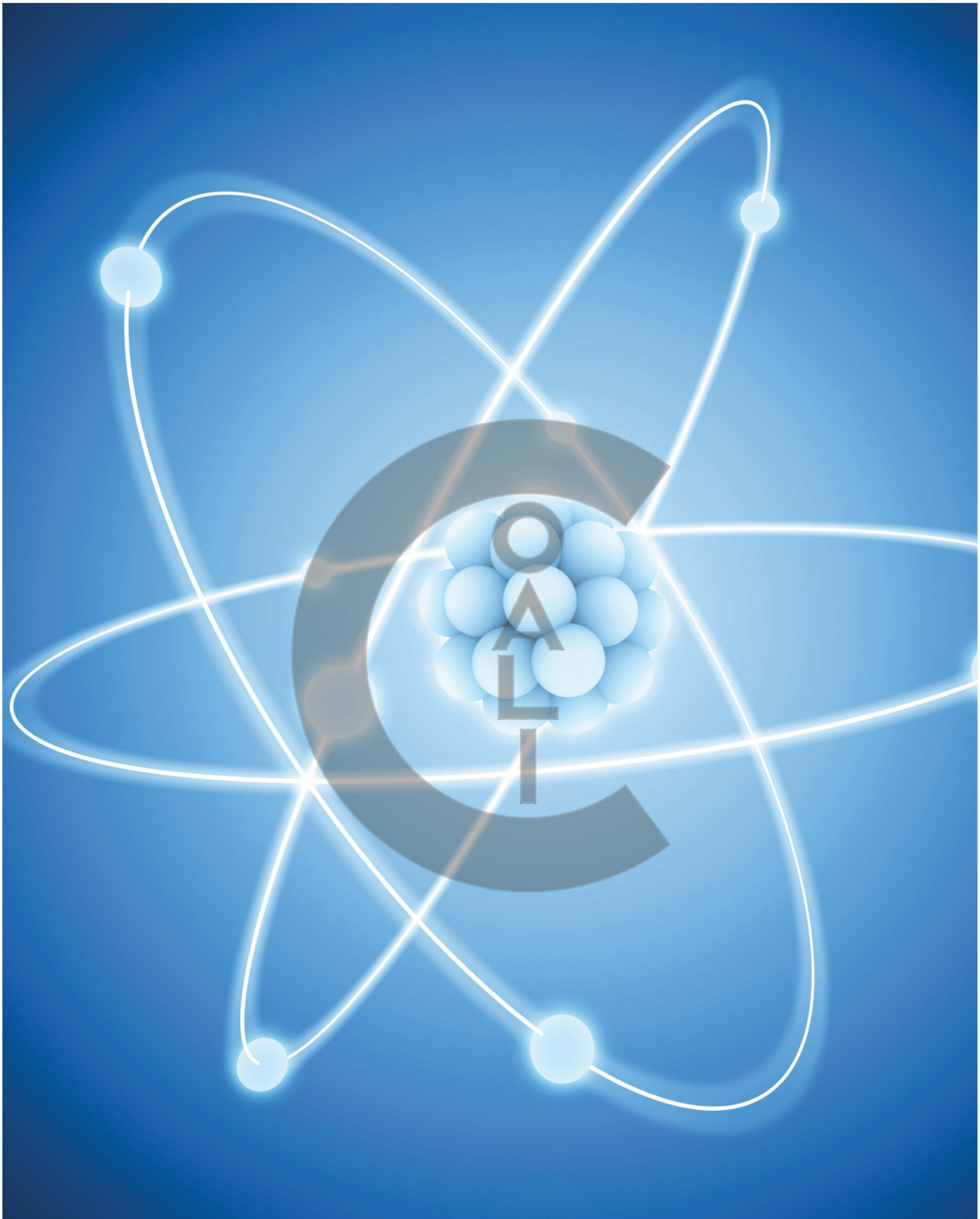
[Total: 12]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand that electromagnetic radiation interacts with matter as photons	28.2			
understand that a photon is a quantum of electromagnetic energy and its energy is given by: $E = hf$	28.2			
use the electronvolt (eV) as a unit of energy, where: 1 eV = 1.60×10^{-19} J	28.2			
understand photoelectric effect	28.2, 28.3			
understand the terms threshold frequency, threshold wavelength and work function	28.4			
use Einstein's photoelectric equation: $hf = \Phi + \frac{1}{2}mv_{max}{}^2$ or $\frac{hc}{\lambda} = \Phi + \frac{1}{2}mv_{max}{}^2$	28.4			
understand why the maximum kinetic energy of photoelectrons is independent of intensity, whereas the photoelectric current is proportional to intensity	28.4			
understand that the photoelectric effect provides evidence for a particulate nature of electromagnetic radiation	28.4			
understand that a photon has momentum, given by: $p = \frac{E}{c}$	28.5			
understand that diffraction provides evidence for the wave-like behaviour of particles (electrons)	28.9, 28.10			
understand that a moving particle has a de Broglie wavelength given by: $\lambda = \frac{h}{p}$	28.10			
understand that there are discrete electron energy levels in isolated atoms (such as atomic hydrogen)	28.7			
understand the appearance and formation of emission and absorption line spectra	28.6, 28.7			
recall and use the relationship: $hf = E_1 - E_2$	28.8			

COALI



〉Chapter 29

Nuclear physics

LEARNING INTENTIONS

In this chapter you will learn how to:

- understand the equivalence between energy and mass as represented by $E = mc^2$ and recall and use this equation
- represent simple nuclear reactions by nuclear equations
- define and use the terms mass defect and binding energy
- sketch the variation of binding energy per nucleon with nucleon number
- explain what is meant by nuclear fusion and nuclear fission
- explain the relevance of binding energy per nucleon to nuclear reactions, including nuclear fusion and nuclear fission
- calculate the energy released in nuclear reactions using $E = \Delta mc^2$
- understand that fluctuations in count rate provide evidence for the random nature of radioactive decay
- understand that radioactive decay is both spontaneous and random
- define activity and decay constant, and recall and use $A = \gamma N$
- define half-life
- use $\lambda = \frac{0.693}{t_{\frac{1}{2}}}$
- understand the exponential nature of radioactive decay, and sketch and use the relationship $x = x_0 e^{-\lambda \tau}$, where x could represent activity, number of undecayed nuclei or received count rate.

BEFORE YOU START

- Background knowledge of radioactivity from Chapter 15 would be useful in the study of this chapter. In pairs, write a summary of what you know.
- Try to remember, then write down, the particles that make up the nucleus and the forces the particles experience.
- Discuss why it is sensible to express the mass of particles in atomic mass units (u).

ENERGY AND THE NUCLEUS

The existence of every living organism on the surface of the Earth, including humans, depends on the light and heat from the Sun. Without the Sun, our planet would be a lifeless rock in space.

The Sun warms our oceans, stirs our atmosphere, creates our climate and, most importantly of all, gives energy to the plants that provide food and oxygen for life on Earth.

How does the Sun produce its energy? The Sun is an active hot ball of gas; it converts mass into energy. The Sun generates about 10^{26} W of radiant power by converting more than a billion kilograms of matter into energy every second. You do not need to worry about the Sun dying out soon – it has lots of mass! The mass of the Sun is about 10^{30} kg. Can you estimate the lifetime of the Sun?

In this chapter, we will examine how nuclear reactions produce energy. We will also look at the stability of nuclei, and how we can model the decay of unstable nuclei with mathematical equations.

Figure 29.1: Our understanding of nuclear physics is important to all life on Earth.

29.1 Balanced equations

When an unstable nucleus undergoes radioactive decay, the nucleus before the decay is often referred to as the **parent** nucleus and the new nucleus formed after the decay process is known as the **daughter** nucleus.

Radioactive decay processes can be represented by balanced equations. As with all equations representing nuclear processes, both nucleon number *A* and proton number *Z* are conserved.

- In α decay, the nucleon number decreases by 4 and the proton number decreases by 2.
- In β^- decay, the nucleon number is unchanged and the proton number increases by 1.
- In β+ decay, the nucleon number is unchanged and the proton number decreases by 1.
- In gamma decay, there is no change in either nucleon number or proton number.

The emission of α- and β-particles can be shown on a graph of nucleon number *A* plotted against proton number *Z*, as shown in Figure 29.2. The graph will appear different if neutron number is plotted against proton number.

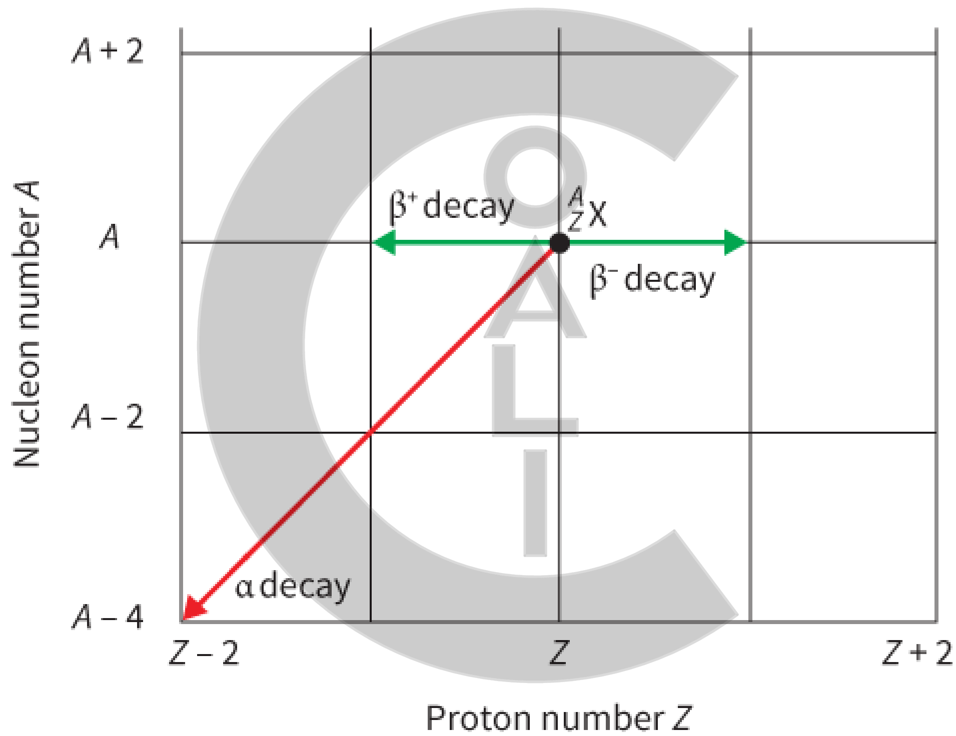


Figure 29.2: Emission of α- and β-particles.

WORKED EXAMPLES

1 Radon is a radioactive gas. The isotope of radon-222 decays by α emission to become a nucleus of polonium (Po). Here is the equation for the decay of a single isotope of radon-222:

$$^{222}_{86}\text{Rn} \rightarrow {}^{218}_{84}\text{Po} + {}^{4}_{2}\text{He}$$

Show that *A* and *Z* are conserved.

Compare the nucleon and proton numbers on both sides of the equation for the decay:

nucleon number *A*: 222 = 218 + 4

proton number *Z*: 86 = 84 + 2

Hint: *Remember that in α decay, A decreases by four and Z decreases by two. Don't confuse nucleon number A with neutron number N.*

In this case, radon-222 is the parent nucleus and polonium-218 is the daughter nucleus.

2 A carbon-14 nucleus (parent) decays by β^- emission to become an isotope of nitrogen (daughter). Here is the equation that represents this decay:

$${}^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}e$$

Show that both nucleon number and proton number are conserved.

Compare the nucleon and proton numbers on both sides of the equation for the decay:

nucleon number A: 14 = 14 + 0

proton number Z: 6 = 7 − 1

Hint: *Remember that in β^- decay, A remains the same and Z increases by 1.*

Questions

1 Study the decay equations given in Worked examples 1 and 2, and write balanced equations for the following:

- **a** A nucleus of radon-220 $({}^{220}_{86}Rn)$ decays by α emission to form an isotope of polonium, Po.
- **b** A nucleus of a sodium isotope $({}^{25}_{11}Na)$ decays by β^- emission to form an isotope of magnesium, Mg.

2 Copy and complete this equation for the β^- decay of a nucleus of argon:

$${}^{41}_{18}Ar \rightarrow K + ?$$

29.2 Mass and energy

In Chapter 15, we saw that energy is released when the nucleus of an unstable atom decays. How can we calculate the amount of energy released by radioactive decay? To find the answer to this, we need to think first about the masses of the particles involved.

We will start by considering a stable nucleus, C. This consists of six protons and six neutrons. Fortunately for us (because we have a lot of this form of carbon in our bodies), this is a very stable nuclide. This means that the nucleons are bound tightly together by the strong nuclear force. It takes a lot of energy to pull them apart.

Figure 29.3 shows the results of an imaginary experiment in which we have done just that. On the left-hand side of the balance is a single $^{12}_{6}C$ nucleus. On the right-hand side are six protons and six neutrons, the result of dismantling the nucleus. The surprising thing is that the balance is tipped to the right. The separate nucleons have **greater** mass than the nucleus itself. This means that the law of conservation of mass appears to have been broken. Have we violated what was thought to be a fundamental law of nature, something that was held to be true for hundreds of years?

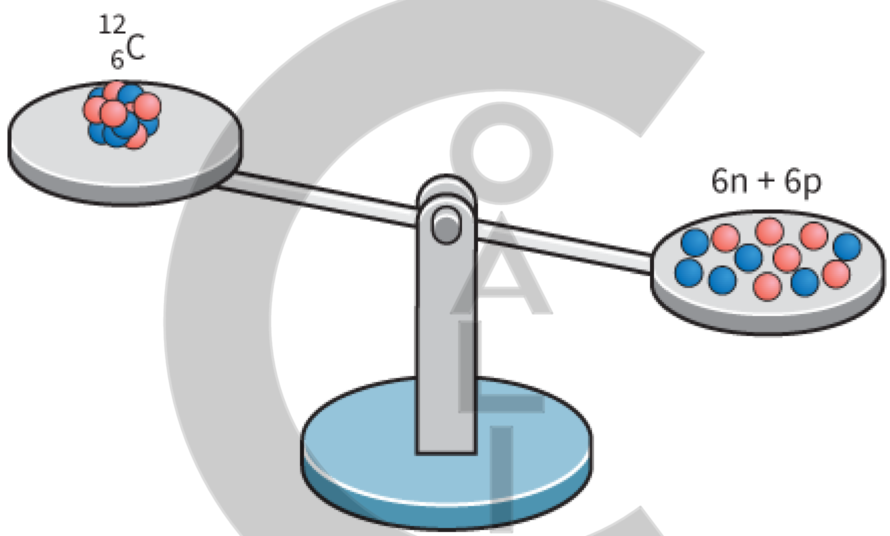


Figure 29.3: The mass of a nucleus is less than the total mass of its component protons and neutrons.

Notice that, in dismantling the $^{12}_{6}C$ nucleus, we have had to do work against the strong nuclear force. The nucleons attract one another with the strong nuclear force when we try to pull them apart. So, we have put energy into the nucleus to pull it apart, and this energy increases the potential energy of the individual nucleons. We can think of the nucleons within the nucleus as sitting in a deep potential well that results from the strong nuclear forces that hold the nucleus together. When we separate nucleons, we lift them out of this potential well, giving them more nuclear potential energy. This potential well is similar to that formed by the electric field around the nucleus; it is this well in which the atomic electrons sit, but it is much, much deeper. This explains why it is much easier to remove an electron from an atom than to remove a nucleon (proton or neutron) from the nucleus.

The problem of **changing** mass remains. To solve this problem, Einstein made the revolutionary hypothesis about energy and mass – to him, they were equivalent. This is not an easy idea. When bodies are in a higher energy state they have more mass than in a lower energy state. A bucket of water at the top of a hill will have more mass than when it is at the bottom because energy has been transferred to it in carrying it up the hill. A tennis ball travelling at 50 m s^{-1} will have more mass than the same tennis ball when stationary. In everyday life, the amount of extra mass is so small that it is not noticeable. However, the large changes in energy that occur in nuclear physics and high-energy physics make the changes in mass significant. Indeed, the increase in mass of particles, such as electrons, as they are accelerated to speeds near to the speed of light is a well-established experimental fact.

Another way to express this is to treat mass and energy as aspects of the same thing. Rather than having separate laws of conservation of mass and conservation of energy, we can combine these two. The total amount of mass and energy in a system is constant. There may be conversions from one to the other, but the total amount of 'mass–energy' remains constant.

Einstein's mass–energy equation

Albert Einstein produced his famous mass–energy equation, which links energy E and mass m:

$$E = mc^2$$

where c is the speed of light in a vacuum (free-space). The value of c is approximately 3.00×10^8 m s^{-1}, but its precise value has been fixed as $c = 299\,792\,458$ m s^{-1}.

Generally, we will be concerned with the changes in mass owing to changes in energy, when the equation becomes:

$$\Delta E = \Delta mc^2$$

where ΔE is the change in energy corresponding to a change, Δm in mass and c is the speed of light in a vacuum.

KEY EQUATION

$$\Delta E = \Delta mc^2$$

You may find this equation written in different forms:

$$E = c^2 \Delta m$$

$$E = mc^2$$

According to Einstein's equation:

- the mass of a system **increases** when energy is **supplied** to it
- the mass of a system **decreases** when energy is **released** from it.

Now, if we know the total mass of particles before a nuclear reaction and their total mass after the reaction, we can work out how much energy is released. Table 29.1 gives the mass in kilograms of each of the particles shown in Figure 29.3. Notice that this is described as the **rest mass** of the particle; that is, its mass when it is stationary. The mass of a particle will be greater when it is moving because of its increase in energy. Nuclear masses are measured to a high degree of precision using mass spectrometers, often to seven or eight significant figures.

Particle	Rest mass / 10^{-27} kg
${}^{1}_{1}p$	1.672 623
${}^{1}_{0}n$	1.674 929
${}^{12}_{6}C$ nucleus	19.926 483

Table 29.1: Rest masses of some particles. It is worth noting that the mass of the neutron is slightly larger than that of the proton (roughly 0.1% greater).

We can use the mass values to calculate the mass that is **released** as energy when nucleons combine to form a nucleus. So, for our particles in Figure 29.3, we have:

$$\begin{aligned}
\text{mass of system before} &= \text{mass of all the seperate nucleons} \\
&= (6 \times 1.672\,623 + 6 \times 1.674\,929) \times 10^{-27}\ \text{kg} \\
&= 20.085\,312 \times 10^{-27}\ \text{kg} \\
\text{mass of system after} &= \text{mass of the carbon-12 nucleus} \\
&= 19.926\,483 \times 10^{-27}\ \text{kg} \\
\text{decreases in the mass of the system} &= \Delta m = (20.085\,312 - 19.926\,483) \times 10^{-27}\ \text{kg} \\
&= 0.158\,829 \times 10^{-27}\ \text{kg}
\end{aligned}$$

When six protons and six neutrons combine to form the nucleus of carbon-12, there is a very small loss of mass Δm, known as the **mass defect**.

The mass defect of a nucleus is equal to the difference between the total mass of the individual separate nucleons and the mass of the nucleus.

The **loss** in mass implies that energy is **released** in this process. The energy released ΔE is given by Einstein's mass–energy equation. Therefore:

$$\begin{aligned}
\Delta E &= \Delta mc^2 = 0.158\,829 \times 10^{-27} \times (3.00 \times 10^8)^2 \\
\Delta E &\approx 1.43 \times 10^{-11}\ \text{J}
\end{aligned}$$

This may seem like a very small amount of energy, but it is a lot on the atomic scale. For comparison, the amount of energy released in a chemical reaction involving a single carbon atom would typically be of the order of 10^{-18} J, more than a million times smaller.

Now look at Worked example 3.

WORKED EXAMPLES

3 Use the following data to determine the minimum energy required to split a nucleus of oxygen-16 ($^{16}_{8}O$) into its separate nucleons. Give your answer in joules (J).

mass of proton = $1.672\,623 \times 10^{-27}$ kg

mass of neutron = $1.674\,929 \times 10^{-27}$ kg

mass of $^{16}_{8}O$ nucleus = $26.551\,559 \times 10^{-27}$ kg

speed of light $c = 3.00 \times 10^8$ m s^{-1}

Step 1 Find the difference Δm in kg between the mass of the oxygen nucleus and the mass of the individual nucleons. The $^{16}_{8}O$ nucleus has 8 protons and 8 neutrons.

$$\begin{aligned}
\Delta m &= \text{final mass} - \text{initial mass} \\
\Delta m &= ((8 \times 1.672\,623 + 8 \times 1.674\,929) - 26.551\,559) \times 10^{-27}\ \text{kg}) \\
\Delta m &\approx 2.20 \times 10^{-28}\ \text{kg}
\end{aligned}$$

There is an **increase** in the mass of this system, therefore, external energy must be **supplied** for the splitting of the oxygen-16 nucleus into its totally free nucleons.

Step 2 Use Einstein's mass–energy equation to determine the energy supplied:

$\Delta E = \Delta mc^2$

$$\text{energy supplied} = 2.20 \times 10^{-28} \times (3.00 \times 10^{8})^{2} \approx 1.98 \times 10^{-11}\ \text{J}$$

The value is the minimum energy. If the energy were to be greater than this value, the surplus energy would appear as kinetic energy of the nucleons.

Mass–energy conservation

Einstein pointed out that his equation $\Delta E = \Delta mc^2$ applied to **all** energy changes, not just nuclear processes. So, for example, it applies to chemical changes too. If we burn some carbon, we start off with carbon and oxygen. At the end, we have carbon dioxide and energy. If we measure the mass of the carbon dioxide, we find that it is very slightly less than the mass of the carbon and oxygen at the start of the experiment. The total potential energy of the system will be less than at the start of the experiment, hence the mass is less. In a chemical reaction such as this, the change in mass is very small, less than a microgram if we start with 1 kg of carbon and oxygen. Compare this with the change in mass that occurs during the fission of 1 kg of uranium, described later. The change in mass in a chemical reaction is a much, much smaller proportion of the original mass, which is why we don't notice it.

Questions

3 The Sun releases vast amounts of energy. Its power output is 4.0×10^{26} W. Estimate how much its mass decreases each second because of this energy loss.

4 **a** Calculate the energy released if a ${}^{4}_{2}\text{He}$ nucleus is formed from separate stationary protons and neutrons. The masses of the particles are given in Table 29.2.

b Calculate also the energy released per nucleon.

Particle	Mass / 10^{-27} kg
${}^{1}_{1}\text{p}$	1.672 623
${}^{1}_{0}\text{n}$	1.674 929
${}^{4}_{2}\text{He}$	6.644 661

Table 29.2: Masses of some particles.

5 The rest mass of a golf ball is 150 g.

Calculate its increase in mass when it is travelling at 50 m s^{-1}. What is this as a percentage of its rest mass?

Another unit of mass

When calculating energy values using $\Delta E = \Delta mc^2$, it is essential to use values of mass in kg, the SI unit of mass. However, the mass of a nucleus is very small, perhaps 10^{-25} kg, and these numbers are awkward. As an alternative, atomic and nuclear masses are often given in a different unit, the **atomic mass unit** (symbol u). You have already met this alternative unit for mass in Chapter 15.

The conversion factor for atomic mass unit u to kilogram (kg) is:

$$1\ \text{u} = 1.660\ 538\ 921(73) \times 10^{-27}\ \text{kg}$$

To convert the mass of a particle from u to kg, you just multiply by the conversion factor shown–usually 1.6605×10^{-27} is sufficiently accurate.

Table 29.3 shows the masses of proton, neutron and some nuclides in u. It is worth noting that the mass in u is close to the nucleon number *A*. For example, the mass of uranium-235 nucleus is 235 u.

Nuclide	Symbol	Mass / u
proton	${}^{1}_{1}p$	1.007 276
neutron	${}^{1}_{0}n$	1.008 665
helium-4	${}^{4}_{2}He$	4.002 602
carbon-12	${}^{12}_{6}C$	12.000 000
potassium-40	${}^{40}_{19}K$	39.963 998
uranium-235	${}^{235}_{92}U$	235.043 930

Table 29.3: Masses of some particles in u. Some have been measured to several more decimal places than are shown here.

Questions

6 **a** The mass of an atom of ${}^{56}_{26}Fe$ is 55.934 937 u. Calculate its mass in kg.

b The mass of an atom of ${}^{16}_{8}O$ is $2.656\,015 \times 10^{-26}$ kg. Calculate its mass in u.

7 Table 29.3 gives the masses (in u) of several particles.

(Avogadro constant $N_A = 6.02 \times 10^{23}$ mol^{-1}.)

Use the table to determine to three significant figures:

a the mass in kg of a helium-4 nucleus

b the mass in gram (g) of 1.0 mole of uranium-235 nuclei.

29.3 Energy released in radioactive decay

Unstable nuclei may emit α- and β-particles with large amounts of kinetic energy. We can use Einstein's mass–energy equation $\Delta E = \Delta mc^2$ to explain the origin of this energy. Take, for example, the decay of a nucleus of uranium-238. It decays by emitting an α-particle and changes into an isotope of thorium:

$$^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\text{He}$$

The uranium nucleus is in a high-energy, relatively unstable state. It emits the α-particle and the remaining thorium nucleus is in a lower, more stable energy state. There is a decrease in the mass of the system. That is, the combined mass of the thorium nucleus and the α-particle is less than the mass of the uranium nucleus. According to Einstein's mass–energy equation, this difference in mass Δm is equivalent to the energy released as kinetic energy of the products. Using the most accurate values available:

$$\begin{aligned}
\text{mass of } {}^{238}_{92}\text{U nucleus} &= 3.952\,83 \times 10^{-25}\text{ kg} \\
\text{total mass of } {}^{234}_{90}\text{Th nucleus and α-particle } ({}^{4}_{2}\text{He}) &= 3.952\,76 \times 10^{-25}\text{ kg} \\
\text{change in mass } \Delta m &= (3.952\,76 - 3.952\,83) \times 10^{-25}\text{ kg} \\
&\approx -7.0 \times 10^{-30}\text{ kg}
\end{aligned}$$

The minus sign shows a decrease in mass, hence, according to the equation $\Delta E = \Delta mc^2$, energy is released in the decay process:

$$\begin{aligned}
\text{energy released} &\approx 7.0 \times 10^{-30} \times (3.0 \times 10^{8})^2 \\
&\approx 6.3 \times 10^{-13}\text{ J}
\end{aligned}$$

This is an enormous amount of energy for a single decay. One mole of uranium-238, which has 6.02×10^{23} nuclei, has the potential to emit total energy equal to about 10^{11} J.

We can calculate the energy released in all decay reactions, including β decay, using the same ideas.

Question

8 A nucleus of beryllium-10 ($^{10}_{4}\text{Be}$) decays into an isotope of boron by β^- emission. The chemical symbol for boron is B.

a Write a nuclear decay equation for the nucleus of beryllium-10.

b Calculate the energy released in this decay and state its form.

(Mass of $^{10}_{4}\text{Be}$ nucleus = $1.662\,38 \times 10^{-26}$ kg; mass of boron isotope = $1.662\,19 \times 10^{-26}$ kg; mass of electron = $9.109\,56 \times 10^{-31}$ kg.)

29.4 Binding energy and stability

We can now begin to see why some nuclei are more stable than others. If a nucleus is formed from separate nucleons, energy is released. In order to pull the nucleus apart, energy must be put in; in other words, work must be done against the strong nuclear force that holds the nucleons together. The more energy involved in this, the more stable the nucleus.

The minimum energy needed to completely pull a nucleus apart into its separate nucleons is known as the **binding energy** of the nucleus.

Take care: this is **not** energy stored in the nucleus. On the contrary, it is the energy that must be put in to the nucleus in order to pull it apart. In the example of $^{12}_{6}\mathrm{C}$ discussed earlier, we calculated the binding energy from the mass difference between the mass of the $^{12}_{6}\mathrm{C}$ nucleus and the masses of the separate protons and neutrons.

In order to compare the stability of different nuclides, we need to consider the **binding energy per nucleon**.

We can determine the binding energy per nucleon for a nuclide as follows:

- Determine the mass defect for the nucleus.
- Use Einstein's mass–energy equation to determine the binding energy of the nucleus by multiplying the mass defect by c^2.
- Divide the binding energy of the nucleus by the number of nucleons to calculate the binding energy per nucleon.

Figure 29.4 shows the variation of binding energy per nucleon with nucleon number *A* for nuclei. The red dot represents the plot for the iron-56 nuclide, which is from Worked example 4. The greater the value of the binding energy per nucleon, the more tightly bound are the nucleons that make up the nucleus. The most striking observation is that not all nuclides are the same – some nuclides are more tightly bound than others.

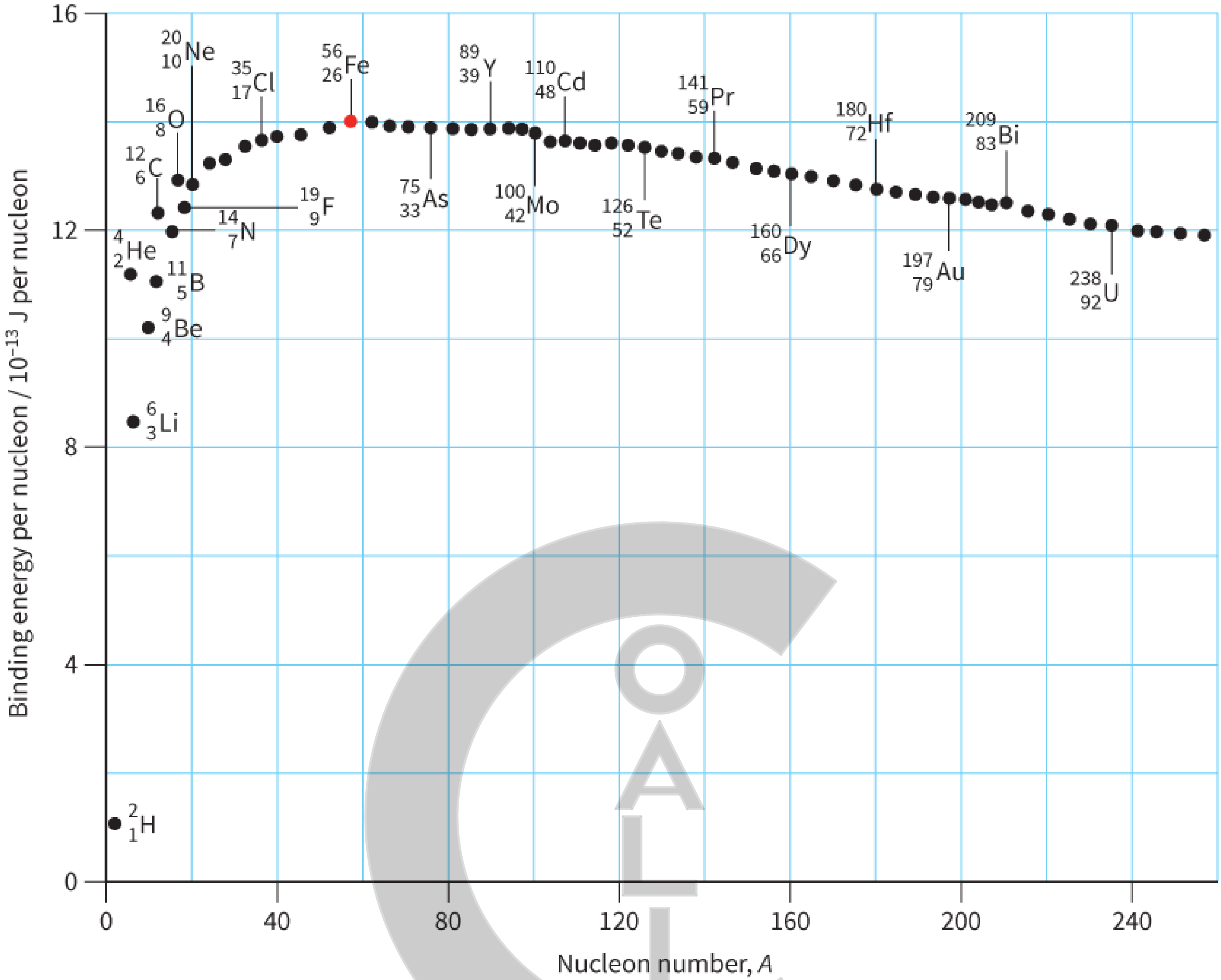


Figure 29.4: This graph shows the binding energy per nucleon for a number of nuclei. The nucleus becomes more stable as binding energy per nucleon increases.

If you further examine this graph, you will see that the general trend is for light nuclei to have low binding energies per nucleon. Note, however, that helium has a much higher binding energy than its place in the Periodic Table might suggest. The high binding energy per nucleon means that it is very stable. Other common stable nuclei include $^{12}_{6}C$ and $^{16}_{8}O$ which can be thought of, respectively, as three and four α-particles bound together (Figure 29.5).

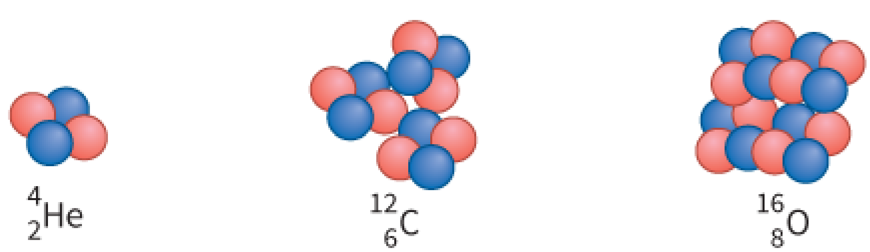


Figure 29.5: More stable nuclei are formed when 'α-particles' are bound together. In $^{12}_{6}C$ and $^{16}_{8}O$ the 'α-particles' do not remain separate, as shown here; rather, the protons and neutrons are tightly packed together.

For nuclides with $A > 20$ approximately, there is not much variation in binding energy per nucleon. The greatest value of binding energy per nucleon is found for $^{56}_{26}\text{Fe}$. This isotope of iron requires the most energy per nucleon to dismantle it into separate nucleons; hence iron-56 is the most stable isotope in nature.

WORKED EXAMPLES

4 Use the following data to calculate the binding energy per nucleon for the nuclide $^{56}_{26}\text{Fe}$

mass of neutron = 1.675×10^{-27} kg

mass of proton = 1.673×10^{-27} kg

mass of $^{56}_{26}\text{Fe}$ nucleus = 9.288×10^{-26} kg

Step 1 Calculate the mass defect.

$$\text{number of neutrons} = 56 - 26 = 30$$

$$\text{mass defect} = (30 \times 1.675 \times 10^{-27} + 26 \times 1.673 \times 10^{-27}) - 9.288 \times 10^{-26}$$

$$\text{mass defect} = 8.680 \times 10^{-28}\ \text{kg}$$

Step 2 Calculate the binding energy of the nucleus using Einstein's mass-energy Zequation.

$$\text{binding energy} = \Delta mc^2 = 8.680 \times 10^{-28} \times (3.00 \times 10^8)^2$$

$$\text{binding energy} = 7.812 \times 10^{-11}\ \text{J}$$

Step 3 Calculate the binding energy per nucleon.

$$\text{binding energy per nucleon} = \frac{7.812 \times 10^{-11}}{56} \approx 14 \times 10^{-13}\ \text{J}$$

Have another look at Figure 29.4. The value matches with the plot of iron-56.

Questions

9 **a** Explain why hydrogen $^{1}_{1}\text{H}$ (proton) cannot appear on the graph shown in Figure 29.4.

b Use Figure 29.4 to estimate the binding energy of the nuclide $^{14}_{7}\text{N}$

10 The mass of a $^{8}_{4}\text{Be}$ nucleus is 1.33×10^{-26} kg. For the nucleus of $^{8}_{4}\text{Be}$, determine:

a the mass defect in kg

b the binding energy of the nucleus in MeV

c the binding energy (in MeV) per nucleon for the nucleus.

Binding energy, fission and fusion

We can use the binding energy graph to help us decide which nuclear processes – fission, fusion, radioactive decay – are likely to occur (Figure 29.6).

Fission

Fission is the process in which a massive nucleus splits to form two smaller fragments (rather than simply emitting α- or β-radiation).

The isotope of uranium-235 can split spontaneously, but such an event is very rare. However, in a process known as **induced** fission, uranium-235 can be made split by absorbing a slow-moving neutron. A typical nuclear reaction is shown:

$$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{236}_{92}U \rightarrow ^{142}_{56}Ba + ^{92}_{36}Kr + 2^{1}_{0}n$$

The uranium-235 nucleus captures the neutron and becomes a highly unstable nucleus of uranium-236. In a very short period of time, typically a few microseconds, the fission of uranium-236 results in barium-142, krypton-92 and two fast-moving neutrons. Energy is released in the reaction as kinetic energy because the total mass of the system decreases. This is what we would expect from Einstein's mass-energy equation. There is now another alternative way of interpreting this reaction. If we look at Figure 29.6, we see that these two fragments have greater binding energy per nucleon than the original uranium nucleus. Hence, if the uranium nucleus splits in this way, energy will be released. The total binding energy of $^{142}_{56}Ba$ and $^{92}_{36}Kr$ is greater than the binding energy of $^{235}_{92}U$ - the difference is the energy released. (Note: the neutron is a lone nucleon, so it has zero binding energy.)

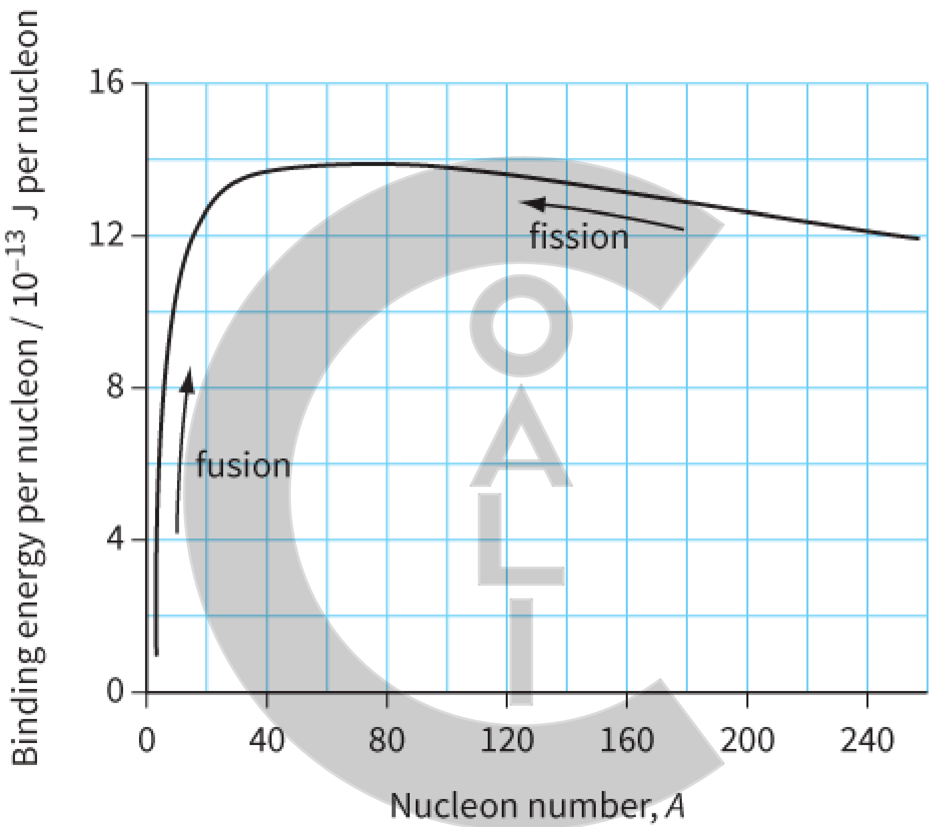


Figure 29.6: Both fusion and fission are processes that tend to increase the binding energy per nucleon of the particles involved.

Fusion

Fusion is the process by which two very light nuclei join together to form a heavier nucleus. This is the process by which energy is released in the Sun, when hydrogen nuclei fuse to form helium nuclei. When two light nuclei join together, the final binding energy of the nucleus formed is greater than the total binding energy of the fusing-nuclei – once again, the difference is the energy released in the fusion reaction. The high binding energy of the $^{4}_{2}He$ nuclide means that it is rare for these nuclei to fuse.

The following fusion reaction is one of the many taking place inside the core of stars, including our Sun:

$$^{2}_{1}H + ^{1}_{1}p \rightarrow ^{3}_{2}He$$

A deuterium nucleus $({}^{2}_{1}\mathrm{H})$ joins together with a proton $({}^{1}_{1}\mathrm{p})$ to make the helium-3 nucleus. The binding energy of deuterium nucleus is 2.2 MeV, and the binding energy of helium-3 nucleus is 7.7 MeV. The energy released in this fusion reaction is 5.5 MeV, which is the difference in the two binding energies. It is worth noting that the binding energy per nucleon of the helium-3 nucleus is greater that of the deuterium nucleus – fusion increases the binding energy per nucleon, as shown on Figure 29.6.

Questions

11 Use the binding energy graph (Figure 29.6) to suggest why fission is unlikely to occur with 'light nuclei' ($A < 20$) and why fusion is unlikely to occur for heavier nuclei ($A > 40$).

12 Use the information given in the fusion section, to determine the binding energy (in MeV) per nucleon of each particle in the following fusion reaction:

$${}^{2}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{p} \rightarrow {}^{3}_{2}\mathrm{He}$$

Comment on your answers.

29.5 Randomness and radioactive decay

Listen to a counter connected to a Geiger–Müller (GM) tube that is detecting the radiation from a weak source, so that the count rate is about one count per second. Each count represents the detection of a single α-particle or a β-particle or a γ-ray photon. You will notice that the individual counts do not come regularly.

The counter beeps or clicks in a random, irregular manner. If you try to predict when the next clicks will come, you are unlikely to be right.

You can see the same effect if you have a ratemeter, which can measure faster rates (Figure 29.7). The needle fluctuates up and down. Usually, a ratemeter has a control for setting the 'time constant'–the time over which the meter averages out the fluctuations. Usually, this can be set to 1 s or 5 s. The fluctuations are smoothed out more on the 5 s setting.

Figure 29.7: The time constant of this ratemeter can be adjusted to smooth out rapid fluctuations in the count rate.

Figure 29.8 shows a graph of count rate against time, with a smoothing of a few seconds. The count rate decreases with time as the number of radioactive nuclei that are left decreases. The fluctuations either side are caused by the randomness of the decay.

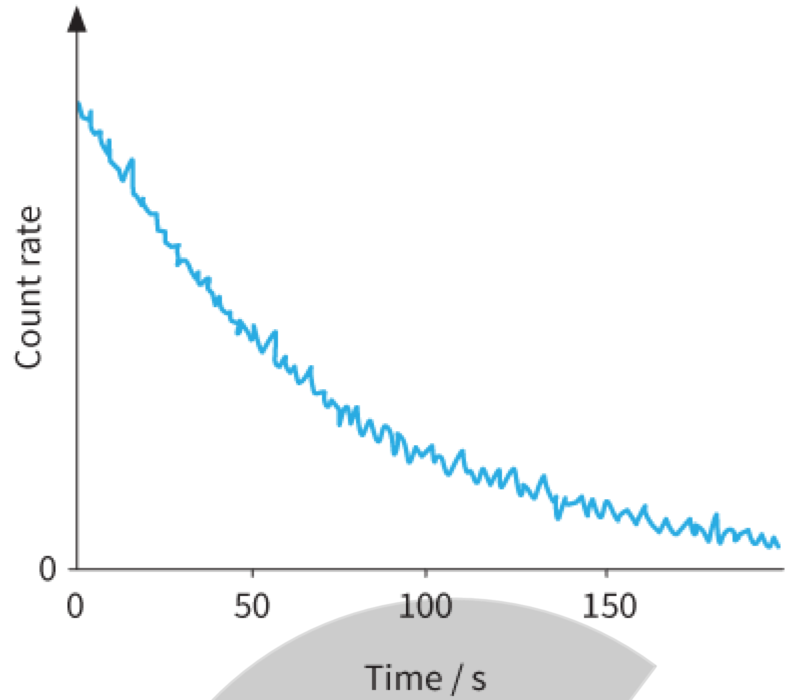


Figure 29.8: Count rate showing randomness of decay.

So, it is apparent that radioactive decay is a random, irregular phenomenon. But is it completely unpredictable? Well, not really. We can measure the average rate of decay. We might measure the number of counts detected in 1000 s, and then calculate the average number per second. We cannot be sure about this average rate either, because the number of counts in 1000 s will fluctuate, too. All of our measurements of radioactive decay are inherently uncertain and imprecise but, by taking averages over a sufficiently long time period, we can reduce or smooth out the random fluctuations to reveal the underlying pattern.

Spontaneous decay

Radioactive decay occurs within the unstable nucleus of an atom. A nucleus emits radiation and becomes the nucleus of an atom of a different element. We cannot predict, for a particular nucleus, when it will happen. If we sit and stare at an individual nucleus, we cannot see any change that will tell us that it is getting ready to decay. And if it doesn't decay in the first hour when we are watching it, we cannot say that it is any more likely to decay in the next hour. What is more, we cannot affect the probability of an individual nucleus decaying, for example, by changing its temperature.

This is slightly odd, because it goes against our everyday experience of the way things around us change. We observe things changing. They gradually age, die, rot away. But this is not how things are on the scale of atoms and nuclei. Many of the atoms of which we are made have existed for billions of years, and will still exist long after we are gone. The nucleus of an atom does not age.

If we look at a very large number of atoms of a radioactive substance, we will see that the number of undecayed nuclei gradually decreases. However, we cannot predict when an **individual** nucleus will decay. Each nucleus 'makes up its own mind' when to decay, independently from its neighbours. This is because neighbouring nuclei do not interact with one another (unlike neighbouring atoms). The nucleus is a tiny fraction of the size of the atom, and the nuclear forces do not extend very far outside the nucleus. So, one nucleus cannot affect a neighbouring nucleus by means of the nuclear force. Being inside a nucleus is a bit like living in a house in the middle of nowhere; you can just see out into the garden, but everything is darkness beyond and the next house is 1000 km away.

The fact that individual nuclei decay independently of their neighbours and of environmental factors, accounts for the random pattern of clicks that we hear from a Geiger counter and the fluctuations of the needle on the ratemeter. Radioactive decay is both **spontaneous** and **random**.

Nuclear decay is spontaneous because:

- the decay of a particular nucleus is not affected by the presence of other nuclei
- the decay of nuclei cannot be affected by chemical reactions or external factors such as temperature and pressure.

Nuclear decay is **random** because:

- it is impossible to predict when a particular nucleus in a sample is going to decay
- each nucleus in a sample has the same chance of decaying per unit time.

29.6 The mathematics of radioactive decay

We have seen that radioactive decay is a random, spontaneous process. Because we cannot say when an individual nucleus will decay, we have to start thinking about very large numbers of nuclei. Even a tiny speck of radioactive material will contain more than 10^{15} nuclei. Then we can talk about the average number of nuclei that we expect to decay in a particular time interval; in other words, we can find out the **average** decay rate. Although we cannot make predictions for individual nuclei, we can say that certain types of nuclei are more likely to decay than others. For example, a nucleus of carbon-12 is stable; carbon-14 decays gradually over thousands of years; carbon-15 nuclei last, on average, a few seconds.

So, because of the spontaneous nature of radioactive decay, we have to make measurements on very large numbers of nuclei and then calculate averages. One quantity we can determine is the probability that an individual nucleus will decay in a particular time interval. For example, suppose we observe one million nuclei of a particular isotope. After one hour, 200 000 have decayed. Then the probability that an individual nucleus will decay in one hour is 0.2 or 20%, since 20% of the nuclei have decayed in this time. (Of course, this is only an approximate value, since we might repeat the experiment and find that only 199 000 decay because of the random nature of the decay. The more times we repeat the experiment, the more reliable our answer will be.)

We can now define the decay constant:

The probability that an individual nucleus will decay per unit time interval is called the **decay constant**, λ.

For the example, we have:

$$\text{decay constant } \lambda = 0.20\ \text{h}^{-1}$$

Note that, because we are measuring the probability of decay per unit time interval, λ has units of h^{-1} (or s^{-1}, day^{-1}, year^{-1}, etc.).

The **activity** A of a radioactive sample is the rate at which nuclei decay or disintegrate.

Activity is measured in decays per second (or h^{-1}, day^{-1}). An activity of one decay per second is one becquerel (1 Bq):

$$1\ \text{Bq} = 1\ \text{s}^{-1}$$

Clearly, the activity of a sample depends on the decay constant λ of the isotope under consideration. The greater the decay constant (the probability that an individual nucleus decays per unit time interval), the greater is the activity of the sample. It also depends on the number of undecayed nuclei N present in the sample.

For a sample of N undecayed nuclei, we have:

$$A = -\lambda N$$

where λ is the decay constant of the isotope and N is the number of undecayed nuclei.

KEY EQUATION

Activity A is given by:

$$A = -\lambda N$$

Activity A is equal to rate of decay of nuclei; therefore $A = \lambda N$.

The minus sign indicates that the number of undecayed nuclei decreases with time. We can omit this minus sign if we just want to determine the magnitude of the activity. So, in calculations, we can just use $A = \lambda N$.

We can also think of the activity as the number of α- or β-particles emitted from the source per unit time. Hence, we can also write the activity A as:

$$A = \frac{\Delta N}{\Delta t}$$

where ΔN is equal to the number of emissions (or decays) in a small time interval of Δt.

WORKED EXAMPLE

5 A radioactive source emits β-particles. The source has an activity of 2.8×10^7 Bq. Estimate the number of β-particles emitted in a time interval of 2.0 minutes. State one assumption made.

Step 1 Write down the given quantities in SI units.

$A = 2.8 \times 10^7$ Bq $\quad \Delta t = 120$ s

Step 2 Determine the number of β-particles emitted.

$$A = \frac{\Delta N}{\Delta t}$$

$$\Delta N = A \times \Delta t = 2.8 \times 10^7 \times 120$$
$$= 3.36 \times 10^9 \approx 3.4 \times 10^9$$

We have assumed that the activity remains constant over a period of 2.0 minutes.

6 A sample consists of 1000 undecayed nuclei of a nuclide whose decay constant is 0.20 s^{-1}. Determine the initial activity of the sample. Estimate the activity of the sample after 1.0 s.

Step 1 Since activity $A = \lambda N$, we have:

$A = 0.20 \times 1000 = 200\ s^{-1} = 200$ Bq

Step 2 After 1.0 s, we might expect 800 nuclei to remain undecayed.

The activity of the sample would then be:

$A = 0.2 \times 800 = 160\ s^{-1} = 160$ Bq

(In fact, it would be slightly higher than this. Since the rate of decay decreases with time all the time, less than 200 nuclei would decay during the first second.)

Count rate

Although we are often interested in finding the activity of a sample of radioactive material, we cannot usually measure this directly. This is because we cannot easily detect **all** of the radiation emitted. Some will escape past our detectors, and some may be absorbed within the sample itself. A (Geiger–Muller) GM tube placed in front of a radioactive source therefore only detects a fraction of the activity. The further it is from the source, the smaller the count rate. Therefore, our measurements give a received **count rate** R that is significantly lower than the activity A. If we know how efficient our detecting system is, we can deduce A from R. If the level of background radiation is significant, then it must be subtracted to give the **corrected** count rate.

Questions

13 A sample of carbon-15 initially contains 500 000 undecayed nuclei. The decay constant for this isotope of carbon is 0.30 s^{-1}.

Calculate the initial activity of the sample.

14 A small sample of radium gives a received count rate of 20 counts per minute in a detector. It is known that the counter detects only 10% of the decays from the sample. The sample contains 1.5×10^9 undecayed nuclei. Calculate the decay constant of this form of radium.

15 A radioactive sample is known to emit α-, β- and γ-radiations.

Suggest **four** reasons why the count rate measured by a Geiger counter placed next to this sample would be lower than the activity of the sample.

29.7 Decay graphs and equations

The activity of a radioactive substance gradually diminishes as time goes by. The atomic nuclei emit radiation and become different substances. The pattern of radioactive decay is an example of a very important pattern found in many different situations, a pattern called **exponential decay**. Figure 29.9 shows the decay graphs for three different isotopes, each with a different rate of decay.

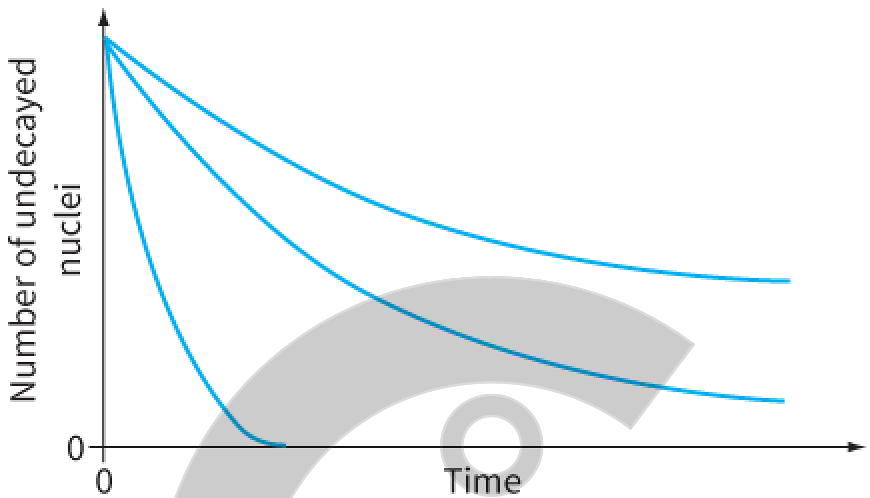


Figure 29.9: Some radioactive materials decay faster than others.

Although the three graphs look different, they all have something in common – their shape. They are curved lines having a special property. If you know what is meant by the **half-life** of an isotope, then you will understand what is special about the shape of these curves.

The half-life $t_{\frac{1}{2}}$ of an isotope is the mean time taken for half of the active nuclei in a sample to decay.

In a time equal to one half-life, the activity of the sample will also halve. This is because activity is directly proportional to the number of undecayed nuclei ($A \propto N$). It takes the same amount of time again for half of the remainder of the nuclei to decay, and a third half-life for half of the new remainder to decay (Figure 29.10).

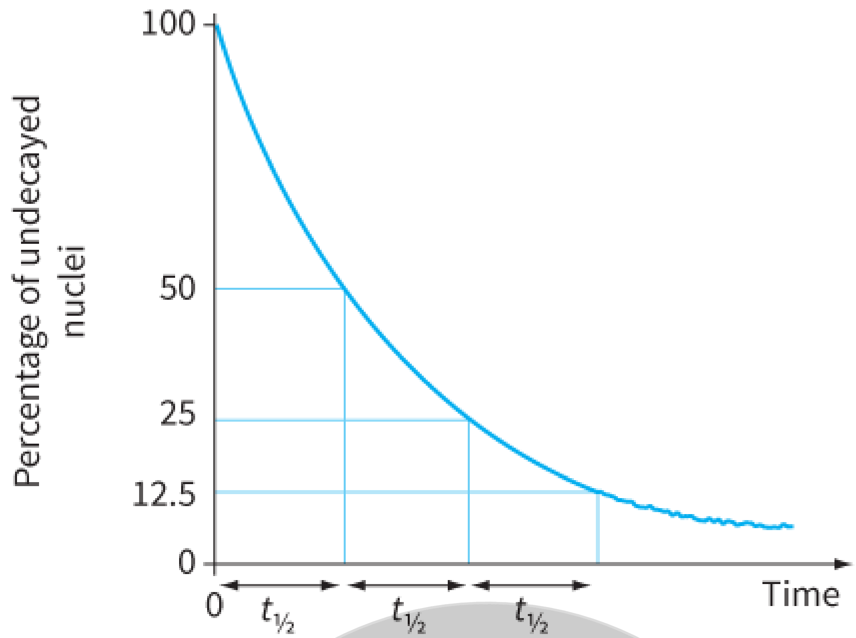


Figure 29.10: All radioactive decay graphs have the same characteristic shape.

In principle, the graph never reaches zero; it just gets closer and closer. In practice, when only a few undecayed nuclei remain, the graph will cease to be a smooth curve (because of the random nature of the decay) and it will eventually reach zero. We use the idea of half-life because we cannot say when a sample will have completely decayed.

Mathematical equations for radioactive decay

We can write an equation to represent the graph shown in Figure 29.10. If we start with N_0 undecayed nuclei, then the number N that remain undecayed after time t is given by:

$$N = N_0 \, e^{-\lambda t}$$

In this equation, λ is the decay constant of an isotope, as before. (You may also see this written as $N = N_0 \exp(-\lambda t)$.) Note that you must take care with units. If λ is in s^{-1}, then the time t must be in s.

The symbol e represents the number e = 2.71828…, a special number in the same way that π is a special number. You will need to be able to use the e^x button on your calculator to solve problems involving e.

PRACTICAL ACTIVITY 31.1

Determining half-life

If you are to determine the half-life of a radioactive substance in the laboratory, you need to choose something that will not decay too quickly or too slowly. In practice, the most suitable isotope is protactinium-234, which decays by emitting β^--radiation. This is available in a bottle containing a solution of a uranium compound (uranyl(VI) nitrate) (Figure 29.11). By shaking the bottle, you can separate the protactinium into the top layer of solvent in the bottle. The counter allows you to measure the decay of the protactinium.

After recording the number of counts in consecutive 10-second intervals over a period of a few minutes, you can then draw a graph, and use it to find the half-life of protactinium-234.

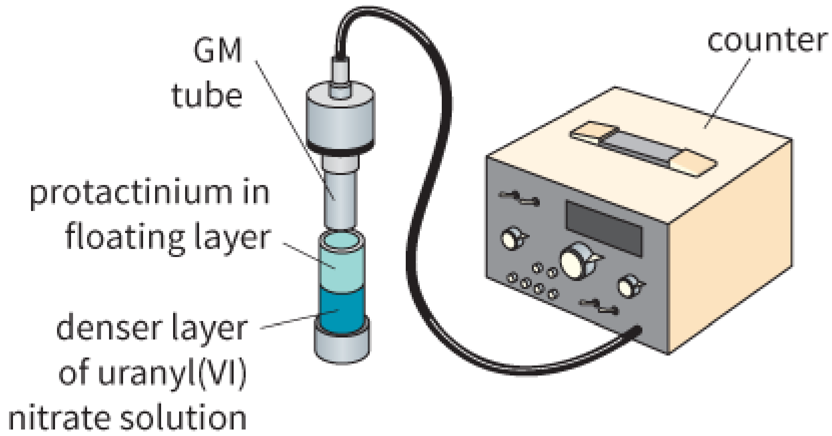


Figure 29.11: Practical arrangement for observing the decay of protactinium-234.

The activity A of a sample is directly proportional to the number of undecayed nuclei N. Hence the activity of the sample decreases exponentially:

$$A = A_0 e^{-\lambda t} \quad (A_0 \text{ is the activity at time } t = 0.)$$

Usually, we measure the corrected count rate R in the laboratory rather than the activity or the number of undecayed nuclei. Since the count rate is a fraction of the activity, it too decreases exponentially with time:

$$R = R_0 e^{-\lambda t} \quad (R_0 \text{ is the corrected count rate at time } t = 0.)$$

KEY EQUATION

$$x = x_0 e^{-\lambda t}$$

where x can represent activity A, number of undecayed nuclei N or received count rate R.

(λ is the decay constant and x is the quantity left at time t.)

Now look at Worked examples 7 and 8.

WORKED EXAMPLES

7 Suppose we start an experiment with 1.0×10^{15} undecayed nuclei of an isotope for which λ is equal to $0.02\ s^{-1}$. Determine the number of undecayed nuclei after 20 s.

Step 1 In this case, we have $N_0 = 1.0 \times 10^{15}$, $\lambda = 0.02\ s^{-1}$ and $t = 20$ s. Substituting in the equation gives:

$$N = 1.0 \times 10^{15}\ e^{-0.02 \times 20}$$

Step 2 Use the e^x button and calculate N.

$$N = 1.0 \times 10^{15}\ e^{-0.02 \times 20} = 6.7 \times 10^{14}$$

8 A sample initially contains 1000 undecayed nuclei of an isotope whose decay constant $\lambda = 0.10\ min^{-1}$. Draw a graph to show how the sample will decay over a period of 10 minutes.

Step 1 We have $N_0 = 1000$ and $\lambda = 0.10\ min^{-1}$. Hence, we can write the equation for this decay:

$N = 1000\ e^{-0.10 \times t}$

Step 2 Calculate values of the number N of undecayed nuclei at intervals of 1.0 min (60 s); this gives Table 29.4 and the graph shown in Figure 29.12.

t / min	0	1.0	2.0	3.0	4.0	5.0
N	1000	905	819	741	670	607
t / min	6.0	7.0	8.0	9.0	10.0	
N	549	497	449	407	368	

Table 29.4: For Worked example 8.

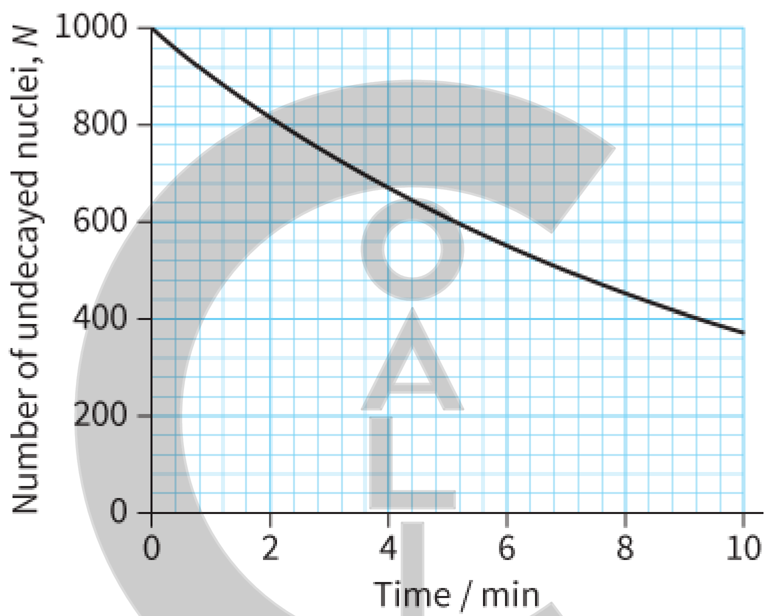


Figure 29.12: Radioactive decay graph.

Questions

16 The isotope nitrogen-13 has a half-life of 10 min. A sample initially contains 8.0×10^{10} undecayed nuclei.

a Write down an equation to show how the number undecayed N depends on time t.

b Calculate how many undecayed nuclei will remain after 10 min, and after 20 min.

c Determine how many nuclei will decay during the first 30 min.

17 A sample of an isotope for which $\lambda = 0.10\ s^{-1}$ contains 5.0×10^9 undecayed nuclei at the start of an experiment. Determine:

a the number of undecayed nuclei after 50 s

b its activity after 50 s.

18 The value of λ for protactinium-234 is $9.6 \times 10^{-3}\ s^{-1}$. Table 29.5 shows the number of undecayed nuclei N in a sample.

Copy and complete Table 29.5. Draw a graph of N against t, and use it to find the half-life $t_{\frac{1}{2}}$ of protactinium-234.

t / *s*	0	20	40	60	80	100	120	140
N	400	330						

Table 29.5: Data for Question 18.

29.8 Decay constant λ and half-life $t_{\frac{1}{2}}$

An isotope that decays rapidly has a short half-life $t_{\frac{1}{2}}$. Its decay constant must be large, since the probability per unit time of an individual nucleus decaying must be high. What is the connection between the decay constant and the half-life?

If $e^x = y$, then $x = \ln y$

In a time equal to one half-life $t_{\frac{1}{2}}$, the number of undecayed nuclei is halved. Hence the equation:

$$N = N_0 e^{-\lambda t}$$

becomes:

$$\frac{N}{N_0} = e^{(-\lambda t_{\frac{1}{2}})} = \frac{1}{2}$$

Therefore:

$$\begin{aligned} e^{\lambda t_{\frac{1}{2}}} &= 2 \\ \lambda t_{\frac{1}{2}} &= \ln 2 \\ &\approx 0.693 \end{aligned}$$

The half-life of an isotope and the decay constant are inversely proportional to each other. That is:

$$\begin{aligned} \lambda &= \frac{\ln 2}{t_{\frac{1}{2}}} \\ &= \frac{0.693}{t_{\frac{1}{2}}} \end{aligned}$$

Thus, if we know either $t_{\frac{1}{2}}$ or λ, we can calculate the other. For a nuclide with a very long half-life, we might not wish to sit around waiting to measure the half-life; it is easier to determine λ by measuring the activity (and using $A = \lambda N$) and use that to determine $t_{\frac{1}{2}}$.

Note that the units of λ and $t_{\frac{1}{2}}$ must be compatible; for example, λ in s^{-1} and $t_{\frac{1}{2}}$ in s.

KEY EQUATION

Half-life and decay constant are related as follows:

$$\begin{aligned} \lambda &= \frac{\ln 2}{t_{\frac{1}{2}}} \\ &= \frac{0.693}{t_{\frac{1}{2}}} \end{aligned}$$

Questions

19 Figure 29.13 shows the decay of an isotope of caesium, ${}^{134}_{55}\text{Cs}$. Use the graph to determine the half-life of this nuclide in years, and hence find the decay constant in year^{-1}.

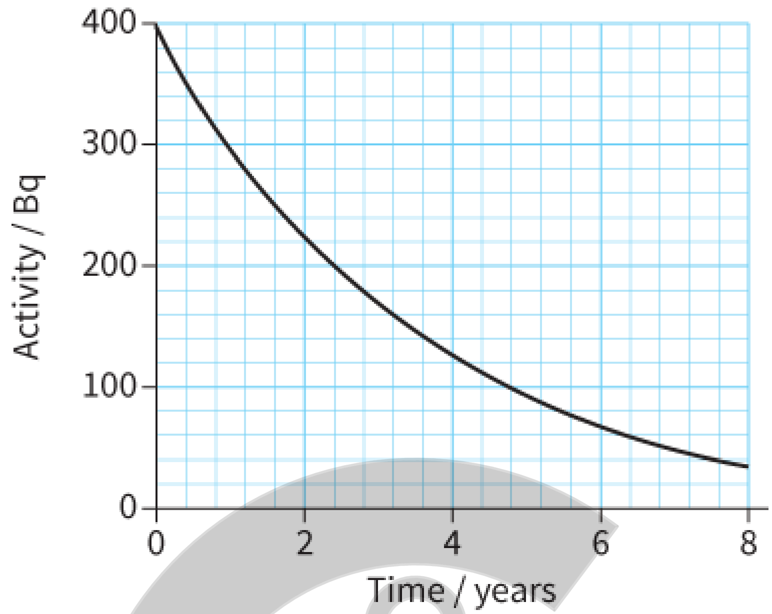


Figure 29.13: Decay graph for an isotope of caesium. For Question 19.

20 The decay constant of a particular isotope is $3.0 \times 10^{-4}\ \text{s}^{-1}$. Calculate how long it will take for the activity of a sample of this substance to decrease to one-eighth of its initial value.

21 The isotope ${}^{16}_{7}\text{N}$ decays with a half-life of 7.4 s.

a Calculate the decay constant for this nuclide.

b A sample of N initially contains 5000 nuclei. Calculate how many will remain after a time of:

i 14.8 s

ii 20.0 s.

22 A sample contains an isotope of half-life $t_{\frac{1}{2}}$.

a Show that the fraction *f* of nuclei in the sample that remain undecayed after a time *t* is given by the equation:

$$f = \left(\frac{1}{2}\right)^n \text{ where } n = \frac{t}{t_{\frac{1}{2}}}$$

b Calculate the fraction *f* after each of the following times:

i $t_{\frac{1}{2}}$

ii $2t_{\frac{1}{2}}$

iii $2.5t_{\frac{1}{2}}$

iv $8.3t_{\frac{1}{2}}$

REFLECTION

Without looking at your textbook, list all equations that contain decay constant λ.

What information can you get from the gradient of a graph of *N* against *t*?

Have a competition with a classmate. Use the internet for about 5 mins to find an isotope with the shortest half-life and the longest half-life.

What did this competition reveal about you as a learner?

SUMMARY

Nuclear reactions can be represented by equations of the form:

$$^{14}_{6}C \rightarrow ^{14}_{7}N + ^{0}_{-1}e$$

Einstein's mass–energy equation $\Delta E = \Delta mc^2$ relates mass changes to energy changes.

The mass defect is equal to the difference between the mass of the separate nucleons and that of the nucleus.

The mass of nuclear particles may be measured in atomic mass unit (u), where:

$$1 \text{ u} \approx 1.660 \times 10^{-27} \text{ kg}$$

The binding energy of a nucleus is the minimum energy required to break up the nucleus into separate nucleons.

The binding energy per nucleon indicates the relative stability of different nuclides.

The variation of binding energy per nucleon shows that energy is released when light nuclei undergo fusion and when heavier nuclei undergo fission, because these processes increase the binding energy per nucleon and, hence, result in more stable nuclides.

Nuclear decay is a spontaneous and random process. This unpredictability means that count rates tend to fluctuate, and we have to measure average quantities.

The half-life $t_{\frac{1}{2}}$ of an isotope is the mean time taken for half of the active nuclei in a sample to decay.

The decay constant λ is the probability that an individual nucleus will decay per unit time interval.

The activity A of a sample is related to the number of undecayed nuclei in the sample N by: $A = \lambda N$

The decay constant and half-life are related by the equation:

$$\lambda = \frac{\ln 2}{t_{\frac{1}{2}}} = \frac{0.693}{t_{\frac{1}{2}}}$$

We can represent the exponential decrease of a quantity with time t by an equation of the form:

$$x = x_0 e^{-\lambda t}$$

where x can be activity A, count rate R or number of undecayed nuclei N.

EXAM-STYLE QUESTIONS

1 Which expression is correct for determining the energy (in electronvolt eV) produced from a mass change of 1 u? **[1]**

A $1.0 \times (3.00 \times 10^8)^2$

B $1.66 \times 10^{-27} \times (3.00 \times 10^8)^2$

C $1.66 \times 10^{-27} \times (3.00 \times 10^8)^2 \times 1.60 \times 10^{-19}$

D $1.66 \times 10^{-27} \times \frac{(3.00 \times 10^8)^2}{1.60 \times 10^{-19}}$

2 A student determines the half-life of an isotope to be 66 ± 5 s.

What is the absolute uncertainty in the decay constant? **[1]**

A $8.0 \times 10^{-4}\ s^{-1}$

B $1.1 \times 10^{-3}\ s^{-1}$

C $5.3 \times 10^{-2}\ s^{-1}$

D $7.6 \times 10^{-2}\ s^{-1}$

3 An antiproton is identical to a proton except that it has negative charge. When a proton and an antiproton collide, they are annihilated and two photons are formed. In annihilation, all the mass of the particles is converted into energy.

a Calculate the energy released in the reaction. **[3]**

b Calculate the energy released if 1 mole of protons and 1 mole of antiprotons were annihilated by this process. **[3]**

(Mass of a proton = mass of an antiproton = 1.67×10^{-27} kg.)

[Total: 6]

4 Calculate the mass that would be annihilated to release 1 J of energy. **[2]**

5 In a nuclear reactor, the mass converted to energy takes place at a rate of 70 μg s^{-1}. Calculate the maximum power output from the reactor assuming that it is 100% efficient. **[3]**

6 The equation shows the radioactive decay of radon-222.

$$^{222}_{86}Rn \rightarrow {}^{218}_{84}Po + {}^{4}_{2}\alpha + \gamma$$

Calculate the total energy output from this decay and state what forms of energy are produced. **[6]**

(Mass of $^{222}_{86}Rn$ = 221.970 u, mass of $^{222}_{84}Po$ = 217.963 u, mass of $^{4}_{2}\alpha$ = 4.002u, 1 u is the unified atomic mass unit = 1.660×10^{-27} kg.)

(Hint: find the mass defect in u, then convert to kg.)

7 A carbon-12 atom consists of six protons, six neutrons and six electrons. The unified atomic mass unit (u) is defined as $\frac{1}{12}$ the mass of the carbon-12 atom.

Calculate:

a the mass defect in kilograms **[2]**

b the binding energy **[2]**

c the binding energy per nucleon. **[2]**

(Mass of a proton = 1.007 276 u, mass of a neutron = 1.008 665 u, mass of an electron = 0.000 548 u.)

[Total: 6]

8 The fusion reaction that holds most promise for the generation of electricity is the fusion of tritium ${}^{3}_{1}H$ and deuterium ${}^{2}_{1}H$. The following equation shows the process:

$${}^{3}_{1}H + {}^{2}_{1}H \rightarrow {}^{4}_{2}He + {}^{1}_{1}H$$

Calculate:

a the change in mass in the reaction **[3]**

b the energy released in the reaction **[2]**

c the energy released if one mole of deuterium were reacted with one mole of tritium. **[2]**

(Mass of ${}^{3}_{1}H$ = 3.015 500 u, mass of ${}^{2}_{1}H$ = 2.013 553 u, mass of ${}^{4}_{2}H$ = 4.001 50 u; mass of ${}^{1}_{1}H$ = 1.007 276 u.)

[Total: 7]

9 The initial activity a sample of 1 mole of radon-220 is 8.02×10^{21} s^{-1}. Calculate:

a the decay constant for this isotope **[3]**

b the half-life of the isotope. **[2]**

[Total: 5]

10 The graph of count rate against time for a sample containing indium-116 is shown.

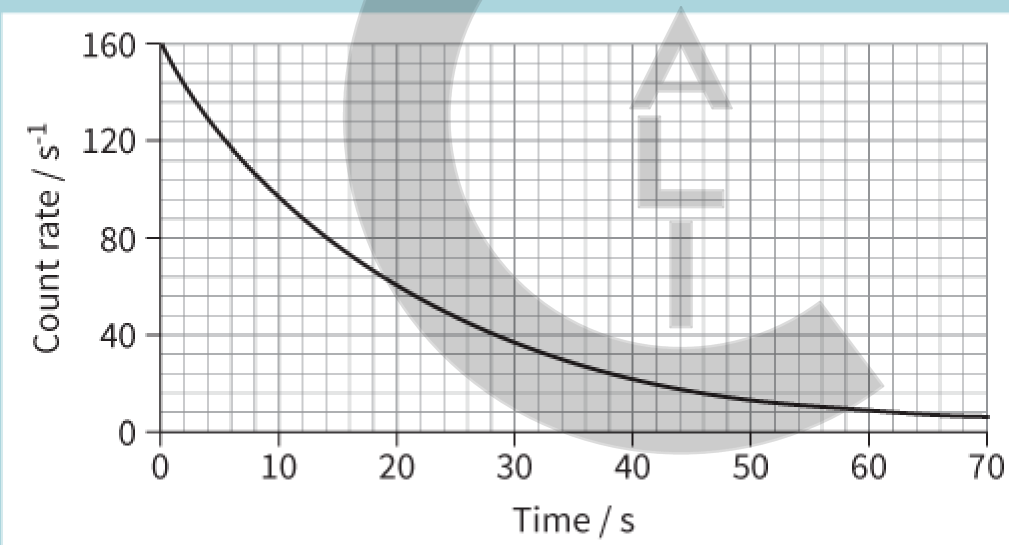


Figure 29.14

a Use the graph to determine the half-life of the isotope. **[2]**

b Calculate the decay constant. **[2]**

[Total: 4]

11 The proportions of different isotopes in rocks can be used to date the rocks. The half-life of uranium-238 is 4.9×10^{9} years. A sample has 99.2% of the proportion of this isotope compared with newly formed rock.

a Calculate the decay constant in y^{-1} for this isotope of uranium. **[2]**

b Calculate the age of the rock in years. **[3]**

[Total: 5]

12 The table shows the received count rate when a sample of the isotope vanadium-52 decays.

Time / min	0	1	2	3	4	5	6	7	8
Count rate / s^{-1}	187	159	134	110	85	70	60	56	40

Table 29.6

a **i** Sketch a graph of the count rate against the time. **[2]**

ii Comment on the scatter of the points. **[1]**

b From the graph, determine the half-life of the isotope. **[1]**

c Describe the changes to the graph that you would expect if you were given a larger sample of the isotope. **[2]**

[Total: 6]

13 This question is about the nucleus of uranium-235 ($^{235}_{92}U$), which has a mass of 3.89×10^{-25} kg.

a State the number of protons and neutrons in this nucleus. **[1]**

b The radius *r* of a nucleus is given by the equation:

$$r = 1.41 \times 10^{-15} A^{\frac{1}{3}}$$

where *A* is the nucleon number of the nucleus.

Calculate the density of the $^{235}_{92}U$ nucleus. **[3]**

c Explain why the total mass of the nucleons is different from the mass of the U nucleus. **[2]**

d Without calculations, explain how you can determine the binding energy per nucleon for the uranium-235 nucleus from its mass and the masses of a proton and a neutron. **[4]**

[Total: 10]

14 **a** Explain what is meant by **nuclear fusion** and explain why it only occurs at very high temperatures. **[3]**

b The main reactions that fuel the Sun are the fusion of hydrogen nuclides to form helium nuclides. However, other reactions do occur. In one such reaction, known as the triple alpha process, three helium nuclei collide and fuse to form a carbon-12 nucleus.

i Explain why temperatures higher than those required for the fusion of hydrogen are needed for the triple alpha process. **[1]**

ii Calculate the energy released in the triple alpha process. **[3]**

(Mass of a helium ($^{4}_{2}He$) nucleus = 4.001 506 u, mass of a carbon ($^{12}_{6}C$) nucleus = 12.000 000 u, 1 u = 1.660×10^{-27} kg.)

[Total: 7]

15 The isotope of polonium, $^{218}_{81}Po$, decays by the emission of an α-particle with a half-life of 183 s.

a In an accident at a reprocessing plant some of this isotope, in the form of dust, is released into the atmosphere.

Explain why a spillage in the form of a dust is far more dangerous to health than a liquid spillage. **[2]**

b It is calculated that 2.4 g of the isotope is released into the atmosphere. The molar mass of polonium is 218 g mol^{-1}.

Calculate the initial activity of the released polonium. **[4]**

c It is felt that it would safe to re-enter the laboratory when the activity falls to background, about 10 Bq.

Calculate how many hours must pass before it is safe to re-enter the laboratory. **[3]**

[Total: 9]

16 A nuclear reactor is fuelled by fission of uranium. The output from the reactor is 200 MW. The following equation describes a typical fission reaction:

$$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{236}_{92}U \rightarrow ^{87}_{35}Br + ^{146}_{57}La + 3^{1}_{0}n$$

a State and explain into what form the majority of the energy released in the reaction is transformed. **[2]**

b **i** Calculate the energy released in the reaction. The kinetic energy of the captured neutron is negligible. **[2]**

ii Assume that the energy released in this fission is typical of all fissions of U-236. Calculate how many fissions occur each second. **[1]**

iii Calculate the mass of uranium-235 that is required to run the reactor for 1 year. **[3]**

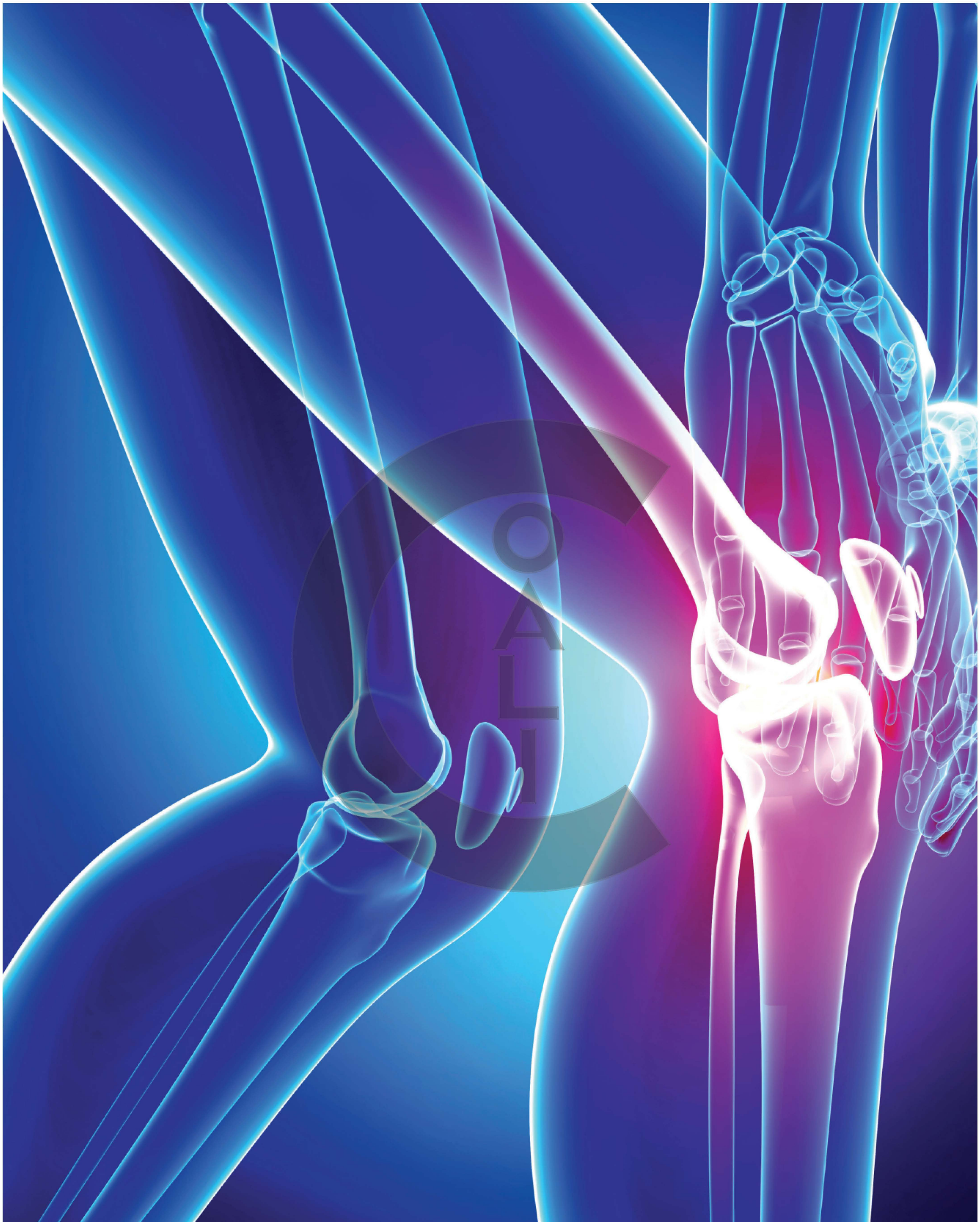
(Mass of $^{235}_{92}U$ = 3.90 × 10^{-25} kg, mass of $^{87}_{35}Br$ = 1.44 × 10^{-25} kg, mass of $^{146}_{57}La$ = 2.42 × 10^{-25} kg, mass of neutron = 1.67 × 10^{-27} kg, 1 year = 3.15 × 10^{7} s, molar mass of uranium-235 = 235 g mol^{-1}.)

[Total: 8]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand Einstein's mass-energy equation: $E = mc^2$	29.2			
understand the terms mass defect and binding energy	29.2, 29.4			
understand the significance of the binding energy per nucleon against nucleon number graph	29.4			
understand fission and fusion	29.4			
calculate the energy released in nuclear reactions using $\Delta E = \Delta mc^2$	29.3			
understand that radioactive decay is both spontaneous and random	29.5			
understand the terms activity, decay constant and half-life	29.6, 29.7			
use the equations: $A = \lambda N$ and $\lambda = \frac{0.693}{t_{\frac{1}{2}}}$ and $x = x_0 e^{-\lambda t}$	29.6, 29.7, 29.8			
understand the exponential decay of activity, undecayed nuclei and count rate.	29.6, 29.7			



〉Chapter 30

Medical imaging

LEARNING INTENTIONS

In this chapter you will learn how to:

- explain how X-ray beams are produced and controlled
- explain how ultrasound is produced and detected
- explain how ultrasound images are produced, revealing internal structures
- describe how conventional and CT scan X-ray images are produced
- explain the principles of positron emission tomography.

BEFORE YOU START

This chapter brings together many strands that you have met during the course. Working with a partner, make a few key notes to remind yourself about the following: magnetic and electric fields, nuclear decay equations, matter and antimatter annihilation, energy and momentum of photons, exponential decay.

APPLYING PHYSICS

In this book, you have learned many important ideas from physics. You may have noticed that the same big ideas keep reappearing – for example, the idea of a field of force (magnetic, electric, gravitational), or the idea of energy transmitted as waves, or the idea that matter is made of particles with forces acting between them. This is an important characteristic of physics; ideas that are used in one area prove to be useful in another. Hopefully, you will see many of these connections now that you are approaching the end of your course.

Physics is also useful. It is applied in many areas of life. In this chapter, we look at one of these areas: medical imaging. This topic covers a range of techniques that doctors use to see inside our bodies. The best known is X-rays, good for showing up bones (Figure 30.1), and the subject of the first part of this chapter. The sections that follow will look at the physics behind two other medical diagnostic techniques: ultrasound scanning and PET scanning. In this chapter, you will make use of several important aspects of physics that you have studied earlier in the course, including sound as a wave, electromagnetic radiation, the behaviour of charged particles, magnetic fields and the annihilation of matter and antimatter.

Many modern techniques use ionising radiation (X-rays, γ-rays). It is well known that ionising radiation affects living tissue. Why is it justifiable to expose patients to this radiation knowing that there is the potential for harm? Why is it justifiable to expose radiographers to the radiation? What precautions are taken?

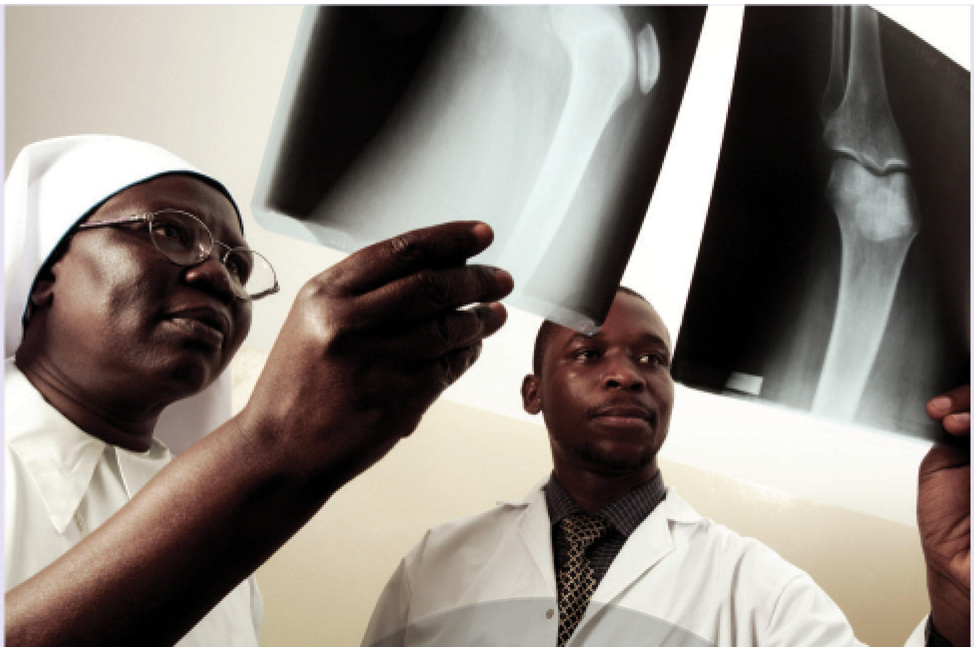


Figure 30.1: A radiographer and a doctor examine X-ray images of a patient's leg.

30.1 The nature and production of X-rays

X-rays are a form of electromagnetic radiation. They belong to the short-wavelength, high-frequency end of the electromagnetic spectrum, beyond ultraviolet radiation (Figure 30.2).

X-rays have wavelengths in the range 10^{-8} m to 10^{-13} m and are effectively the same as gamma-rays (γ-rays); the difference is the way they are produced:

- X-rays are produced when fast-moving electrons are rapidly decelerated. As the electrons slow down, their kinetic energy is transformed to photons of electromagnetic radiation.
- γ-rays are produced by radioactive decay. Following alpha (α) or beta (β) emission, a gamma photon is often emitted by the decaying nucleus (see Chapter 15).

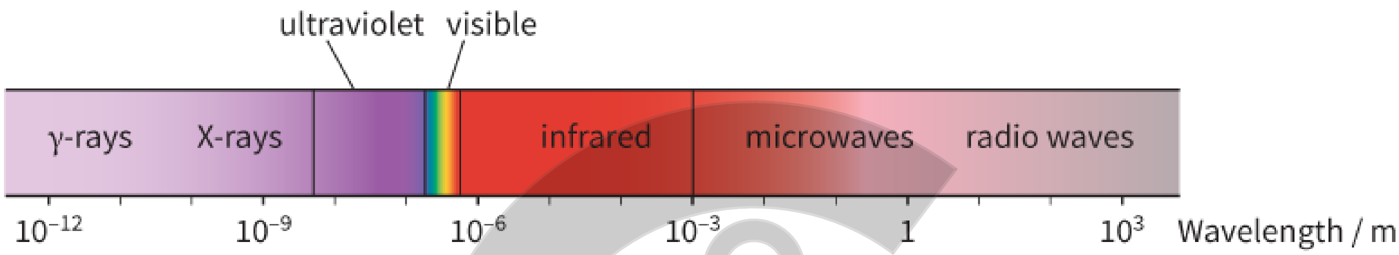


Figure 30.2: The electromagnetic spectrum; X-rays and γ-rays lie at the high-frequency, short-wavelength end of the spectrum.

As with all electromagnetic radiation, we can think of X-rays either as waves or as photons (see Chapter 28). X-rays travel in straight lines through a uniform medium.

X-ray tube

Figure 30.3a shows a patient undergoing a pelvic X-ray to check for bone degeneration. The X-ray machine is above the patient; it contains the X-ray tube that produces the X-rays that pass downwards through the patient's body.

In the early years of X-ray diagnosis the pictures were captured on photographic film. Increasingly, the images (as shown in Figure 30.3b) are detected electronically. This has the major advantages of being able to view the image immediately and also the images being rapidly sent via the internet (or a hospital intranet) to other doctors for second opinions.

Figure 30.4 shows the principles of the modern X-ray tube. The tube itself is evacuated, and contains two electrodes:

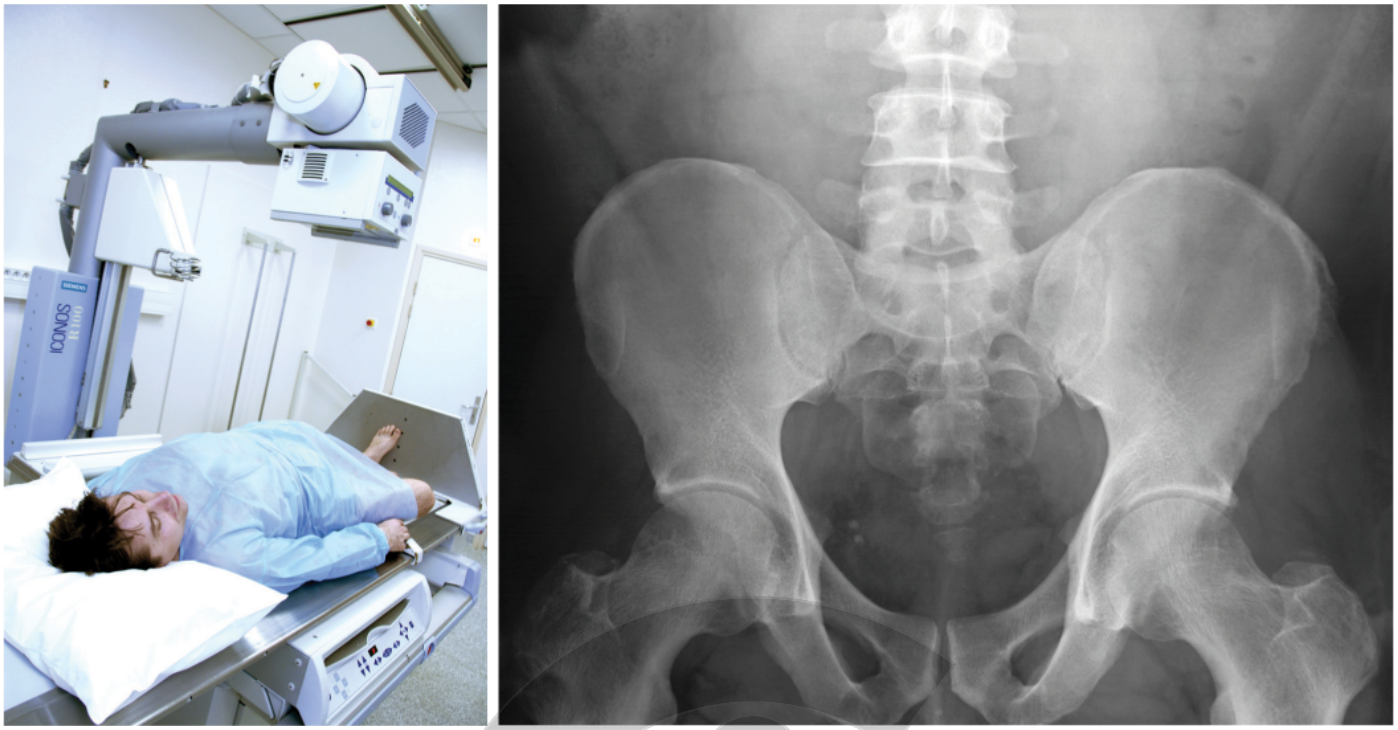


Figure 30.3: **a** A general-purpose X-ray system. **b** A typical X-ray image produced by such a machine, showing the region around the pelvis.

- Cathode – the heated filament acts as the cathode (negative) from which electrons are emitted.
- Anode – the rotating anode (positive) is made of a hard metal such as tungsten. (The anode metal is often referred to as the 'target metal'.)

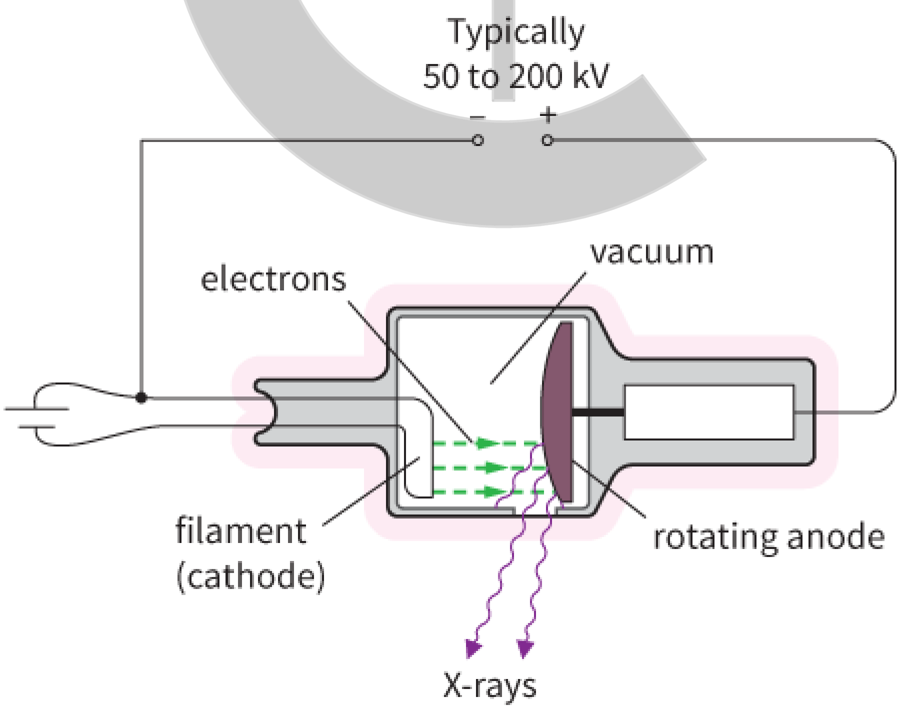


Figure 30.4: A simplified diagram of an X-ray tube.

An external power supply produces a voltage of up to 200 kV between the two electrodes. This accelerates a beam of electrons across the gap between the cathode and the anode. The kinetic energy of an electron arriving at the anode is 200 keV. When the electrons strike the anode at high speed, they lose some of their kinetic energy in the form of X-ray photons, which emerge in all directions. Part of the outer casing, the window, is thinner than the rest and allows X-rays to emerge into the space outside the tube. The width of the X-ray beam can be controlled using metal tubes beyond the window to absorb X-rays. This produces a parallel-sided beam called a **collimated beam**.

Only a small fraction, about 1%, of the kinetic energy of the electrons is converted to X-rays. Most of the incident energy is transferred to the anode, which becomes hot. This explains why the anode rotates; the region that is heated turns out of the beam so that it can cool down by radiating heat to its surroundings. Some X-ray tubes have water circulating through the anode to remove this excess thermal energy.

When an electron strikes the anode, it will be attracted towards the nucleus of an atom in the anode. This will cause it to change both speed and direction–in other words, it decelerates. A decelerating electron (or any other charged particle) loses energy by emitting electromagnetic radiation. The result is a single X-ray photon or, more usually, several photons. The electron interacts with more nuclei until it has lost all its energy and it comes to a halt.

The energy E gained by the electron when it is accelerated through a potential difference of V between the cathode and the anode is given by $E = eV$. This is the maximum energy that an X-ray photon can have, and so the maximum X-ray frequency f_{max} that can be produced can be calculated using the formula $E = hf$. So:

$$f_{max} = \frac{eV}{h}$$

Questions

1 **a** Summarise the energy changes that take place in an X-ray tube.

b An X-ray tube is operated with a potential difference of 80 kV between the cathode and the tungsten anode. Calculate the kinetic energy (in electronvolts and joules) of an electron arriving at the anode. Estimate the impact speed of such an electron (assume that the electron is non-relativistic).

2 Determine the minimum wavelength of X-rays emitted from an X-ray tube operated at a voltage of 120 kV.

X-rays of a whole range of energies are produced. The lowest energy X-rays will not have sufficient energy to penetrate through the body, so will have no effect on the resulting image. However, they will contribute to the overall X-ray dose that the patient receives. These X-rays must be filtered out; this is done using aluminium absorbers across the window of the tube.

Controlling intensity

The **intensity** of an X-ray beam is a measure of the energy passing through unit area (see the next topic). To increase the intensity of a beam, the current in the X-ray tube must be increased. Since each electron that collides with the anode produces X-rays, a greater current (more electrons per second) will produce a beam of greater intensity (more X-ray photons per second). A more intense beam means that the X-ray image will be formed in a shorter time.

30.2 X-ray attenuation

As you can see if you look back to Figure 30.1, bones look white in an X-ray photograph. This is because they are good absorbers of X-rays, so that little radiation arrives at the photographic film to cause blackening. Flesh and other soft tissues are less absorbing, so the film is blackened. Modern X-ray systems use digital detectors instead of photographic films. The digital images are easier to process, store and transmit using computers.

X-rays are a form of ionising radiation; that is, they ionise the atoms and molecules of the materials they pass through. In the process, the X-rays transfer some or all of their energy to the material, and so a beam of X-rays is gradually absorbed as it passes through a material.

The gradual decrease in the intensity of a beam of X-rays as it passes through matter is called **attenuation**. We will now look at the pattern of attenuation of X-rays as they travel through matter.

Decreasing intensity

Given that intensity is the rate of energy transfer per unit cross-sectional area, we can see that intensity is related to power by the equation:

$$I = \frac{P}{A}$$

where P is power and A is the cross-sectional area normal to the radiation. The unit of intensity is $\mathrm{W\,m^{-2}}$.

The intensity of a collimated beam of X-rays decreases as it passes through matter. Picture a beam entering a block of material. Suppose that, after it has passed through 1 cm of material, its intensity has decreased to half its original value. Then, after it has passed through 2 cm, the intensity will have decreased to one quarter of its original value (half of a half), and then, after 3 cm, it will be reduced to one eighth. You should recognise this pattern $\left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\right)$ as a form of exponential decay.

We can write an equation to represent the attenuation of X-rays as they pass through a uniform material as follows:

$$I = I_0 \, e^{-\mu x}$$

where I_0 is the initial intensity (before absorption), x is the thickness of the material, I is the transmitted intensity and μ is the **attenuation** (or **absorption**) **coefficient** of the material. Figure 30.5 shows this pattern of absorption. It also shows that bone is a better absorber of X-rays than flesh; it has a higher attenuation coefficient. (The attenuation coefficient also depends on the energy of the X-ray photons.)

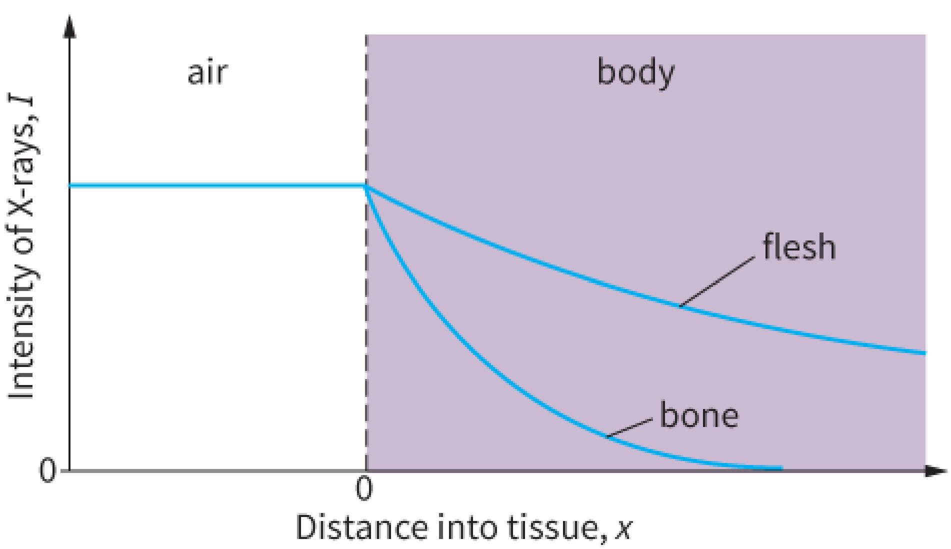


Figure 30.5: The absorption of X-rays follows an exponential pattern.

KEY EQUATION

$$I = I_0 \, e^{-\mu x}$$

Attenuation of X-rays as they pass through a uniform material.

The unit of the attenuation coefficient μ is m^{-1} (or cm^{-1} etc.).

Now look at Worked example 1.

Half thickness

If we compare the graphs (or equations) for the attenuation of X-rays as they pass through a material with the decay of a radioactive nuclide or with the discharge of a capacitor we see that they are all exponential decays. From Chapter 29, you should become familiar with the concept of the half-life of a radioactive isotope (the time taken for half the nuclei in any sample of the isotope to decay). In a similar manner, we refer to the half-thickness of an absorbing material. This is the thickness of material that will reduce the transmitted intensity of an X-ray beam of a particular frequency to half its original value.

WORKED EXAMPLE

1 The attenuation (absorption) coefficient of bone is 600 m^{-1} for X-rays of energy 20 keV. A beam of such X-rays has an intensity of 20 W m^{-2}. Calculate the intensity of the beam after passing through a 4.0 mm thickness of bone.

Step 1 Write down the quantities that you are given; make sure that the units are consistent.

I_0 = 20 W m^{-2}

x = 4.0 mm = 0.004 m

μ = 600 m^{-1}

Step 2 Substitute in the equation for intensity and solve.

Hint: *Calculate the exponent (the value of* $-\mu x$*) first.*

$$\begin{aligned} I &= I_0 e^{-\mu x} \\ &= 20 \times e^{-(600 \times 0.004)} = 20 \times e^{-2.4} \\ &= 1.8 \text{ W m}^{-2} \end{aligned}$$

So, the intensity of the X-ray beam will have been reduced to about 10% of its initial value after passing through just 4.0 mm of bone.

Questions

3 Use the equation $I = I_0 \, e^{-\mu x}$ to show that the half-thickness $x_{½}$ is related to the attenuation coefficient μ by:

$$x_{1/2} = \frac{\ln 2}{\mu}$$

4 An X-ray beam transfers 400 J of energy through an area of 5.0 cm^2 each second. Calculate its intensity in W m^{-2}.

5 An X-ray beam of initial intensity 50 W m^{-2} is incident on soft tissue of attenuation coefficient 1.2 cm^{-1}. Calculate its intensity after it has passed through a 5.0 cm thickness of tissue.

30.3 Improving X-ray images

The X-ray systems in use in hospitals and clinics today are highly developed pieces of technology. They do not simply show bones against a background of soft tissue. They can also show very fine detail in the soft tissue, including the arrangement of blood vessels.

Two aims of radiographers are:

- to reduce as much as possible the patient's exposure to harmful X-rays
- to improve the **contrast** of the image, so that the different tissues under investigation show up clearly in the image.

Reducing dosage

X-rays, like all ionising radiation, can damage living tissue, causing mutations that can lead to the growth of cancerous tissue. It is therefore important that the dosage is kept to a minimum.

A radiographer may choose to record the X-ray image on film or digitally. X-rays are only weakly absorbed by photographic film, so, historically, patients had to be exposed to long and intense doses of X-rays. Today, **intensifier screens** are used. These are sheets of a material that contains phosphor, a substance that emits visible light when it absorbs X-ray photons. The film is sandwiched between two intensifier screens. Each X-ray photon absorbed results in several thousand light photons, which then blacken the film. This reduces the patient's exposure by a factor of 100 to 500.

In digital systems, **image intensifiers** are also used (Figure 30.6). The incoming X-rays strike a phosphor screen, producing visible light photons. These then release electrons (by the photoelectric effect) from the photocathode. The electrons are accelerated and focused by the positively charged anode so that they strike a screen, which then gives out visible light. The image on this screen can be viewed via a television camera. At the same time, the image can be stored electronically. Digital systems have the advantage that images can be easily stored, shared and viewed.

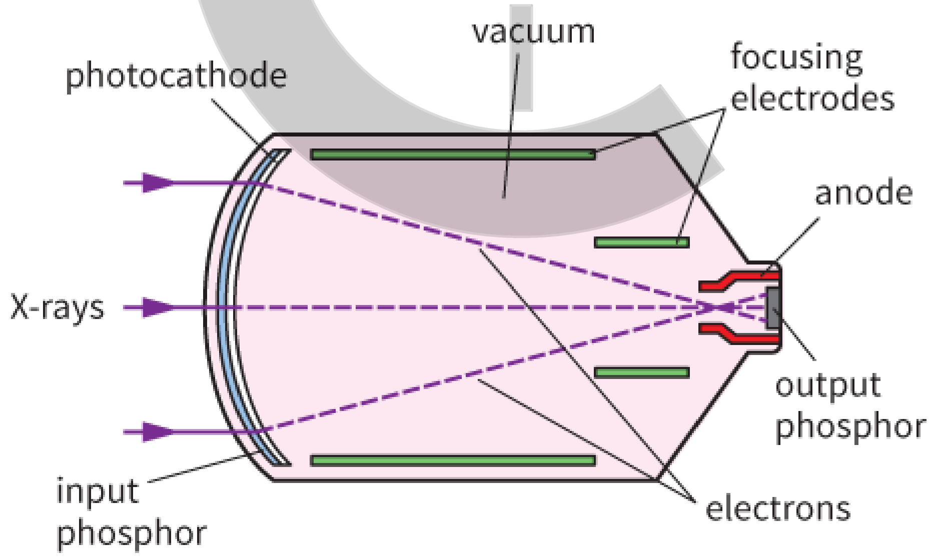


Figure 30.6: An X-ray image intensifier.

Image intensifiers are particularly useful in a technique called **fluoroscopy**. A continuous X-ray beam is passed through the patient onto a fluorescent screen where a real-time image is formed. Using an image intensifier ensures that the patient is not exposed to dangerous levels of X-rays over a long period.

Improving contrast

Good contrast is said to be achieved if there is a clear difference in the blackening of the photographic film as the X-ray passes through different types of tissue. The contrast is largely determined by the hardness of the X-rays. Bone is a good absorber of the radiation. If the doctor is diagnosing a break in a bone, he or she will use hard X-rays. In contrast, investigation of the tissue of the breast, where the tissue is a poor absorber, will require a longer exposure, using much softer (long-wavelength, low-frequency) X-rays.

As we have seen, different tissues show up differently in X-ray images. In particular, bone can readily be distinguished from soft tissue such as muscle because it is a good absorber of X-rays. However, it is often desirable to show up different soft tissues that absorb X-rays equally. In order to do this, **contrast media** are used.

A contrast medium is a substance, such as iodine or barium, which is a good absorber of X-rays. The patient may swallow a barium-containing liquid (a ‘barium meal’), or have a similar liquid injected into the tissue of interest. This tissue is then a better absorber of X-rays and its edges show up more clearly on the final image.

Figure 30.7 shows an X-ray image of the intestine of a patient who has been given a barium meal. The large pale areas show where the barium has accumulated. Other parts of the intestine have become smeared with barium, and this means that the outline of the tissue shows up clearly.

Contrast media are elements with high values of atomic number Z. This means that their atoms have many electrons with which the X-rays interact, so they are more absorbing. Soft tissues mostly consist of compounds of hydrogen, carbon and oxygen (low Z values), while bone has the heavier elements calcium and phosphorus, and contrast media have even higher Z values – see Table 30.1.

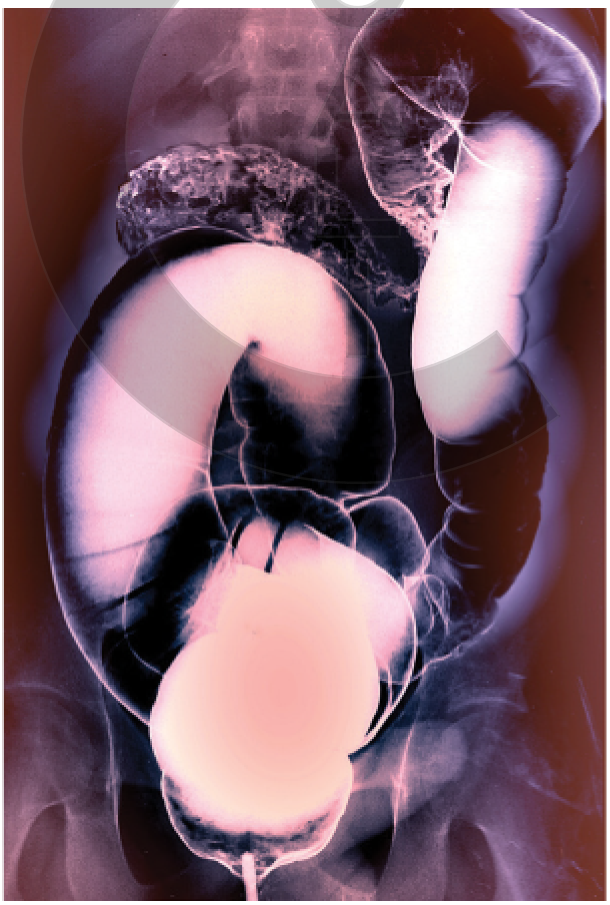


Figure 30.7: X-ray image of a patient’s intestine after taking a barium meal. Barium shows up as pale in this image, which has also been artificially coloured to highlight features of interest.

Substance	**Elements (*Z* values)**	**Average *Z***
soft tissue	H (1), C (6), O (8)	7
bone	H (1), C (6), O (8), P (15), Ca (20)	14
contrast media	I (53), Ba (56)	55

Table 30.1: Proton (atomic) numbers of the constituents of different tissues, and of contrast media.

Questions

6 The data in Table 30.2 shows how the attenuation coefficient μ depends on the energy of the X-rays in bone and muscle. When making a diagnostic X-ray image, it is desirable that bone should be clearly distinguished from muscle. Use the data in Table 30.2 to explain why it would be best to use lower energy (50 keV) X-rays for this purpose.

Maximum X-ray energy	**Bone: μ / cm^{-1}**	**Muscle: μ / cm^{-1}**
4.0 MeV	0.087	0.049
250 keV	0.32	0.16
100 keV	0.60	0.21
50 keV	3.32	0.54

Table 30.2: Data for Questions 7 and 8.

7 When low-energy X-rays are used, the attenuation coefficient μ is (roughly) proportional to the cube of the proton number Z of the absorbing material. Use the data in Table 30.2 to show that bone absorbs X-rays eight times as strongly as muscle.

30.4 Computerised axial tomography

A conventional X-ray image has an important limitation. Because an X-ray is essentially a two-dimensional shadow image, it shows the bones, organs, and so on at different depths within the body superimposed on each other. For example, in Figure 30.8, it is difficult to distinguish the bones of the front and back of the ribcage. This can be overcome by taking several images at different angles. An experienced radiographer can then study these images and deduce what is going on inside the patient.

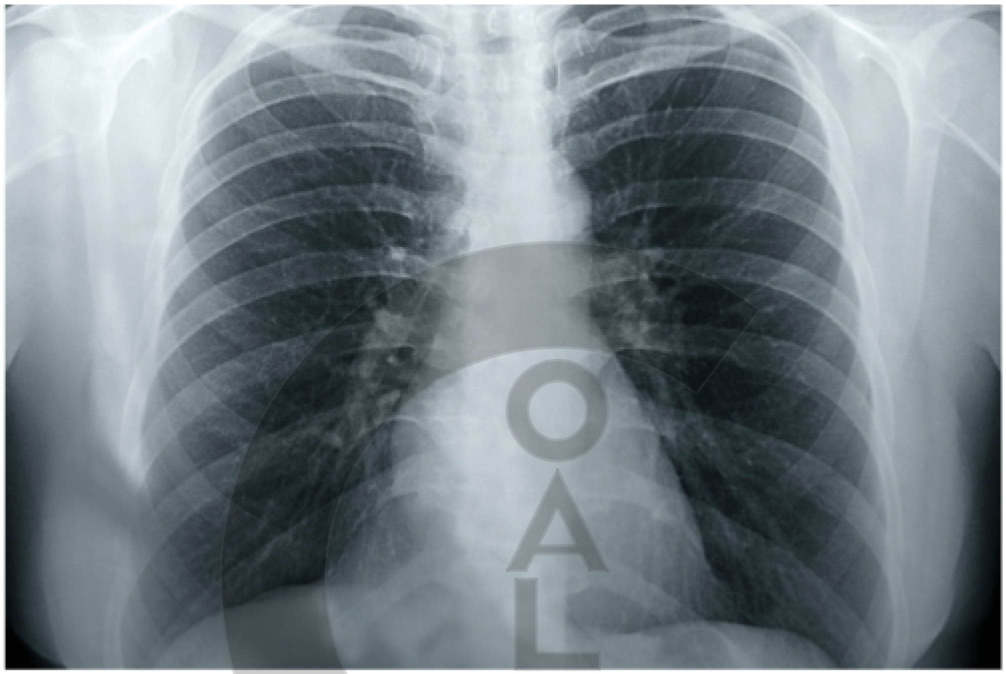


Figure 30.8: This chest X-ray shows the difficulty of distinguishing one bone from another when they overlap.

An ingenious technique for extending this approach was invented by Geoffrey Hounsfield and his colleagues at EMI in the UK in 1971. They developed the **computerised axial tomography** scanner (CAT scanner or CT scanner). Figure 30.9 illustrates the principle of a modern scanner.

- The patient lies in a vertical ring of X-ray detectors.
- The X-ray tube rotates around the ring, exposing the patient to a fan-shaped beam of X-rays from all directions.
- Detectors opposite the tube send electronic records to a computer.
- The computer software builds up a three-dimensional image of the patient.
- The radiographer can view images of 'slices' through the patient on the computer screen.

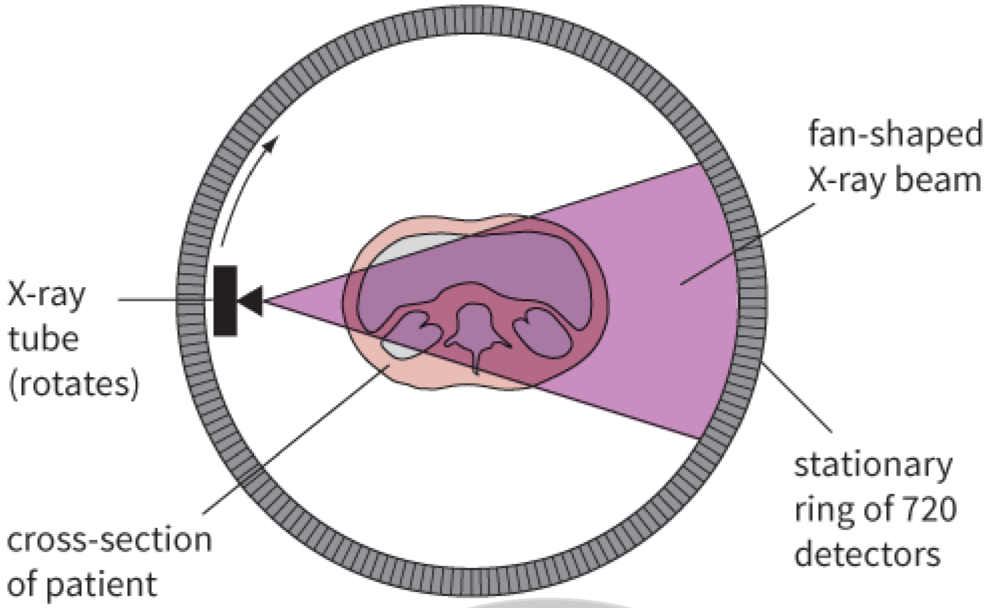


Figure 30.9: Operation of a modern CT scanner. The X-ray tube rotates around the patient while the detectors are stationary.

CT scanners have undergone many developments since they were first invented. In a fifth-generation scanner, the patient's bed slides slowly through the ring of detectors as the X-ray tube rotates. The tube thus traces out a spiral path around the patient, allowing information to be gathered about the whole body.

Figure 30.10 shows a child undergoing a CT scan. On the monitor you can see a cross-section of the patient's head.

Figure 30.10: A boy undergoes a CT scan in an investigation of an eye condition.

This technique is called **computerised axial tomography** because it relies on a computer to control the scanning motion and to gather and manipulate the data to produce images; because the X-ray tube rotates around an axis and because it produces images of slices through the patient – the Greek word **tomos** means **slice**.

Advantages of a CT scan

Although single X-ray images still have many uses (and they can be made very quickly), CT scans have a number of advantages:

- They produce images that show three-dimensional relationships between different tissues.
- They can distinguish tissues with quite similar densities (attenuation coefficients).

So, for example, a CT scan can show up the precise position, shape and size of a tumour. This allows it to be precisely targeted in treatment with high-energy X-rays or γ-rays.

However, it is worth noting that a CT scan involves using X-rays and any exposure to ionising radiation carries a risk for the patient. These risks are fairly small; it is estimated, with modern scanning equipment, that the radiation dose received is about one-third the dose received from background radiation in a year, or is equivalent to the dose received on four long-haul flights. Nevertheless, it is important to be aware of the dangers, particularly if there are other underlying health problems or if a woman is pregnant.

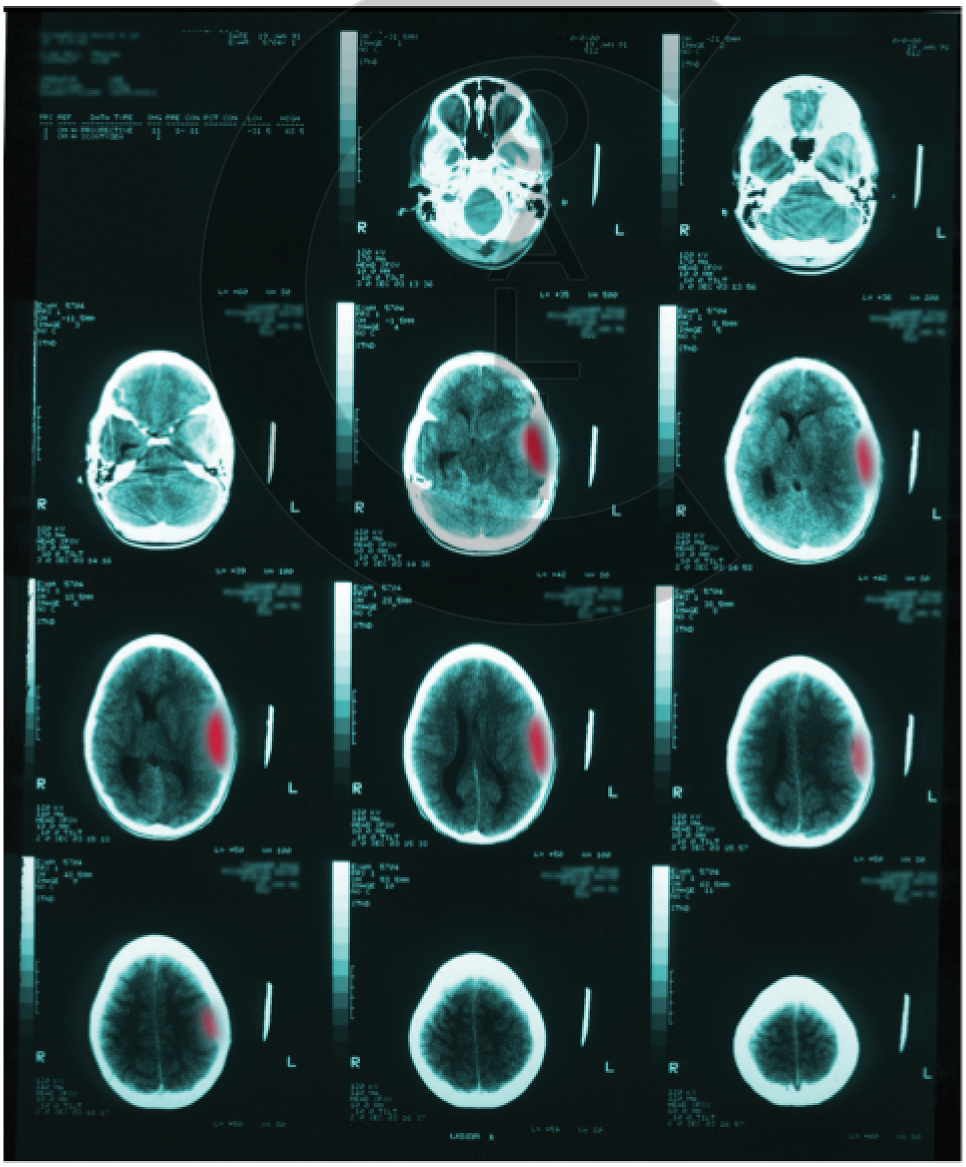


Figure 30.11: Sections through the head of a 10-year-old boy. You can see the haematoma (bruising) arising from being struck on the side of the head; this causes pressure on his brain.

Questions

8 Suggest why a patient may be asked to hold his or her breath during a CT scan.

9 A patient with an injury to the skull, perhaps as a result of a road accident, is likely to undergo a CT scan. Explain why a CT scan is preferable to a conventional X-ray in a case like this.

30.5 Using ultrasound in medicine

Ultrasound scanning is routinely used to check the condition of a baby in the womb (Figure 30.12). There do not seem to be any harmful side-effects associated with this procedure, and it can provide useful information on the baby's development. Indeed, for many children, their first appearance in the family photo album is in the form of an ante-natal (before birth) scan!

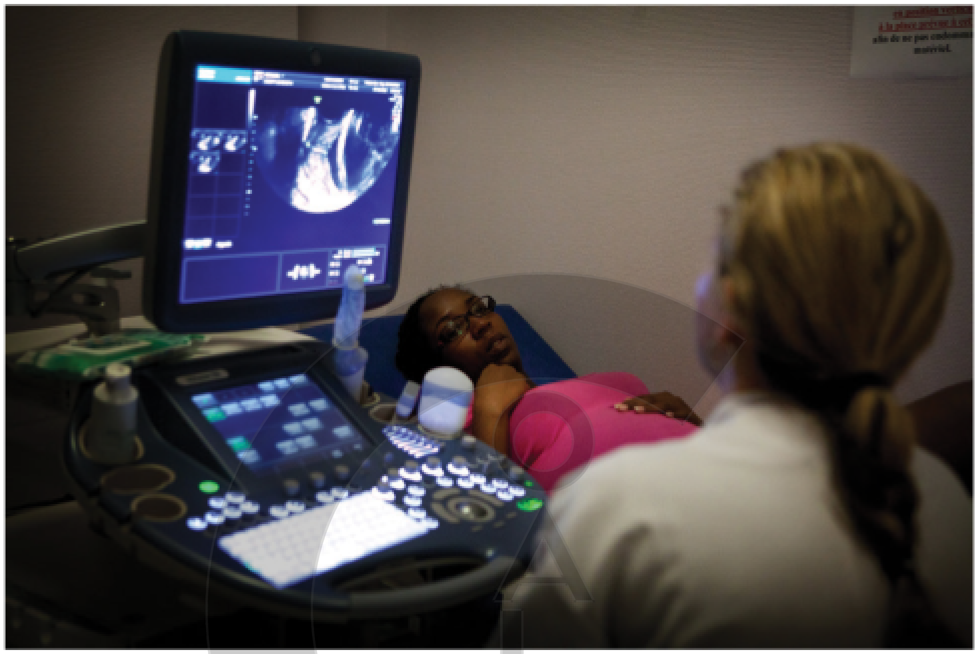


Figure 30.12: An expectant mother undergoes an ultrasound scan. The image of her baby is built up by computer and appears on the monitor.

This technique has many other uses in medicine. It can be used to investigate heart problems; the changes in frequency caused by the Doppler effect is used to investigate the flow of blood through the heart. Gallstones or kidney stones (two very painful complaints) can also be detected using ultrasound and so men as well as women may undergo this type of scan.

The technique of ultrasound scanning is rather similar to the way in which sailors use echo sounding and echo location to detect the seabed and shoals of fish. Ultrasound waves are directed into the patient's body. These waves are partially reflected at the boundaries between different tissues and the reflected waves are detected and used to construct the image.

Working with ultrasound

Ultrasound is any sound wave that has a frequency above the upper limit of human hearing. This is usually taken to mean frequencies above 20 kHz (20000 Hz), although the limit of hearing decreases with age to well below this figure. In medical applications, the typical frequencies used are in the megahertz range.

Sound waves are longitudinal waves. They can only pass through a material medium; they cannot pass through a vacuum. The speed of sound (and hence of ultrasound) depends on the material. In air, it is approximately 330 m s^{-1}; it is higher in solid materials. A typical value for body tissue is 1500 m s^{-1}. Using the wave equation $v = f\lambda$, we can calculate the wavelength of 2.0 MHz ultrasound waves in tissue:

$$\lambda = \frac{v}{f} = \frac{1500}{2.0 \times 10^6} = 7.5 \times 10^{-4}\ \text{m} \approx 1\text{mm}$$

This means that 2.0 MHz ultrasound waves will be able to distinguish detailed features whose dimensions are of the order of 1 mm. Higher-frequency waves have shorter wavelengths and these are used to detect smaller features inside the body. Unfortunately, higher-frequency waves are absorbed more strongly and so a more intense beam must be used.

Producing ultrasound

Like audible sound, ultrasound is produced by a vibrating source. The frequency of the source is the same as the frequency of the waves it produces. In ultrasound scanning, ultrasonic waves are produced by a varying electrical voltage in a transducer. The same device also acts as a detector. (You should recall from Chapter 25 that a transducer is any device that changes energy from one form to another.)

At the heart of the transducer is a **piezo-electric crystal**, such as quartz. This type of crystal has a useful property: when a voltage is applied across it in one direction, it shrinks slightly – see Figure 30.13a. When the voltage is reversed, it expands slightly. So, an alternating voltage with frequency *f* causes the crystal to contract and expand at the same frequency *f*. We say that the voltage induces a strain in the crystal. In the best piezo-electric substances, the maximum value of strain is about 0.1%; in other words, the crystal's width changes by about one part in a thousand.

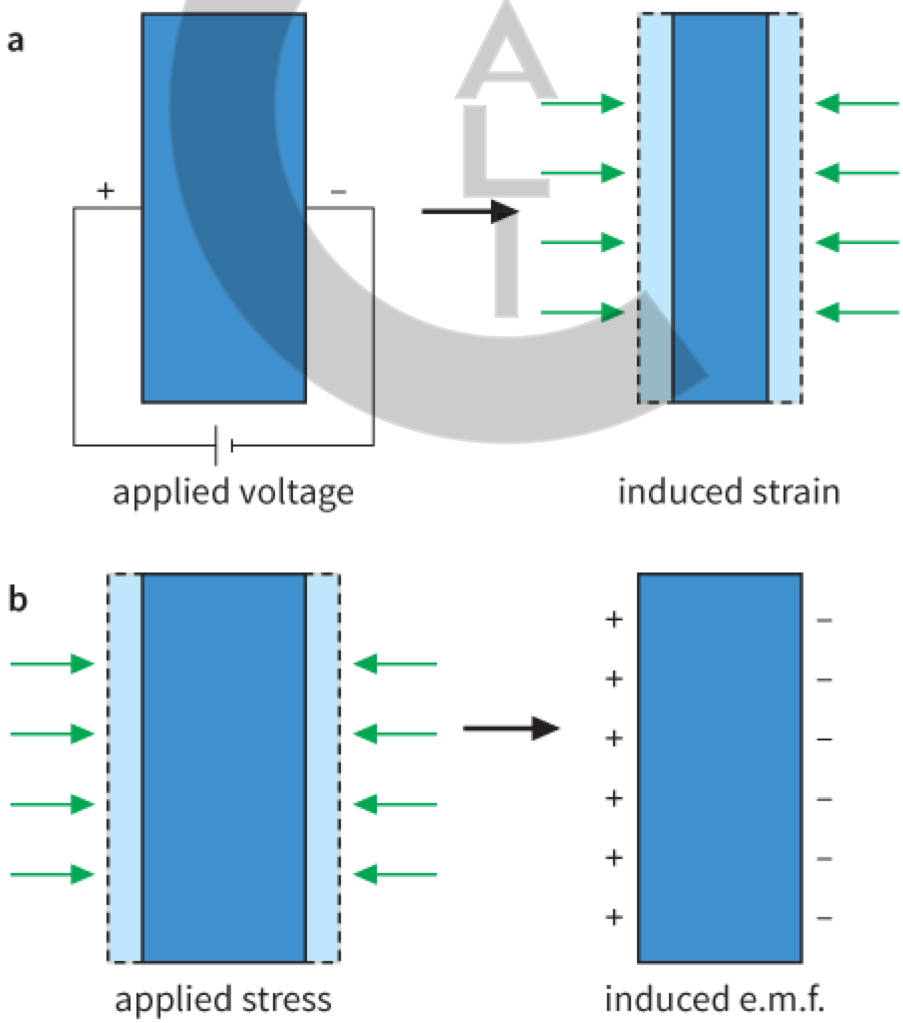


Figure 30.13: The **piezo-electric effect**. **a** An applied voltage causes a piezo-electric crystal to contract or expand. **b** An applied stress causes an induced e.m.f. across the crystal.

In a piezo-electric transducer, an alternating voltage is applied across the crystal, which then acts as the vibrating source of ultrasound waves. A brief pulse of ultrasound waves is sent into the patient's body; the transducer then receives an extended pulse of reflected ultrasound waves.

Detecting ultrasound

The transducer also acts as the detector of reflected ultrasound waves. It can do this because the piezo-electric effect works in reverse: a varying stress applied to the crystal produces a varying e.m.f. across the crystal – see Figure 30.13b. To maximise the effect, the frequency of the waves must match the resonant frequency of the crystal.

The optimum size of the crystal is half the wavelength $\left(\frac{\lambda}{2}\right)$ of the ultrasound waves.

Figure 30.14 shows the construction of a piezo-electric ultrasound transducer. Note the following features:

- The crystal is now usually made of polyvinylidene difluoride. Previously, quartz and lead zirconate titanate were used.
- The outer case supports and protects the crystal.
- At the base is the acoustic window, made from a material that is a good transmitter of ultrasound.
- Behind the crystal is a large block of damping material (usually epoxy resin). This helps to stop the crystal vibrating when a pulse of ultrasound has been generated. This is necessary so that the crystal is not vibrating when the incoming, reflected ultrasound waves reach the transducer.

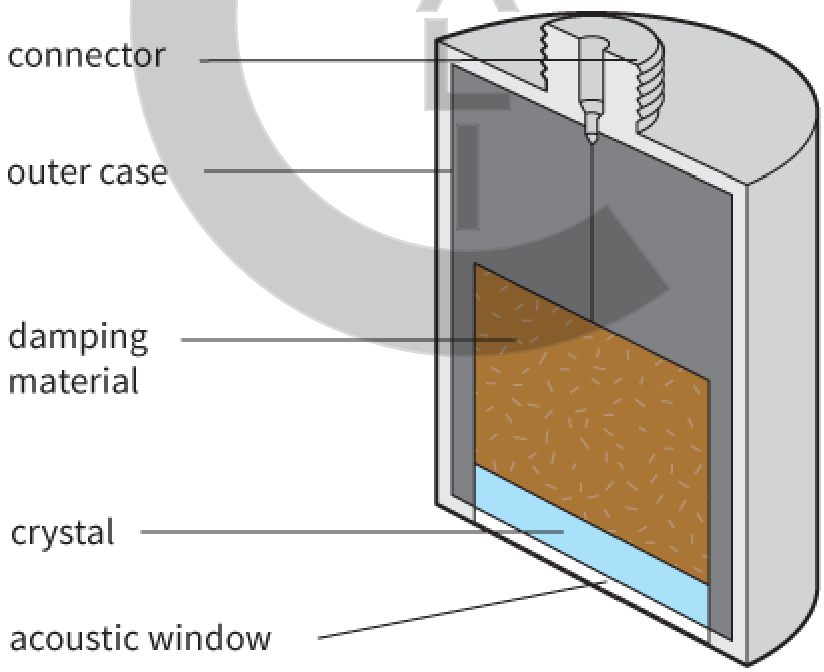


Figure 30.14: A section through an ultrasound transducer.

Questions

10 Quartz is an example of a piezo-electric material. The speed of sound in quartz is 5700 m s^{-1}.

a Calculate the wavelength of ultrasound waves of frequency 2.1 MHz in a quartz crystal.

b If the crystal is to be used in an ultrasound transducer, its thickness must be half a wavelength. Calculate the thickness of the transducer.

11 Piezo-electric crystals have many applications other than in ultrasound scanning. For example, they are used in:

a gas lighters (to produce a spark)

b inkjet printers (to break up the stream of ink into droplets)

c guitar pickups (to connect the guitar to an amplifier)

d the auto-focus mechanism of some cameras (to move the lens back and forth).

For each of these examples, state whether the piezo-electric effect is being used to transfer energy in the vibrations of the crystal to electrical energy or the other way round.

30.6 Echo sounding

The principle of an ultrasound scan is to direct ultrasound waves into the body. These pass through various tissues and are partially reflected at each boundary where the wave speed changes. The reflected waves are then detected and used to construct an internal image of the body.

Figure 30.15 shows what happens when a beam of ultrasound reaches a boundary between two different media. The beam is partially refracted (that is, the transmitted beam has changed direction) and partially reflected. This diagram should remind you of the way in which a ray of light is refracted and reflected when it strikes the boundary between two media. It is the change in speed that causes the refraction of a wave.

For ultrasound, we are interested in the fraction of the incident intensity of ultrasound that is reflected at the boundary. This depends on the **acoustic impedance** *Z* of each material. This quantity depends on the density ρ and the speed of sound *c* in the material. Acoustic impedance is defined as follows:

$$\text{acoustic impedence} = \text{density} \times \text{speed of sound}$$

$$Z = \rho c$$

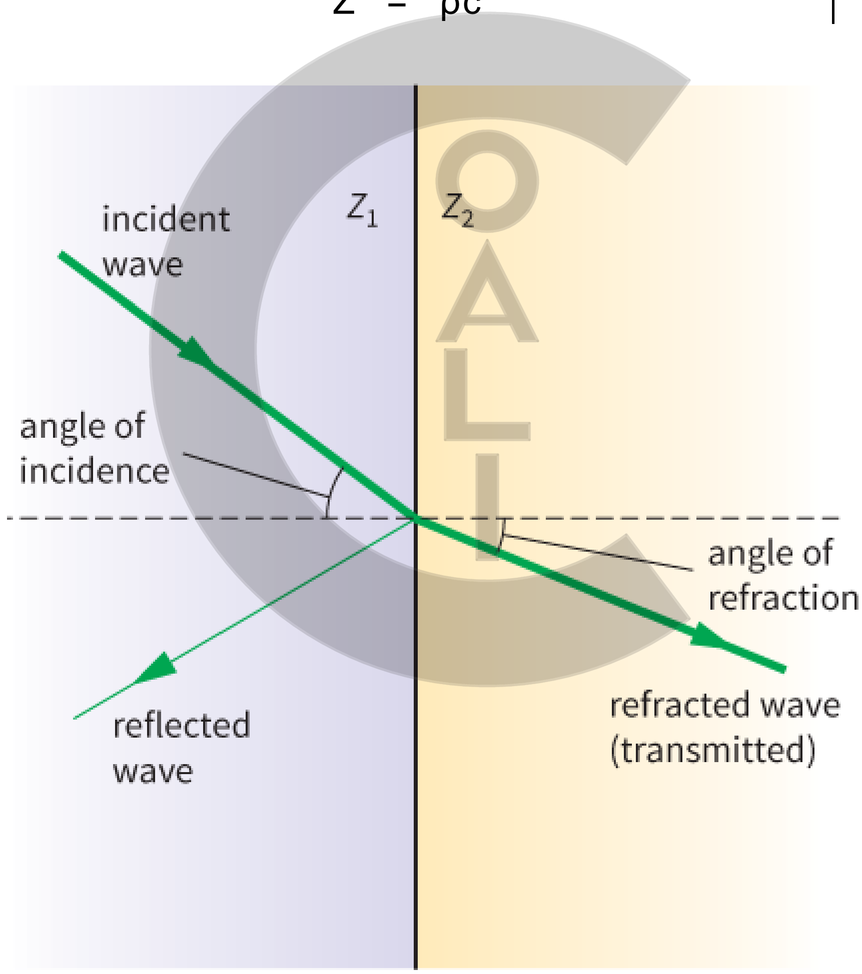


Figure 30.15: An ultrasound wave is both refracted and reflected when it strikes the boundary between two different materials.

where *Z* is the acoustic impedance of a material, ρ is the density of the substance and *c* is the speed of the ultrasound in the material.

Since the unit of density is kg m^{-3} and the unit of speed is m s^{-1}, the unit of acoustic impedance *Z* is kg m^{-2} s^{-1}.

KEY EQUATION

$$\text{acoustic impedence} = \text{density} \times \text{speed of sound}$$
$$Z = \rho c$$

Table 30.3 shows values of ρ, c and Z for some materials that are important in medical ultrasonography.

Material	Density / kg m^{-3}	Speed of sound / m s^{-1}	Acoustic impedance / 10^6 kg m^{-2} s^{-1}
air	1.3	330	0.0004
water	1000	1500	1.50
Biological			
blood	1060	1570	1.66
fat	925	1450	1.34
soft tissue (average)	1060	1540	1.63
muscle	1075	1590	1.71
bone (average; adult)	1600	4000	6.40
Transducers			
barium titanate	5600	5500	30.8
lead zirconate titanate	7650	3790	29.0
quartz	2650	5700	15.1
polyvinylidene difluoride	1780	2360	4.20

Table 30.3: The density (ρ), speed of sound in air (c) and acoustic impedance (Z) of some materials important in medical scanning.

Calculating reflected intensities

When an ultrasound beam reaches the boundary between two materials, the greater the **difference** in acoustic impedances, the greater the reflected fraction of the ultrasound waves. For normal incidence (that is, angle of incidence = 0°) the ratio of the reflected intensity I_r to the incident intensity I_0 is given by:

$$\frac{I_r}{I_0} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2}$$
$$\Rightarrow \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2$$

where I_0 is the intensity of the incident ultrasonic beam, I is the intensity of the reflected beam, and Z_1 and Z_2 are the acoustic impedances of the two materials (see Figure 30.16).

The ratio $\frac{I_r}{I_0}$ indicates the fraction of the intensity of the beam that is reflected.

KEY EQUATION

$$\frac{I_r}{I_0} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2} \Rightarrow \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2$$

Intensity reflection fraction of the boundary between two materials.

Comparing acoustic impedances

A big change in acoustic impedance gives a large fraction of reflected intensity. Inspection of Table 30.3 shows that:

- a very large fraction $\left(\frac{I_r}{I_0} \approx 99.95\%\right)$ of the incident ultrasound will be reflected at an air–tissue boundary
- a large fraction will be reflected at a tissue–bone boundary (as shown in Worked example 2)
- very little will be reflected at a boundary between soft tissues including fat and muscle.

This means that bone shows up well in an ultrasound scan, but it is difficult to see different soft tissues (Figure 30.16). Another problem is that the patient's skin is in contact with air, and 99.95% of the ultrasound will be reflected before it has entered the body. To overcome this, the transducer must be 'coupled' to the skin using a gel whose impedance matches that of the skin. This process of **impedance matching** explains why the patient's skin is smeared with gel before a scan.

WORKED EXAMPLE

2 A beam of ultrasound is normally incident on the boundary between muscle and bone. Use Table 30.3 to determine the fraction of its intensity that is reflected.

Step 1 Write down the values of Z_1 (for muscle) and Z_2 (for bone).

$Z_1 = 1.71 \times 10^6$ kg m^{-2} s^{-1}

$Z_2 = 6.40 \times 10^6$ kg m^{-2} s^{-1}

Step 2 Substitute these values in the equation for $\frac{I_r}{I_0}$.

Hint: *We can use this equation because we know that the angle of incidence = 0°.*

$$\frac{I_r}{I_0} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2}$$

$$\Rightarrow \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2$$

$$= \left(\frac{6.40 - 1.71}{6.40 + 1.71}\right)^2$$

$$= \left(\frac{4.69}{8.11}\right)^2$$

$$= 0.578^2$$

$$= 0.33$$

Hint: *We can ignore the factor of 10^6 in the Z values because this is a factor common to all the values, so they cancel out.*

So, 33% of the intensity of ultrasound will be reflected at the muscle–bone boundary.

The acoustic impedance of the gel is typically 1.65×10^6 kg m^{-2} s^{-1} and that of skin is 1.71×10^6 kg m^{-2} s^{-1}. With gel between the skin and the transducer, the percentage of the intensity reflected is 0.03%.

The poor match of impedance between air and tissue means that ultrasound cannot penetrate the lungs. The operator must take care to avoid any bubbles of gas in the intestines. Bones are also difficult to see through. For an ultrasound scan of the heart, the probe must be directed through the gap between two ribs.

As ultrasound waves pass through the body, they are gradually absorbed. Their absorption follows the same exponential pattern as we saw earlier for X-rays. The intensity I decreases with distance x according to the equation:

$$I = I_0 e^{-\alpha x}$$

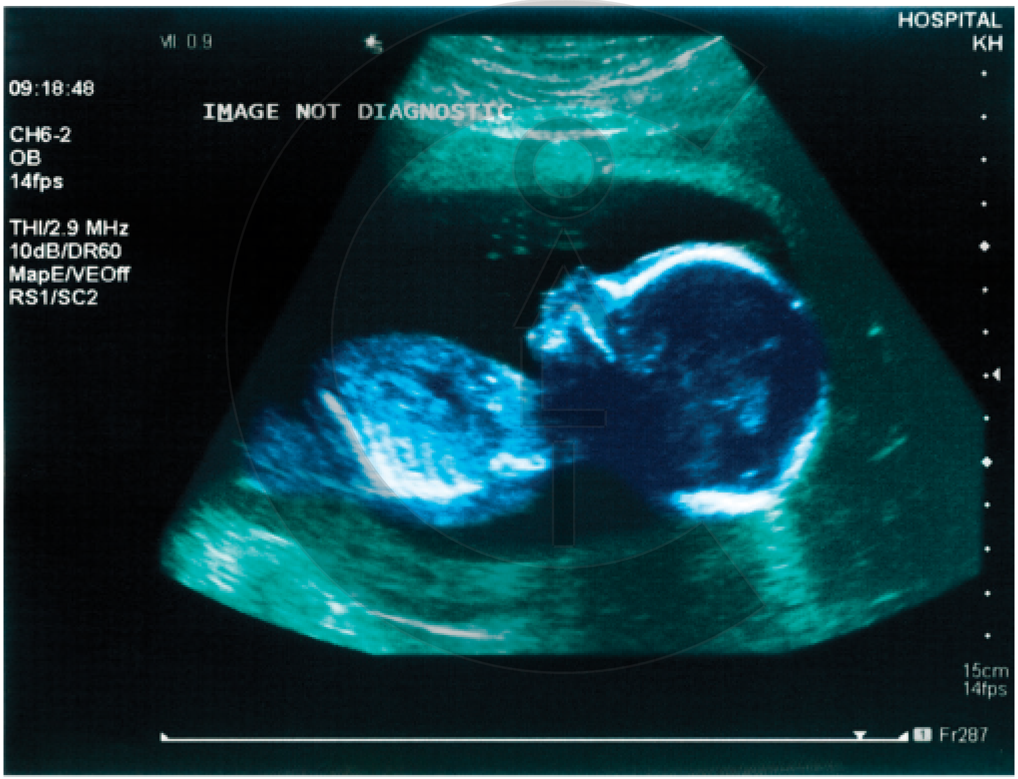


Figure 30.16: Ultrasound scan of a foetus at 20 weeks; the baby's skin is clearly visible, as are its bony skull and ribs.

where I_0 is the intensity of the incident ultrasonic beam, I is the intensity of the reflected beam, α is the absorption coefficient, and x is the distance travelled through the material.

Here, α is the absorption coefficient, equivalent to the quantity μ in the absorption equation for X-rays; its value varies with the nature of the tissue through which the ultrasound is passing, and with the frequency of the ultrasound. In practice, absorption is not a serious problem in an ultrasound scan as scanning relies on the reflection of ultrasound at the boundaries between different tissues.

KEY EQUATION

$$I = I_0 e^{-\alpha x}$$

Attenuation of ultrasound.

Questions

12 Calculate the acoustic impedance of muscle tissue. (Density = 1075 kg m^{-3}; speed of sound = 1590 m s^{-1}.)

13 Determine the fraction of the intensity of an ultrasound beam that is reflected when a beam is incident normally on a boundary between water and fat. (Use values from Table 30.3.)

14 The ultrasound image shown in Figure 30.25 clearly shows the baby's skin and some bones. Explain why these show up clearly while softer organs inside its body do not.

15 Explain why ultrasound cannot readily be used to examine the brain. Suggest one or more alternative scanning techniques that can be used for this.

30.7 Ultrasound scanning

There are several different types of ultrasound scan that are used in practice. To illustrate the basic principles, we will concentrate on the A-scan and the B-scan.

A-scan

This is the simplest type of scan. A pulse of ultrasound is sent into the body and the reflected 'echoes' are detected and displayed on an oscilloscope or computer screen as a voltage–time graph.

A pulse generator controls the ultrasound transducer. It is also connected to the time base of the oscilloscope. Simultaneously, the pulse generator triggers a pulse of ultrasound that travels into the patient and starts a trace on the screen. Each partial reflection of the ultrasound is detected and appears as a spike on the screen (see Figure 30.17).

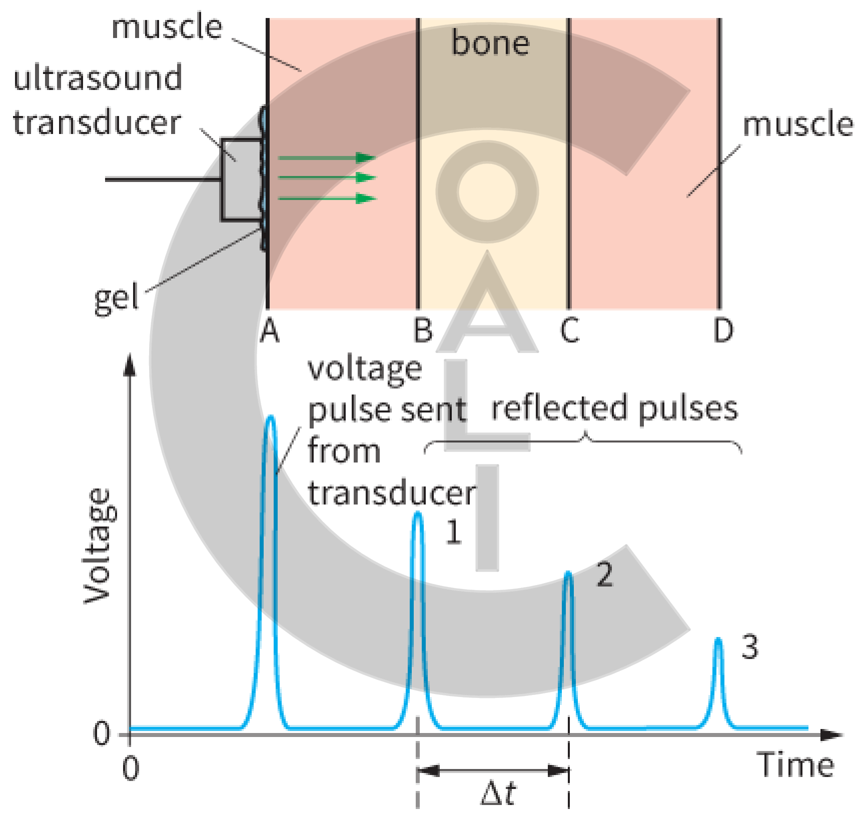


Figure 30.17: An A-scan. Information about the depth of reflecting tissues can be obtained from the positions of the spikes along the time axis; their relative amplitudes can indicate the nature of the reflecting surfaces.

In Figure 30.17, pulses 1, 2 and 3 are reflected at the various boundaries. Pulse 1 is the reflection at the muscle–bone boundary at B. Pulse 2 is the reflection at the bone–muscle boundary at C. The time Δt is the time taken for the ultrasound to travel **twice** the thickness of the bone. Finally, pulse 3 is the reflection at the muscle–air boundary at D. The thickness of the bone can be determined from this A-scan.

The time interval between pulses 1 and 2 = Δt

$$\text{thickness of bone} = \frac{\text{distance travelled by ultrasound}}{2} = \frac{c\Delta t}{2}$$

where *c* is the speed of the ultrasound in the bone (see Worked example 3).

WORKED EXAMPLE

3 In a particular A-scan, similar to Figure 30.26, the time interval between pulses 1 and 2 is 12 μs. The speed of ultrasound in bone is about 4000 m s^{-1}. Determine the thickness of the bone.

Step 1 Determine the distance travelled by the ultrasound in the time interval of 12 μs.

$$\text{distance} = \text{speed} \times \text{time}$$
$$\text{distance} = 4000 \times 12 \times 10^{-6}$$
$$= 4.8 \times 10^{-2}\ \text{m}$$

Step 2 Calculate the thickness of the bone.

Hint: *The distance you have just calculated must be halved because the ultrasound has to travel through the bone twice.*

$$\text{thickness of bone} = \frac{4.8 \times 10^{-2}}{2}$$
$$= 2.4 \times 10^{-2}\ \text{m} = 2.4\ \text{cm}$$

Because ultrasound waves are gradually attenuated as they pass through the body (their energy is absorbed so that their amplitude and intensity decrease), the echoes from tissues deeper in the body are weaker and must be amplified.

A-scans are used for some straightforward procedures such as measuring the thickness of the eye lens.

B-scan

In a B-scan, a detailed image of a cross-section through the patient is built up from many A-scans. The ultrasound transducer is moved across the patient's body in the area of interest. Its position and orientation are determined by small sensors attached to it.

Each reflected pulse is analysed to determine the depth of the reflecting surface (from the time of echo) and the nature of the surface (from the amplitude of the reflected wave). A two-dimensional image is then built up on a screen by positioning dots to represent the position of the reflecting surfaces and with brightness determined by the intensity of the reflection, brighter dots indicating more reflected ultrasound (see Figure 30.18).

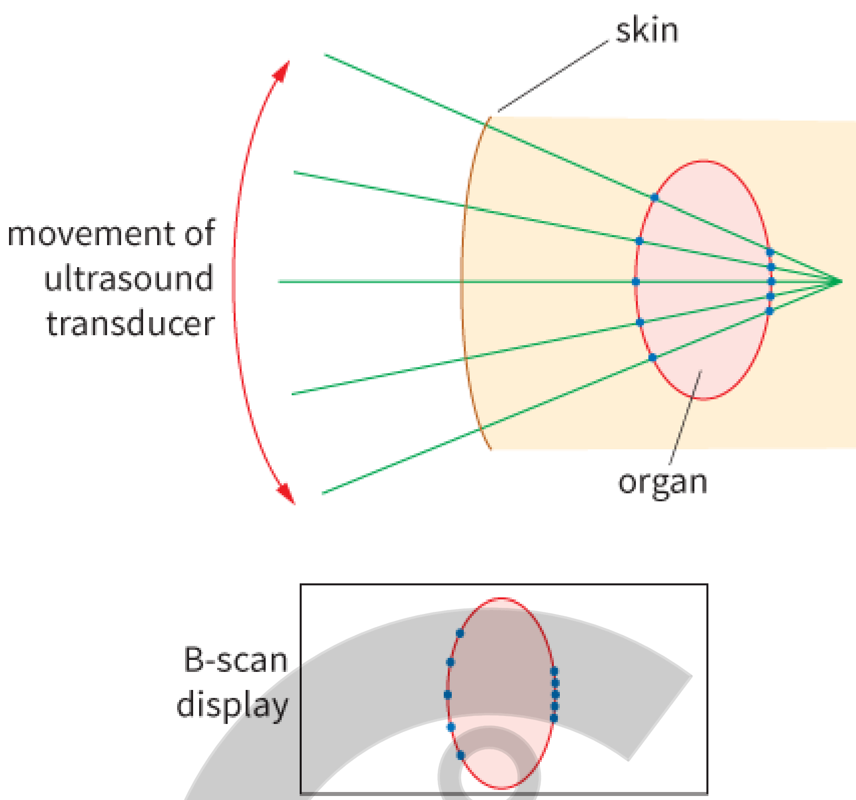


Figure 30.18: In a B-scan, dots are produced on the screen rather than pulses as in the A-scan. By moving the transducer, a series of dots on the screen traces out the shape of the organ being examined.

Figure 30.19 shows the result of a typical B-scan. Because it takes several seconds for the scanner to move across the body, problems can arise if the organs of interest are moving–this gives a blurred image.

Figure 30.19: An ultrasonic B-scan of an abnormal thyroid gland.

Questions

16 Two consecutive peaks in an ultrasound A-scan are separated by a time interval of 0.034 ms. Calculate the distance between the two reflecting surfaces. (Assume that the speed of sound in the tissue between the two surfaces is 1540 m s^{-1}.)

17 Explain why an ultrasound B-scan, rather than X-rays, is used to examine a foetus.

30.8 Positron Emission Tomography

Positron Emission Tomography or PET scanning is another tool in the diagnostic toolbox of modern medicine. It has a range of uses: investigating, diagnosing and monitoring treatment of cancers, heart disease, gastrointestinal disorders and brain function.

The principle operation of PET is different from CT and ultrasound scanning. CT and ultrasound look at the patient from the outside, whereas PET looks at the patient from the inside. A small amount of tracer, sometimes referred to as a **radiotracer**, is injected into a vein, travels round the body and is absorbed by organs and tissues. It is the radiation from this that is used to produce the image.

Radiotracers

There are several different radiotracers used in PET, an example being a glucose based molecule, onto which a radioactive-nuclide, fluorine-18, is attached. This substance is known as fluorodeoxyglucose. The fluorine-18 nuclide decays by emitting a β^+-particle, a positron. The advantage of using a glucose-based tracer is that it is taken up at different rates by different tissues or organs. Cancer cells are more metabolically active than surrounding healthy cells, consequently they absorb glucose at a higher rate and thus emit radiation at a greater rate. This will then appear on the screen as a bright area, allowing doctors to identify diseases and also determine the progress and effectiveness of any treatment used for the disease. PET scans are not only used for the detection of cancers but are a diagnostic tool in investigating blood flow, heart disease and brain injuries, and they are also being used to investigate Alzheimer's disease and other forms of dementia.

PET scans are unique in that they are able to pinpoint molecular activity within the patient's body, rather than looking at the body from outside. Consequently, they can identify disease in its earliest stages, meaning that there is a greater chance of successful treatment. They can also be used to track a patient's immediate and ongoing response to treatments.

What happens in positron emission?

PET scanners require a radioactive isotope that decays by β^+ emission, the emission of a positron, the antiparticle of the electron, which you met in Chapter 15. Most β^+ emitters are not naturally occurring isotopes and are made by firing protons at target nuclei.

The positron moves through the patient's tissue and within a very short distance (significantly less than a millimetre) it will encounter an electron. The pair will annihilate and their mass becomes pure energy in the form of two γ-rays that move apart in opposite directions. The concept of mass-energy is discussed in detail in Chapter 29.

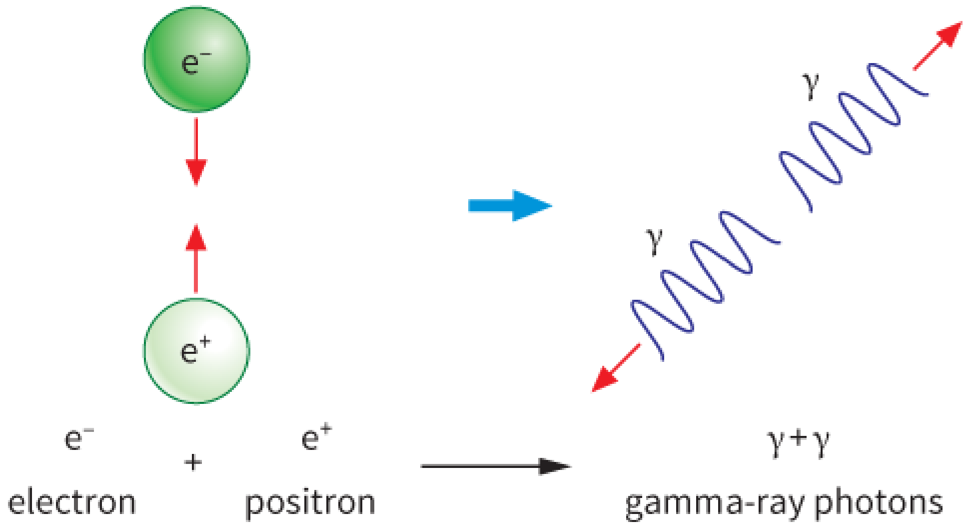


Figure 30.20: Energy is released in the annihilation of a positron and an electron.

In the annihilation process, as in all collisions, both mass-energy and momentum are conserved. The initial kinetic energy of the positron is small – negligible compared to their rest mass-energy – hence, the γ-ray photons have a specific energy and a specific frequency that are determined, solely, by the mass-energy of the positron–electron pair.

The energy of a photon is given by:

$$E = hf$$

where *h* is Planck's constant, and *f* is the frequency of the photon.

The momentum of a photon is given by:

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

where *c* is the speed of electromagnetic radiation in a vacuum.

KEY EQUATION

$$\text{energy of a photon} = E = hf$$

$$\text{momentum of a photon} = p = \frac{E}{c}$$

The production of suitable radioisotopes

You will remember from your work on magnetic fields that a charged particle entering a magnetic field at right angles to the field will travel in a circular path. The cyclotron works on this principle, however, the particles are continuously accelerated by an alternating electric field as they go round the circle, thus they travel in a spiral path before they are released and collide with the target nuclei. Figure 30.21 shows an early cyclotron. The principle of the cyclotron is shown in Figure 30.22.

In the production of fluorine-18, oxygen-18 nuclei are bombarded with protons and the following reaction takes place:

$$^{18}_{8}\mathrm{O} + ^{1}_{1}\mathrm{p} \rightarrow ^{18}_{9}\mathrm{Fl} + ^{1}_{0}\mathrm{n} + \gamma$$

The isotope, fluorine-18, has a half-life of just under two hours. This means the patient is not subjected to radiation for a long period of time. However, it also means that the radiotracer needs to be made up freshly, probably on site, to be most effective.

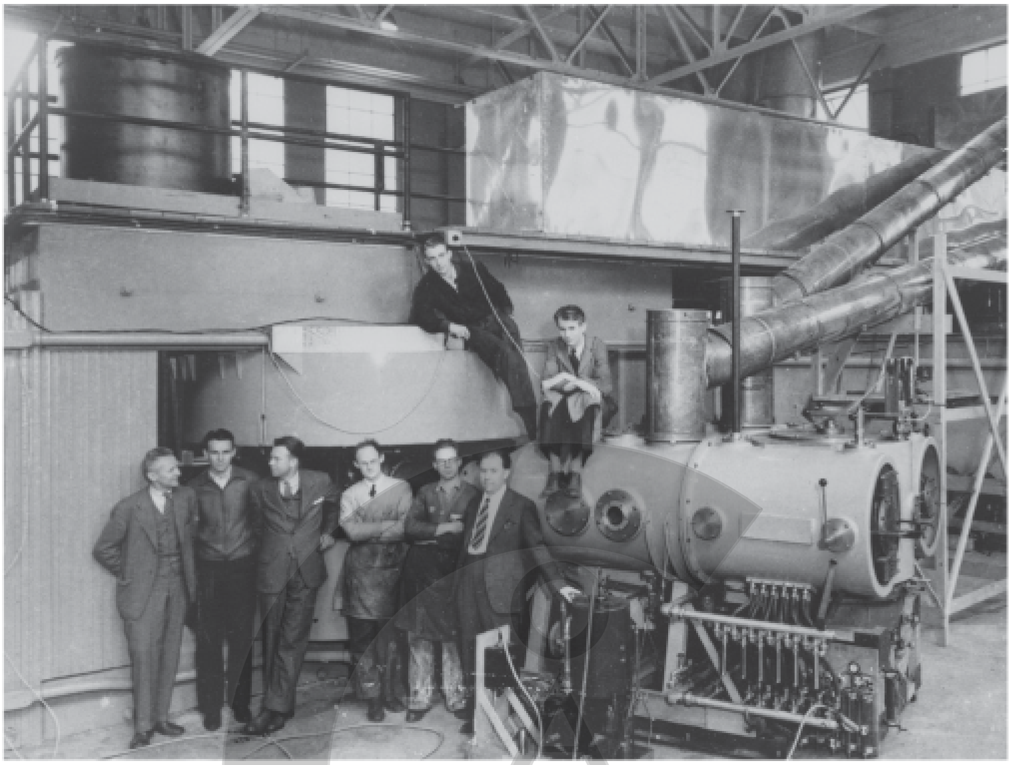


Figure 30.21: The cyclotron at the Lawrence Radiation Laboratory, Berkeley, soon after completion in 1939.

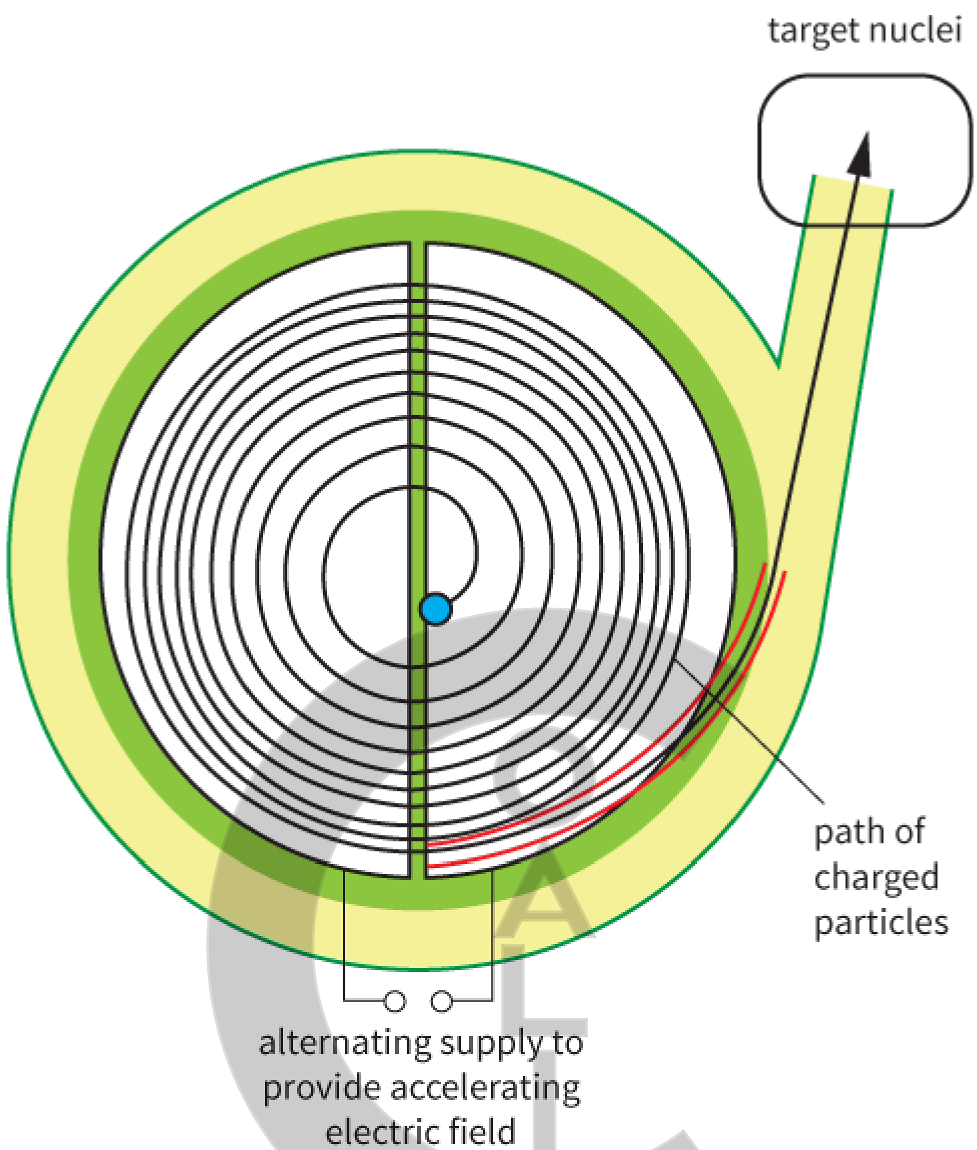


Figure 30.22: The principle of the cyclotron. Note, the magnetic field is into the plane of the page.

Questions

18 Suggest the reason why, in PET scanning, it is important that the positron meets an electron within a very short distance from its point of emission.

19 Explain why the γ-rays produced in positron–electron annihilation must travel at 180° to each other.

20 Fluorine-18 decays by β^+ emission. Write a nuclear equation to show this decay.

21 **a** Calculate the energy released when a positron and an electron annihilate.

(Mass of an electron = mass of a positron = 9.1×10^{-31} kg.)

b Calculate the frequency of the γ-rays emitted.

c Calculate the momentum of the one of the γ-rays emitted.

Detecting the γ-rays

The patient being scanned is placed on a bed with a series of rings of detectors, in a donut type shape. The patient on the bed is moved through the detectors, so that a series of images of 'slices' through the patient are made in similar manner to those made by a CT scan. Indeed, PET scans are often combined with CT scans so that more information is gathered.

Figure 30.23: A patient being prepared for a PET/CT scan by a radiologist. Note the donut shaped ring, which contains the detectors and through which the patient will be moved.

The detectors of the γ-ray photons consist of two parts: a crystal that scintillates and a photomultiplier. When a high energy γ-ray photon is incident on the crystal, an electron is excited into a very high energy state. As the electron travels through the crystal, it loses energy and excites more electrons; these electrons then decay back to their original state, emitting visible light photons. The photons produced by the scintillator are then converted into an electrical signal by the photomultiplier tube – these signals are then fed to a computer that can plot back where the photon pair was originally produced.

Reconstruction of the image

Figure 30.24 shows a simplified view of the detectors in a PET scanners. They form a series of rings around the patient. The γ-ray photons, formed by an electron–positron annihilation, travel from a point very near to the event. They travel in a straight line and in opposite directions and strike the detectors as shown. A line (known as the **line of response**) can be drawn, joining the two detectors. Using the time lapse between the two photons arriving at the detectors, the position on the line of response can be established. In practice, there are many annihilations and sophisticated computers analyse the data and convert it into an image. The numbers of photons arriving from a particular point determine the concentration of the tracer at that point. Where there are many arriving per unit time, it means that there is a high concentration of tracer and this will appear as a bright point on the image.

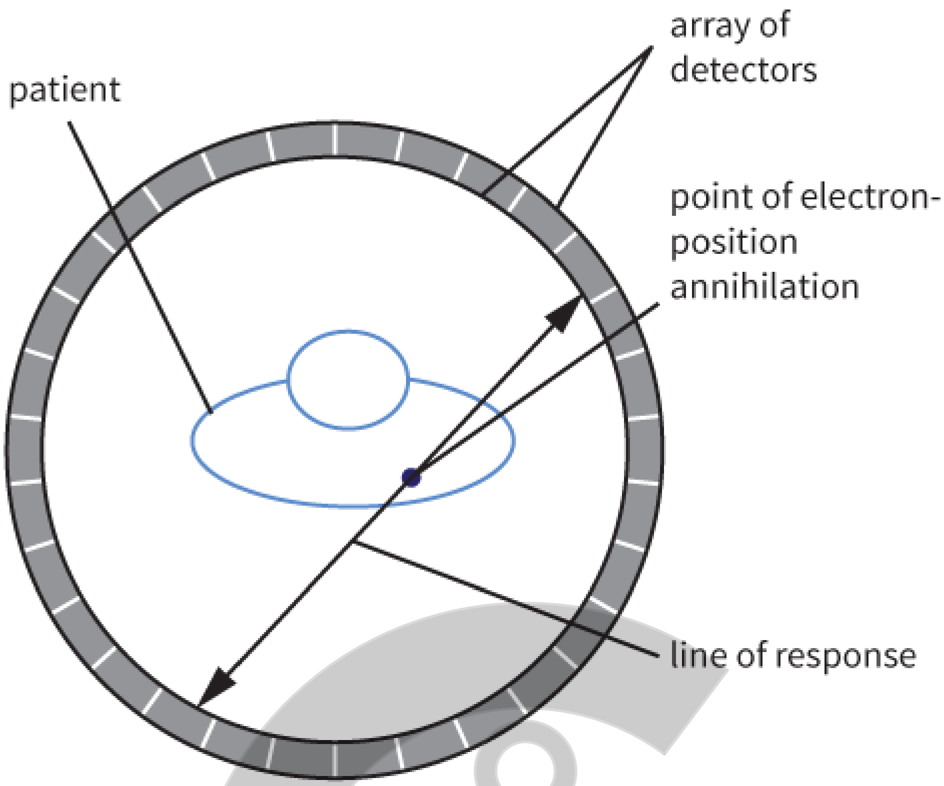


Figure 30.24: The arrangement of detectors in a PET scanner.

REFLECTION

It is about 120 years since X-rays were discovered. Modern medicine has many methods for looking inside the bodies of people who are unwell or have suffered injuries. Use the internet to find as many different methods as you can. Try and draw a timeline to show when these methods were developed.

What did you learn about yourself as you worked on this activity? Did you find it a useful way of learning?

SUMMARY

X-rays are short wavelength, high frequency, electromagnetic radiation.

X-rays are formed when electrons are decelerated.

Intensity is the power transmitted per unit cross-sectional area.

Intensity of a beam of X-rays shows an exponential decrease as the beam passes through matter.

Image intensifiers and contrast media are used to improve the quality of X-ray images.

Ultrasound is a longitudinal wave with a frequency greater than 20 kHz.

Transducers that use the piezo-electric effect are used to generate and detect ultrasound.

Acoustic impedance of a material (Z) depends on the density of the material and the speed of the sound in the material and is given by the formula:

$$Z = \rho c$$

The fraction of the intensity of an ultrasound reflected at a boundary between different materials is given by the formula:

$$\frac{I_r}{I_0} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2}$$

An impedance matching gel is used to avoid a large degree of reflection occurring as the ultrasound travels from the air to the skin of the patient.

Radioactive tracers are used in diagnosing tumours.

In a PET scan, the tracer emits positrons (β^+-particles).

When a particle and an antiparticle meet, they annihilate.

In an annihilation event, both momentum and mass–energy are conserved.

The energy of an X-ray photon is given by the formula $E = hf$ and its momentum $p = \frac{E}{c}$

EXAM-STYLE QUESTIONS

1 X-rays are produced by firing electrons at a metal anode target.

Which statement is correct? **[1]**

A The frequencies of the characteristic spectrum lines are determined by the potential used to accelerate the electrons.

B The frequencies of the characteristic spectrum lines are determined by the metal used to make the target.

C The maximum frequency of the braking radiation is determined by the metal used to make the target.

D The minimum frequency of the braking radiation is determined by the potential used to accelerate the electrons.

2 Which statement about PET scanning is correct? **[1]**

A A positron is emitted by the radiotracer that interacts with an electron in the detector producing two γ-rays that move apart at 180° to each other.

B A positron is emitted by the radiotracer that interacts with an electron in the detector producing two γ-rays that move apart at right angles to each other.

C A positron is emitted by the radiotracer that interacts with an electron in the surrounding tissue producing two γ-rays that move apart at 180° to each other.

D A positron is emitted by the radiotracer that interacts with an electron in the surrounding tissue producing two γ-rays that move apart at right angles to each other.

3 **a** Explain what is meant by ionising radiation and explain why it can be harmful to humans. **[2]**

b Which of the following scans use ionising radiation? **[2]**

- X-ray shadow imaging
- ultrasound A-scan
- ultrasound B-scan
- PET scan
- CT scan

[Total: 4]

4 Calculate the minimum wavelength (in air) of X-rays produced when the accelerating potential across the source is 20 kV. **[2]**

5 Explain why a gel is used between the skin and the transducer when an ultrasound scan of a foetus is taken. **[2]**

6 For ultrasound of frequency 3.5 MHz, the acoustic impedance of muscle is 1.78×10^6 kg m^{-2} s^{-1}, and that of soft tissue is 1.63×10^6 kg m^{-2} s^{-1}.

Calculate the percentage of the incident ultrasound reflected at a muscle–soft tissue boundary. **[3]**

7 A transducer produces ultrasonic waves of frequency 800 kHz. The speed of sound in the crystal is 5200 m s^{-1}.

Calculate the optimum thickness for the crystal. **[2]**

8 State and explain two reasons why full-body CT scans are not offered for regular checking of healthy patients. **[2]**

9 **a** Explain, with the aid of a simple, labelled diagram, how X-rays are produced. **[5]**

[3]

b Discuss the energy changes in the production of X-rays.

[Total: 8]

10 This graph shows the spectrum of X-rays produced from an X-ray source.

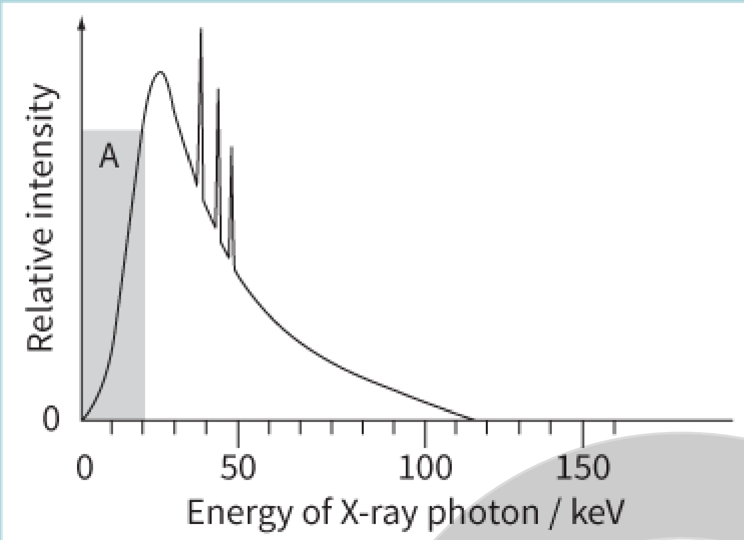


Figure 30.25

a Describe the process by which:

i the three sharp peaks of high-intensity X-rays are produced **[2]**

ii the broad band of X-rays is produced. **[2]**

b The X-rays in the shaded region, labelled A, are filtered out using an aluminium filter. Explain:

i why it is advantageous to filter these X-rays out **[2]**

ii why aluminium is a suitable material to filter them out. **[2]**

c Calculate the maximum frequency of X-rays produced by this tube. **[3]**

[Total: 11]

11 a An X-ray beam, containing X-rays with a variety of frequencies and that has an intensity of 4.0 × 105 W, is incident on an aluminium plate of thickness 5.0 cm. The average linear attenuation coefficient is 250 m^{-1}.

i Calculate the intensity of the transmitted beam. **[3]**

ii Explain the advantages of passing the X-rays through this aluminium plate prior to their being incident on a patient. **[3]**

[Total: 6]

12 a Explain what is meant by acoustic impedance and outline its role in the use of ultrasound scans. **[3]**

b Brain tissue has a density of 1.04×10^3 kg m^{-3} and ultrasound travels at 1.58×10^3 m s^{-1} through it.

Calculate the acoustic impedance of brain tissue. **[2]**

c This is the trace formed on the screen of an oscilloscope when ultrasound is reflected from the front and rear surfaces of the head of a fetus. The time-base of the oscilloscope is set at 10 μs div^{-1}.

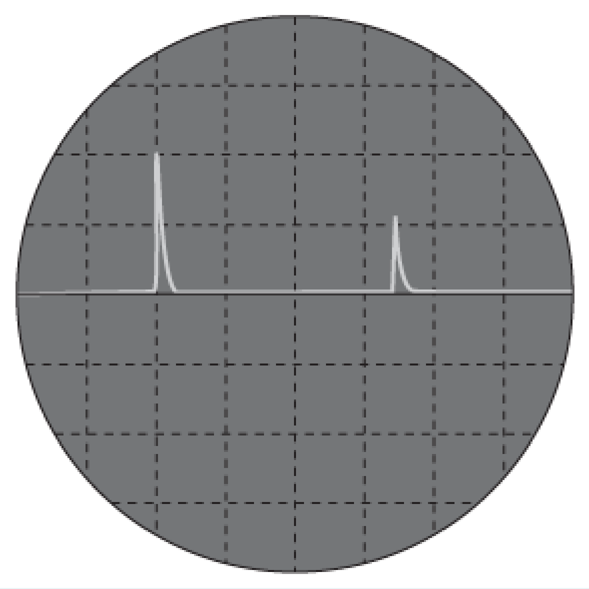


Figure 30.26

i Explain why the second peak is lower than the fetus. **[1]**

ii Calculate the diameter of the head of the fetus. **[3]**

[Total: 9]

13 a Outline the theory of the PET scanner. **[5]**

b It is suggested that a scanner could be designed using the annihilation of a proton and an antiproton.

Calculate:

i the energy released in the proton–antiproton annihilation **[2]**

ii the wavelength of the γ-ray photons produced in the annihilation. **[1]**

c In one type of PET scanner, the tracer isotope is Rb-82. Write an equation for the decay of this isotope. **[2]**

(Mass of proton = 1.67×10^{-27} kg.)

[Total: 5]

14 a i With reference to PET scanning, explain the meaning of the term tracer. **[3]**

ii Explain what is meant by the term line of response and how it is used to identify the precise site of cancerous tissue. **[2]**

b i With reference to PET scanning, explain what is meant by an annihilation event. **[3]**

ii Name the important quantities which are conserved in an annihilation event. **[3]**

[Total: 11]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
understand that X-rays are short-wavelength, high-frequency electromagnetic radiation, produced when electrons are decelerated	30.1			
recall that intensity of an X-ray beam is the power transmitted per unit cross-sectional area	30.1, 30.2			
understand intensity of a collimated X-ray beam decreases exponentially according to the equation $I = I_0 e^{-\mu x}$, where μ is the attenuation coefficient of the medium. μ has units of m^{-1} (or cm^{-1} or mm^{-1})	30.2			
recognise that X-ray images can be improved using image intensifiers and contrast media (such as barium or iodine)	30.3			
understand that ultrasound is a longitudinal wave with a frequency greater than 20 kHz	30.5			
recall that ultrasound transducers use the piezo-electric effect to generate and detect ultrasound waves	30.5			
understand that the acoustic impedance Z of a material depends on its density ρ and the speed c of sound: $Z = \rho c$	30.6			
recall and use the formula for the fraction of the intensity of an ultrasound wave reflected at a boundary: $\frac{I_r}{I_0} = \frac{(Z_2-Z_1)^2}{(Z_2+Z_1)^2}$ or $\frac{I_r}{I_0} = \left(\frac{Z_2-Z_1}{Z_2+Z_1}\right)^2$	30.6			
recognise that to transfer a high proportion of the intensity of an ultrasound pulse into the patient's body, an impedance-matching gel must be used with acoustic impedance almost the same as that of the skin	30.6			
understand that a tracer (radiotracer) is a substance that can be injected into the body and is then absorbed by tissue and organs	30.8			

I can	See topic…	Needs more work	Almost there	Ready to move on
recall that the tracer decays by β^+ emission in a PET scan	30.8			
understand that when a particle meets its antiparticle that annihilation occurs	30.8			
understand that momentum and mass–energy are both conserved in an annihilation event	30.8			
calculate the energy and frequency of the γ-rays emitted in a positron–electron annihilation event	30.8			
understand that the γ-ray photons from an annihilation event travel outside the body and can be detected	30.8			
understand that the positions of the annihilation events can be determined from the detector positions and the arrival times of the photons	30.8			
understand that the tracer concentration at particular tissues or organs can be calculated from the numbers of γ-ray photons arriving per unit time.	30.8			

〉Chapter 31

Astronomy and cosmology

LEARNING INTENTIONS

In this chapter you will learn how to:

- understand the term luminosity as the total power of radiation emitted by a star
- recall and use the inverse square law for radiant flux intensity F in terms of the luminosity L of the source: $F = \frac{L}{4\pi d^2}$
- understand that an object of known luminosity is called a standard candle
- understand the use of standard candles to determine distances to galaxies
- recall and use Wien's displacement law $\lambda_{max} \propto \frac{1}{T}$ to estimate the peak surface temperature of a star
- use the Stefan-Boltzmann law $L = 4\pi\sigma r^2 T^4$
- use Wien's displacement law and the Stefan-Boltzmann law to estimate the radius of a star
- understand that the lines in the emission spectra from distant objects show an increase in wavelength from their known values
- use $\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$ for the redshift of electromagnetic radiation from a source moving relative to an observer
- explain why redshift leads to the idea that the Universe is expanding
- recall and use Hubble's Law $v \approx H_0 d$ and explain how this leads to the Big Bang theory.

BEFORE YOU START

- Your knowledge of electromagnetic waves, including spectra, would be valuable in the understanding of this chapter.
- Can you recall intensity of a wave and its units?
- The idea of the Doppler effect of sound will be extended to spectra from distant stars. When is the observed wavelength shorter, or longer?

LOOKING INTO THE PAST

Figure 31.1 shows galaxies as seen through a powerful telescope. Each galaxy may have as many as 10^{11} stars, and there may be as many as 10^{11} galaxies in the Universe. Light from these galaxies has a finite speed: 3.0×10^8 m s^{-1} in a vacuum. These galaxies are so distant that light from them may have taken billions of years to reach us. So, what we have in this photograph is an image of the past. Andromeda is our closest galaxy. The light from this galaxy would take 2.3 million years to reach us. When we see this galaxy though a telescope, we are looking at its image from 2.3 million years ago! Just to put this into perspective, someone looking at the Earth from this galaxy now, would see a time when our ape-like ancestors roamed the planet.

In this chapter, we will deduce that galaxies further away from us are moving faster. This, in turns, implies that our Universe had a beginning–it was created some 14 billion years ago in an event known as the **Big Bang**. Since then, the fabric of the Universe has been stretching, carrying with it the galaxies.

Can you estimate the size and the mass of the Universe?

Figure 31.1: A cluster of distant galaxies; some created only a few million years after the creation of the Universe.

31.1 Standard candles

All the stars we see in the night sky are from our own galaxy–the Milky Way. Figure 31.2 shows stars in the constellation of Gemini. The stars do not look the same; they differ in brightness and colour. These stars are not all the same distance from us, and they do not all emit the same power. So, we cannot deduce their distance from just how bright they appear in the night sky.

Figure 31.2: Stars have different colours and brightness. Can you tell which star is the closest?

In astronomy, **luminosity** of a star is defined as the **total** radiant energy emitted per unit time. This is the same as the total power emitted by a star. In SI units, luminosity L is measured in W or J s^{-1}. The Sun is the nearest star to us, and astronomers have determined its luminosity to a high degree of accuracy. The luminosity of the Sun (solar luminosity), often written as $L_{\odot}$, is about 3.83×10^{26} W.

As you will see later, you can determine solar luminosity from the intensity of solar radiation reaching the Earth.

In astronomy, a **standard candle** is an astronomical object of known luminosity. Astronomers can determine the distance of a standard candle by measuring the intensity of the electromagnetic radiation arriving at the Earth.

Standard candles have been successfully used to determine the distance of far-flung galaxies. It is amazing that we can do this just by observing the starlight reaching us on Earth.

The two well-known standard candles are Cepheid variable stars and Type 1A supernovae.

Cepheid variable stars

In 1908, Henrietta Leavitt discovered that the brightness of Cepheid variable stars varied periodically, and the period of this variation was related to the average luminosity of the star. By measuring the period, astronomers could determine the luminosity of the star. The star's distance could then be calculated from the observed radiant intensity at the Earth. Finding a Cepheid variable star in a distant galaxy meant that the distance of the galaxy itself could be calculated.

Type 1A supernovae

Type 1A supernovae stars implode rapidly towards the end of their lives, and scatter matter and energy out into space. This implosion event can be brighter than the galaxy itself. The luminosity of the star at the time of the implosion is always the same. From this, astronomers can estimate the star’s distance from the Earth.

31.2 Luminosity and radiant flux intensity

The Sun is the nearest star to the Earth. The second nearest star is Proxima Centauri, 4.0×10^{16} m away. Distances as large as 4.0×10^{16} m are extremely difficult to visualise–they are way beyond any of our day-to-day points of reference. So, astronomers tend to use an alternative unit for distance – the **light-year (ly)**. A light-year is the distance travelled by light in a vacuum in a time of one year. Therefore:

1 ly = speed of light in vacuum × one year in seconds

$$1 \text{ ly} \approx 3.00 \times 10^{8} \times 365 \times 24 \times 3600$$
$$1 \text{ ly} \approx 9.5 \times 10^{15} \text{m}$$

Proxima Centauri is 4.2 ly away. It would take light from Proxima Centauri 4.2 years to reach us.

(Note: You do not need to know about light-years, but they help to visualise vast distances.)

Table 31.1 summarises some data on the brightest stars–do not forget that the Sun is a star too.

Rank order	Name of star	Distance / light-years	Temperature / K	Luminosity / $L_\odot$
1	Sun	1.58×10^{-5}	5800	1.0
2	Sirius	8.6	9900	25
3	Canopus	310	7000	1100
4	Alpha Centauri	4.4	5800	1.5
5	Arcturus	37	4300	170
6	Vega	25	9600	40

Table 31.1: Data on the six brightest stars, including the Sun. The luminosity is given in terms of the solar luminosity $L_\odot$; 1 $L_\odot = 3.83 \times 10^{26}$ W.

We can see from Table 31.1 that the observed brightness of a star is linked to both its distance from the Earth and its luminosity. We would expect a luminous star, such as Canopus, to be bright in the night sky. Alpha Centauri is brighter in the night sky than Arcturus, not because of its luminosity, but because of its closeness to us. You will see later that the luminosity of a star depends not only on its surface temperature but also on its physical size.

Can we relate the brightness of a star to its luminosity? Yes, as long as we understand the underlying assumptions that:

- the power from the star is uniformly radiated through space
- there is negligible absorption of this radiated power between the star and the Earth.

With these assumptions, we can determine the intensity of electromagnetic radiation observed at the Earth.

The observed intensity is known as **radiant flux intensity** *F*. This is defined as the radiant power passing **normally** through a surface per unit area.

Figure 31.3 shows how *F* can be calculated for a star at a distance *d* from its **centre**.

$$\text{radiant flux intensity} = \frac{\text{power of star}}{\text{surface area of sphere}}$$

The power of the star is its luminosity L, and the surface area of a sphere is $4\pi d^2$.

Therefore:

$$F = \frac{L}{4\pi d^2}$$

The SI units for radiant flux intensity are W m^{-2}.

For a given star, the luminosity L is constant, so according to the equation, the radiant flux intensity F obeys an inverse square law with distance d. So, doubling the distance from the **centre** of the star ($2d$) will decrease F by a factor of 4, and tripling the distance ($3d$) will decrease F by a factor of 9, and so on.

You can demonstrate this inverse square law using a bright filament lamp and a light-meter in a darkened laboratory – see Practical Activity 31.1.

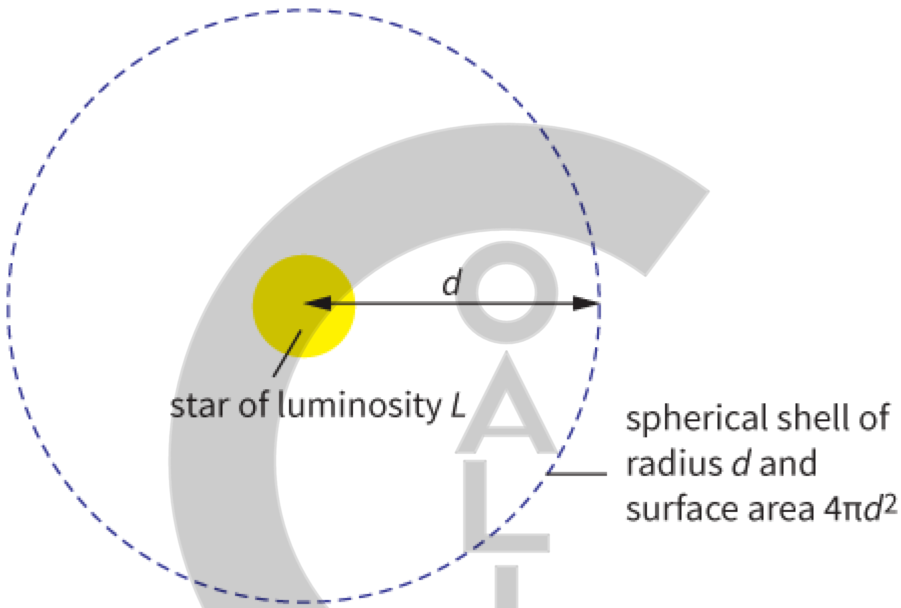


Figure 31.3: The power of the star spreads out uniformly through a spherical shell.

Distance of galaxies

Astronomers on the Earth can determine the radiant flux intensity F of a distant star. The equation $F = \frac{L}{4\pi d^2}$ can then be rearranged to determine the distance d of a star of known luminosity L, for example, a standard candle such as a Cepheid variable star in a distant galaxy.

Now look at Worked examples 1 and 2. In Worked example 1, the radiant flux intensity from the Sun at the Earth is calculated. In Worked example 2, the distance of a star in the Andromeda galaxy is calculated from its radiant flux intensity at the Earth.

WORKED EXAMPLES

1 The radius of the Sun is 6.96×10^8 m and its luminosity is 3.83×10^{26} W.

The orbital radius of the Earth is 1.50×10^{11} m.

Calculate the radiant flux intensity at the surface of the Sun and at the position of the Earth.

Step 1 Calculate the radiant flux intensity at the Sun's surface:

$$F = \frac{L}{4\pi d^2}$$

$$= \frac{3.83\times10^{26}}{4\pi\times(6.96\times10^{8})^2}$$

$$= 6.29\times10^{7}\ \text{W m}^{-2}$$

Step 2 Calculate the radiant flux intensity at the Earth's position.

We can do this using the inverse square law relationship between *F* and *d*.

The distance *increases* by a factor of:

$$\frac{1.50\times10^{11}}{6.96\times10^{8}} = 215.52$$

Therefore, *F* will decrease by a factor of 215.52^2

So,

$$F = \frac{6.29\times10^{7}}{215.52^2}$$

$$= 1.35\times10^{3}\ \text{W m}^{-2}$$

Note: An alternative would be to just use $F = \frac{1}{4\pi d^2}$ with $L = 3.83 \times 10^{26}$ W and $d = 1.50 \times 10^{11}$ m. Try it, you will get the same answer. You do not need the radius of the Sun to get the right answer.

2 The radiant flux intensity, measured at the Earth, from a Cepheid variable star in Andromeda is 1.4×10^{-16} W m^{-2}. The luminosity of the star is 1.0×10^{30} W.

Calculate the distance of this star.

Step 1 Rearrange the equation for radiant flux intensity.

$$F = \frac{L}{4\pi d^2} \Rightarrow d = \sqrt{\frac{L}{4\pi F}}$$

Step 2 Substitute and calculate the distance of the star.

$$d = \sqrt{\frac{1.0\times10^{30}}{4\pi\times1.4\times10^{-16}}}$$

$$= 2.4\times10^{22}\ \text{m}$$

This distance is equivalent to 2.5 million light-years.

PRACTICAL ACTIVITY 31.1

Inverse square law for radiant flux intensity

We can simulate the inverse square law nature of light spreading from a star using a bright filament lamp and a light-meter. Commercial light-meters are not calibrated to show radiant flux intensity *F* in W m^{-2}. Light-meters measure a quantity known as illuminance. We can assume that illuminance, often in a unit known as lux, is directly proportional to radiant flux intensity.

Carry out the experiment in a darkened room.

Measure the illuminance at various distances *d* from the centre of the lamp.

Since radiant flux intensity is inversely proportional to d^2, and directly proportional to illuminance, a graph of illuminance against $\frac{1}{d^2}$ will be a straight line through the origin.

In a laboratory, there will always be some reflection of light from the walls and ceiling. The best place and time for the experiment is outdoors at night!

Questions

Where necessary, take:

$L_{\odot} = 3.83 \times 10^{26}$ W

1 ly ≈ 9.5×10^{15} m

1 State **two** factors that affect radiant flux intensity from a star.

2 The radiant flux intensity F of light from a lamp at a distance of 10 cm is 0.32 W m^{-2}. Calculate F from the same lamp at a distance of 15 cm. State any assumption(s) you make.

3 Use data from Table 31.1 to determine, to two significant figures:

- **a** the distance of Sirius from the Earth in metres.
- **b** the luminosity (in W) of
  - **i** Canopus
  - **ii** Vega.
- **c** the radiant flux intensity measured at the Earth from:
  - **i** Sirius
  - **ii** Alpha Centauri.

4 This question is about Sirius and Arcturus.

With the help of calculations and data from Table 31.1, show that Sirius is brighter than Arcturus.

5 The radiant flux intensity from a star measured at the Earth is 2.7×10^{-9} W m^{-2}. The luminosity of the star is 1300 $L_{\odot}$.

Calculate the distance of this star from the Earth in metres.

31.3 Stellar radii

The Sun can be seen as a glowing ball of gas in the sky. If you briefly look at the Sun through a special filter, like a welder's helmet, you can identify it as a yellow disc in space. The Sun is enormous – it only looks small because it is far away from us. We can determine the diameter of the Sun fairly easily (see Practical Activity 31.2). However, when we look at stars in the night sky, they appear as tiny specks of light – there is no disc to be seen (Figure 31.4). The stars are just too far away. Even the closest stars viewed through powerful telescopes appear as specks of light.

How can astronomers determine the size of stars? In this topic, you will see how two simple laws can be used to determine stellar radii.

Figure 31.4: Even the closest stars appear as specks of light – so how do astronomers determine their size?

PRACTICAL ACTIVITY 31.2

The diameter of the Sun

You can estimate the diameter of our closest star, the Sun, using a simple pin-hole camera. You can make this camera using a shoe-box. One end of the box has a sheet of darkened paper (or aluminum foil) with a tiny hole made with a sharp pin. The opposite end of the box has a sheet of tracing paper, which acts as a screen. A circular image of the Sun is formed on the screen when the camera is pointed towards the Sun. See Figure 31.5.

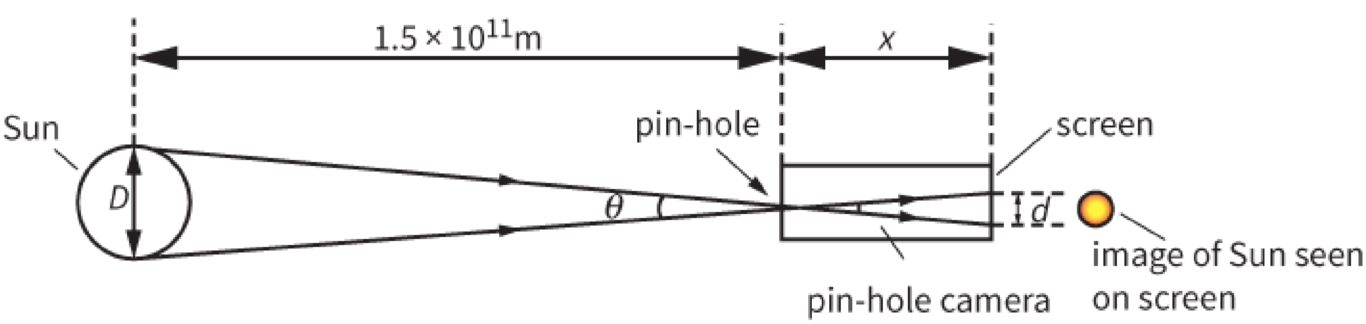


Figure 31.5: You can determine the diameter *D* of the Sun using a simple pin-hole camera.

Measure the distance *x* between the pin-hole and the screen. The distance of the Sun from the Earth is 1.5×10^{11} m. The diameter *D* of the Sun can be determined using simple trigonometry:

$$\tan\theta \approx \frac{d}{x} \approx \frac{D}{1.5\times10^{11}}$$

Therefore:

$$D = \frac{1.5\times10^{11}\times d}{x}$$

Question

6 **a** A student conducted the experiment from Practical Activity 31.2. The results from the experiment are shown below:

x = 300 mm d = 3 mm

Use this data to estimate the diameter of the Sun.

b The actual value for the diameter of the Sun is 1.4×10^{9} m.

Determine the percentage difference between your calculated value and the actual value.

Wien's displacement law

The hottest stars are blueish-white in colour. Cooler stars are a deep shade of red. We can see almost the same effect with the filament of a lamp. Increase the temperature of the filament by increasing the current in the filament. At first, the filament will glow dull red when it is cooler, then reddish-orange, and eventually white as it gets hotter.

There is a link between the observed wavelength of light and temperature. Table 31.2 shows the colour of a star in the night sky and the range of its surface temperature.

A hot object, such as a star, can be modelled as a **black body**. A black body is an idealised object that absorbs all incident electromagnetic radiation falling on it. It has a characteristic emission spectrum and intensity that depend only on its thermodynamic temperature. Figure 31.6 shows typical intensity against wavelength graphs for objects at different temperatures.

Colour of star	Surface temperature of star / K
blue	Greater than 33 000
blue to blue-white	10 000 – 30 000
white	7500 – 10 000
yellowish white	6000 – 7500
yellow	5200 – 6000
orange	3700 – 5200
red	Less than 3700

Table 31.2: The observed colour of a star is related to its temperature.

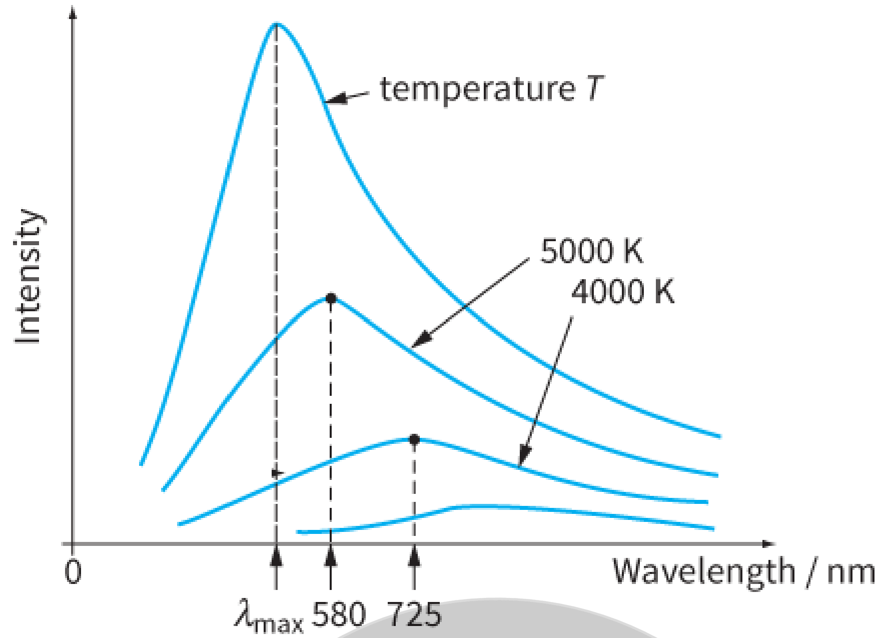


Figure 31.6: The intensity–wavelength graph depends on the temperature of the object. For an object at a thermodynamic temperature *T*, the intensity against wavelength curve peaks at a wavelength λ_{max}.

The higher the temperature of a body:

- the shorter the wavelength at the peak (maximum) intensity
- the greater the intensity of the electromagnetic radiation at each wavelength.

In 1893, German physicist Wilhelm Wien discovered a relationship between the thermodynamic temperature *T* of the object and the wavelength λ_{max} at the peak intensity:

$$\lambda_{max}T = \text{constant}$$

The relationship is known as **Wien's displacement law** . The experimental value of the constant is 2.9×10^{-3} m K.

The surface temperature of the Sun is 5800 K. This gives a λ_{max} value of about 5.0×10^{-7} m or 500 nm. Light of this wavelength appears yellow (which is not surprising for the Sun).

Questions

7 Use the data given in Figure 31.6 to show the validity of Wien's displacement law for 5000 K and 4000 K.

8 For a temperature of 5800 K, the wavelength at peak intensity of electromagnetic radiation is 500 nm. Calculate the surface temperature of a star with wavelength 350 nm at peak intensity.

9 Copy this table.

Star	Surface temperature *T* / K	λ_{max} / nm
Sun	5800	500
Polaris	6000	
Canopus	7000	
Gacrus		810

Use Wien's displacement law to complete the table. Write your answers to two significant figures.

The Stefan-Boltzmann law

A quick inspection of Table 31.1 shows that the luminosity of a star does not depend just on the surface temperature of the star. Luminosity also depends on the physical size of the star–its radius. For example, the super red giant star KY Cygni has a surface temperature of 3500 K but its luminosity is 200 000 times that of our Sun. KY Cygni is cooler than the Sun, but its large surface area makes it very luminous.

The luminosity of a star depends on two factors:

- its surface thermodynamic temperature T
- its radius r.

In 1879, Slovenian physicist Josef Stefan developed an expression for the luminosity L of a star. This is the **Stefan-Boltzmann law**:

$$L = 4\pi\sigma r^2 T^4$$

KEY EQUATIONS

Wien's displacement law:

$$\lambda_{max} T = \text{constant}$$

$$\lambda_{max} \propto \frac{1}{T}$$

Stefan-Boltzmann law:

$$L = 4\pi\sigma r^2 T^4$$

where σ is a constant known as the Stefan-Boltzmann constant. The experimental value for σ is 5.67×10^{-8} W m^{-2} K^{-4}.

Using Wien's displacement law and the Stefan-Boltzmann law to determine stellar radii

The radius of a star can be calculated from Wien's displacement law and the Stefan-Boltzmann law. The procedure would be as follows:

1. Use Wien's displacement law to determine the temperature T of the star. This would involve determining the wavelength γ_{max} at maximum intensity for the star, and then using a reference star (such as the Sun) to determine T.
2. Use the Stefan-Boltzmann law to determine the radius r of the star. The luminosity L of the star can be determined by measuring the radiant flux intensity F of the star.

The procedure is illustrated in Worked example 3.

WORKED EXAMPLE

3 The surface temperature of the Sun is 5800 K and wavelength of light at peak intensity is 500 nm. The wavelength at peak intensity for Sirius-B (a white dwarf star) is 120 nm. The luminosity of this star is 0.056 times that of the Sun. The luminosity of the Sun is 3.83×10^{26} W.

Calculate the radius of Sirius-B.

Step 1 Use Wien's displacement law to calculate the temperature of Sirius-B.

$\lambda_{max}T$ = constant

$5800 \times 500 = T \times 120$

$T = 24\ 167\ \text{K} \approx 24\ 200\ \text{K}$

Step 2 Use the Stefan-Boltzmann law to calculate the radius of Sirius-B.

$$L = 4\pi\sigma r^2 T^4$$

$$r = \sqrt{\frac{L}{4\pi\sigma T^4}}$$

$$= \sqrt{\frac{0.056 \times 3.83 \times 10^{26}}{4\pi \times 5.67 \times 10^{-8} \times 24167^4}}$$

$$\approx 9.4 \times 10^6\ \text{m}$$

Sirius-B is roughly the size of our Earth! It is a very hot star, but not very luminous because of its small size.

Question

10 The luminosity of the star Aldebaran is 520 times that of the Sun. The wavelength of light at peak intensity for Aldebaran is 740 nm and the wavelength of light at peak intensity for the Sun is 500 nm.

a Explain whether Aldebaran is cooler or hotter than the Sun.

b Calculate the ratio:

radius of Aldebaran / radius of the Sun.

31.4 The expanding Universe

The **Big Bang theory** is a model of the evolution of the Universe from an extremely hot and dense state some 13.8 million years ago – the event was called the Big Bang.

The Big Bang was also responsible for the birth of the fabric of space (and time) – this fabric has been expanding ever since then. At the early stages after the Big Bang, fundamental particles (such as quarks) and forces (such as gravitation) came into existence. Subsequent expansion led to cooling and formation of atoms, stars and galaxies. The one question that cosmologists cannot answer (yet) is *why* the Big Bang happened in the first place. There are lots of thoughts and theories, but nothing that can be tested.

In this topic, we will explore evidence for the Big Bang by using the ideas of physics developed in the earlier chapters of this book–notably spectra (Chapter 12) and Doppler effect (Chapter 12).

Hubble's law

Astronomers can see the light from distant galaxies using powerful telescopes. The telescopes can look at the light through a diffraction grating. Analysis of the spectrum of the light from distant galaxies shows that they are all moving away from us. The more distant a galaxy, the faster it moves. How do we know from the spectrum that galaxies are moving away from us (receding)?

We examined the Doppler effect of sound in Chapter 12. The observed wavelength of sound was longer for a receding source and shorter for an approaching source. The same happens with electromagnetic waves. The observed wavelengths of **all** spectral lines from distant galaxies are longer than the ones observed in the laboratory. This is known as **redshift**. Figure 31.7 shows the red-shifting of the absorption spectral lines from a cluster of galaxies some 1 billion light-years away.

The redshift of spectral lines from distant galaxies must imply that all galaxies are receding from us. This is what American astronomer Vesto Slipher discovered in 1917. Another American astronomer, Edwin Hubble, combined his own observations with Slipher's discovery to create **Hubble's law**.

Hubble's law states that the recession speed v of a **galaxy** is directly proportional to its distance d from us.

Therefore,

$$v \propto d$$

or

$$v = H_0 d$$

Figure 31.7: The absorption lines in the spectrum of the galaxies are all shifted to longer wavelengths – redshifted. The top spectrum is the spectrum from a 'stationary source', the Sun.

where v is the recession speed, d is the distance of the galaxy and H_0 is the Hubble constant.

The SI unit for H_0 is second^{-1}, or s^{-1}.

The experimental value for H_0 is about 2.4×10^{-18} s^{-1}. Figure 31.8 shows a recession speed v against distance d graph for galaxies. The straight-line of best fit passes through the origin, and the gradient of the line is equal to H_0.

KEY EQUATION

Hubble's law:

$$v = H_0 d$$

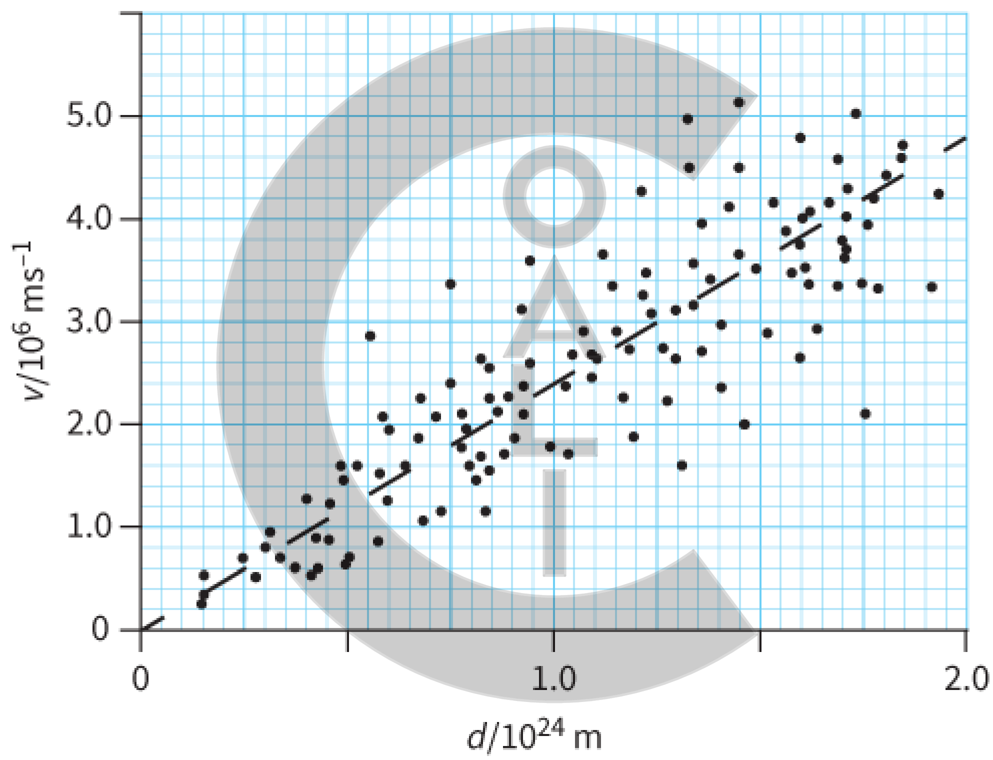


Figure 31.8: Hubble's law shows that recession speed of galaxy ∝ distance from us. The gradient of the best-fit line is equal to H_0 in s^{-1}. The scatter of the data shows considerable uncertainties in the observation.

Question

11 A galaxy is at a distance of 9.5×10^{24} m from us and is moving away with a speed of 2.1×10^7 m s^{-1}.

a Calculate the Hubble constant based on this data.

b Estimate the speed in km s^{-1} of a galaxy at a distance of 1.9×10^{25} m.

Doppler redshift

It is worth pointing out that the term redshift does not imply spectral lines becoming red; all spectral lines show an increase in wavelength. The fractional increase in the wavelength depends on the recession speed v of the source (galaxy).

For non-relativistic galaxies – those moving with speeds far less than the speed of light in a vacuum *c* – we can use the relationship:

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$$

where λ is the wavelength of the electromagnetic waves from the source, $\Delta\lambda$ is the change in the wavelength, *f* is the frequency of the electromagnetic waves from the source, Δf is the change in frequency, *v* is the recession speed of the source and *c* is the speed of light in vacuum.

KEY EQUATION

Doppler redshift:

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$$

Astronomers and cosmologists often assign a value for the term ‘redshift’. For example, a galaxy shows redshift of 7.0 % means that:

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c} \approx 0.070$$

Worked example 4 shows how redshift can be used to determine the speed of a distant galaxy.

WORKED EXAMPLE

4 In the laboratory, an emission spectral line is observed at a wavelength of 656.4 nm. The same spectral line, in the spectrum from a distant galaxy, has wavelength 663.1 nm.

Calculate the speed *v* of the galaxy.

Step 1 Calculate the change in the wavelength of the spectral line.

The observed wavelength is longer; therefore, the galaxy is receding.

$\Delta\lambda = 663.1 - 656.4 = 6.7$ nm

Step 2 Now calculate the speed *v* using the Doppler redshift equation.

Hint: *You do not need to convert the nm to m, because the ratio* $\frac{\Delta\lambda}{\lambda}$ *will be the same; just make sure you use the* ***same*** *unit for* $\Delta\lambda$ *and* λ.

$$\frac{\Delta\lambda}{\lambda} \approx \frac{v}{c}$$

$$\frac{6.7}{656.4} \approx \frac{v}{3.0\times10^8}$$

$$0.0102 \approx \frac{v}{3.0\times10^8}$$

$$v \approx 0.0102 \times 3.0 \times 10^8$$

$$v \approx 3.06 \times 10^6 \text{ m s}^{-1}$$

$$v \approx 3.1 \times 10^6 \text{ m s}^{-1}$$

Questions

12 The fractional change in the wavelength of the observed light from a galaxy is 0.15; its redshift is 15 %.

Calculate its recession speed. State any assumptions made.

13 The Tadpole galaxy has a recession speed of 9400 km s^{-1}.
Calculate the fractional change in the wavelength of the observed spectrum.

Evidence for the Big Bang

All galaxies in the Universe are moving away (receding) from each other, and not from the Earth. An observer in another galaxy will reach the same conclusion. The galaxies have motion because space itself is stretching. This is quite difficult to visualise. The best we can do is to imagine the galaxies as dots on the surface of an ever-expanding balloon (see Figure 31.9).

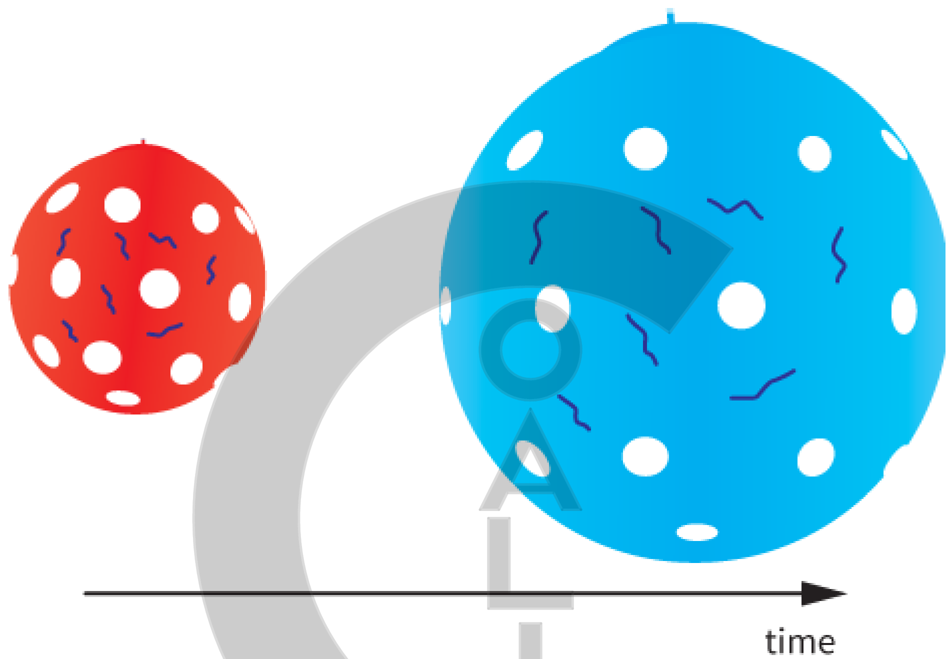


Figure 31.9: The galaxies are modelled as dots on the surface of a balloon. Expansion of the balloon makes all the dots move **away** from each other.

The expanding balloon model can also be used to explain the redshift of light from galaxies. As the Universe expanded, the wavelength of photons was stretched out.

Hubble's law provided the first evidence for the birth, and the subsequent expansion, of the Universe. Distant galaxies appear to be moving faster. However, we must remember that the light has a finite speed, so as we stare deeper into space, we are looking further into the past. The further back in time we go, the faster the galaxies are receding from each other. Rolling back time in this way – like playing a movie in reverse – can only lead to the conclusion that the Universe must have had a beginning … the Big Bang.

How long ago was the Big Bang? We can estimate this from the Hubble constant H_0. In Question 11, the speed of the receding galaxy at a distance of 9.5×10^{25} m was 2.1×10^{7} m s^{-1}. If we assume that this speed has remained unchanged, we can estimate the time when our galaxy and this receding galaxy were at the same place (the time of the Big Bang):

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

$$2.1 \times 10^7 = \frac{9.5 \times 10^{25}}{\text{time}}$$

$$\text{time} = \frac{9.5 \times 10^{25}}{2.1 \times 10^7}$$

$$= 4.52 \times 10^{18}\,\text{s}$$

So, the age of the Universe is roughly 4.5×10^{18} s, or 14 billion years.

Support for the Big Bang theory comes from many other experimental evidences. One of these is worth mentioning here – the temperature of the Universe itself. The expansion of the Universe led to cooling; theories predicted the current temperature of the Universe should be about 2.7 K. Data collected and analysed from telescopes onboard satellites have shown that the peak intensity of the electromagnetic radiation coming from all directions of space occurs at a wavelength of about 1 mm (microwaves). The intensity against wavelength graph is similar to the ones shown in Figure 31.6. According to Wien's displacement law, this corresponds to a temperature of about 3 K.

Physics does make you think. Everything around us, including us, was created during the Big Bang; we could, therefore, suggest that we all have the same age!

Question

14 Use the information given in the table in Question 9 about the Sun to show that the current temperature of the Universe matches with microwaves of wavelength 1 mm at peak intensity.

REFLECTION

Without looking at your textbook, list all the laws from this chapter.

Draw a flow diagram to show how the radius of a star can be determined.

Use the internet to find the most distant object in the Universe and its recession speed.

What was the most important thing you learned personally when working through this chapter?

SUMMARY

Luminosity of a star is defined as the **total** radiant power it emits. Luminosity has the unit watts (W).

Standard candles are used to determine the distance of galaxies. A standard candle is an astronomical object (such as a Cepheid variable star) that has known luminosity.

Radiant flux intensity is defined as the radiant power transmitted normally through a surface per unit area. Radiant flux intensity has the units W m^{-2}. The radiant flux intensity F at a distance d from the centre of a star of luminosity L is given by the equation:

$$F = \frac{L}{4\pi d^2}$$

Wien's displacement law:

$\lambda_{max}T$ = constant

where T is the thermodynamic temperature of the object and λ_{max} is the wavelength at the peak intensity.

The Stefan-Boltzmann law:

$$L = 4\pi\sigma r^2 T^4$$

where σ is a constant known as the Stefan-Boltzmann constant, L is the luminosity of the object (star), r is the radius of the object and T is the surface thermodynamic temperature of the object.

Hubble's law:

The recession speed of a **galaxy** is directly proportional to its distance from us.

The equation for Hubble's law is: $v = H_0 d$

where v is the recession speed, d is the distance of the galaxy and H_0 is the Hubble constant.

Doppler redshift equation:

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$$

where λ is the wavelength of the electromagnetic waves from the source, $\Delta\lambda$ is the change in the wavelength, f is the frequency of the electromagnetic waves from the source, Δf is the change in frequency, v is the recession speed of the source and c is the speed of light in vacuum.

The Big Bang theory is a model of the creation of the Universe, from an extremely hot and dense state, and its subsequent evolution. The redshift of absorption (or emission) spectral lines from distant galaxies provides evidence for the Big Bang.

EXAM-STYLE QUESTIONS

1 Which statement is correct about radiant flux intensity? **[1]**

A It depends on the area of the measuring device.

B It is measured in W m^{-2}.

C It is the same as luminosity.

D It is the total radiant power emitted from a star.

2 A group of astronomers have determined the radiant flux intensity *F* from a star and its distance *d*. The percentage uncertainty in *F* is 1.2 % and the percentage uncertainty in *d* is 2.5 %.

What is the percentage uncertainty in the calculated value of the luminosity of the star? **[1]**

A 1.3 %

B 3.0 %

C 3.7 %

D 6.2 %

3 A particular emission spectral line is measured in the laboratory to have a frequency of 7.3×10^{14} Hz.

a Calculate the wavelength of this spectral line in the laboratory. **[1]**

b Calculate the observed wavelength of this same spectral line in the spectrum of a galaxy moving away from the Earth at a speed of:

i 11 Mm s^{-1} **[3]**

ii 7.0 % the speed of light. **[3]**

c The spectrum of all distant galaxies is redshifted. State and explain what you can deduce about the Universe. **[2]**

[Total: 9]

4 **a** State Hubble's law. **[1]**

b The recession speed *v* against distance *d* graph for some galaxies is shown.

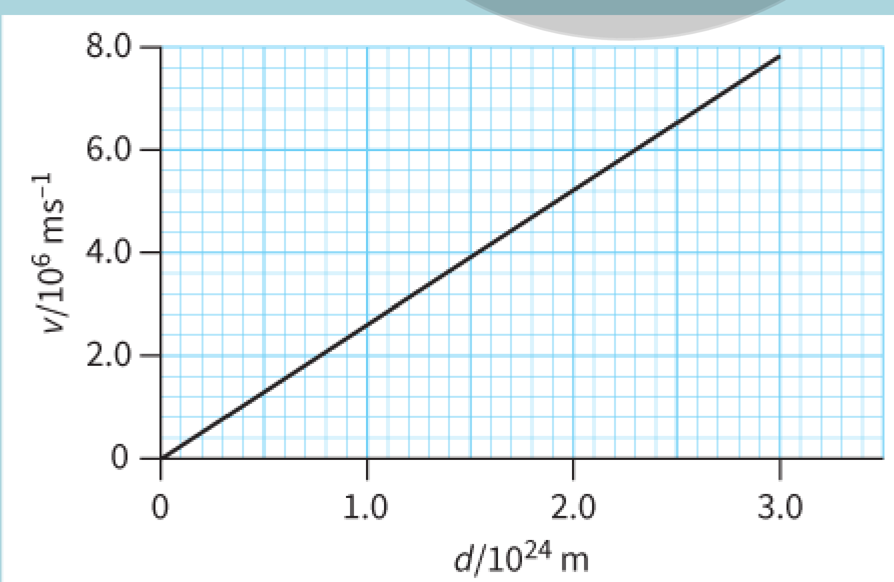


Figure 31.10

Determine the Hubble constant from this graph. Explain your answer. **[3]**

c The Big Bang occurred some 14 billion years ago.

$$1 \text{ year} \approx 3.15 \times 10^7 \text{ s}$$

Estimate the farthest distance we can observe. Explain your answer. **[3]**

[Total: 7]

5 a Define the luminosity of a star. **[1]**

b A red giant is a star bigger than our Sun. Explain how the surface of a red giant star can be cooler than the Sun, yet have a luminosity much greater than the Sun. **[2]**

c An astronomer has determined the surface temperature of a white dwarf star to be 7800 K and its radius as 8.5×10^6 m. Calculate the luminosity of this star. **[3]**

d The surface temperature T of a star depends on the wavelength λ_{max} at the peak intensity of the emitted radiation from the star.

The T against λ_{max} graph for a cluster of stars in our galaxy is shown.

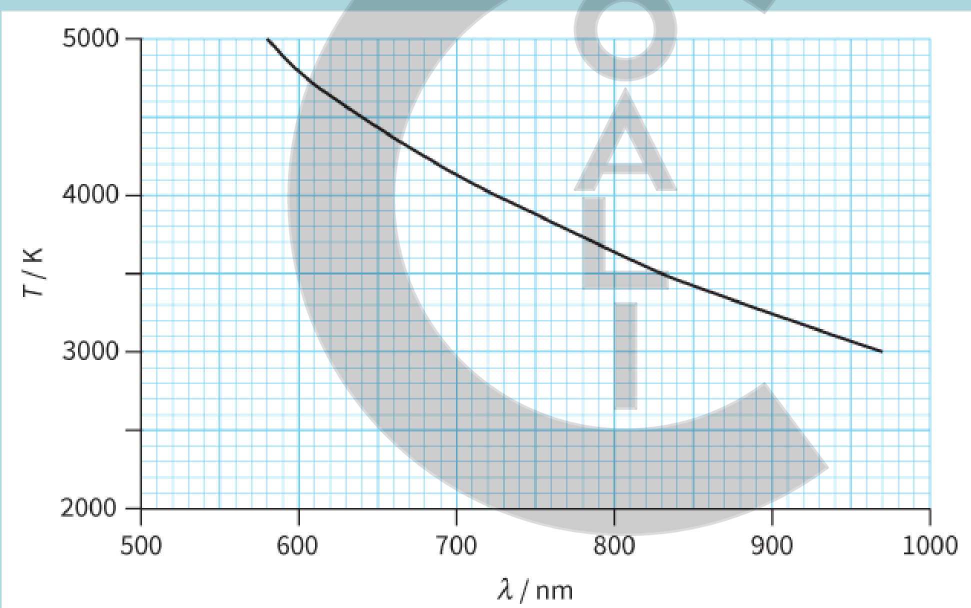


Figure 31.11

i Use the graph to show Wien's displacement law is obeyed. **[2]**

ii Estimate the surface temperature of a star with $\lambda_{max} = 400 \pm 10$ nm. In your answer, include the absolute uncertainty. **[4]**

[Total: 12]

6 Light from a galaxy is passed through a diffraction grating. The diagram shows part of the emission spectrum.

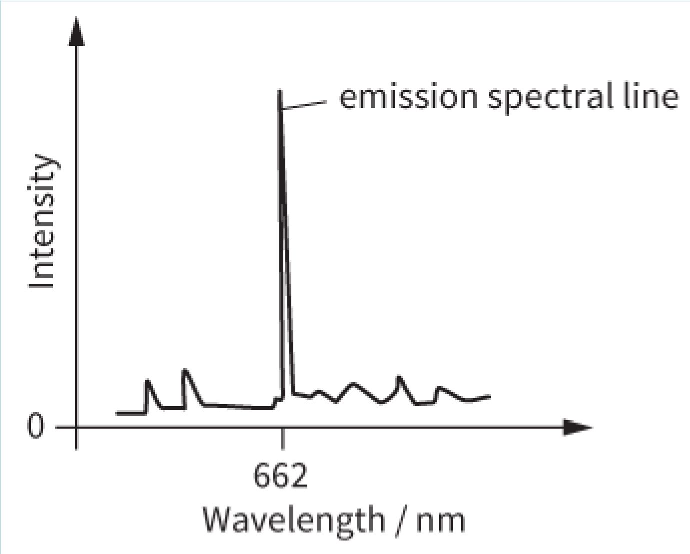


Figure 31.12

The strong emission spectral line has wavelength 662 nm.

a Calculate the energy of a photon of wavelength 662 nm. **[2]**

b Explain how a spectral line is produced by electrons within atoms. **[2]**

c In the laboratory, the same spectral line has wavelength 656 nm.

i Calculate the speed of the galaxy. **[3]**

ii State the direction of travel of the galaxy. **[1]**

d State and explain what the wavelength of the same spectral line would be for a much more distant galaxy. **[2]**

[Total: 10]

7 a Define radiant flux intensity. **[1]**

b State the relationship between radiant flux intensity *F* and distance *d* from the centre of a star. **[1]**

c Neptune is the farthest known planet from the Sun in the Solar System. Its distance from the Sun is 30 times greater than the distance of the Earth from the Sun. The radiant flux intensity from the Sun at the Earth is 1400 W m^{-2}.

A space probe is close to Neptune.

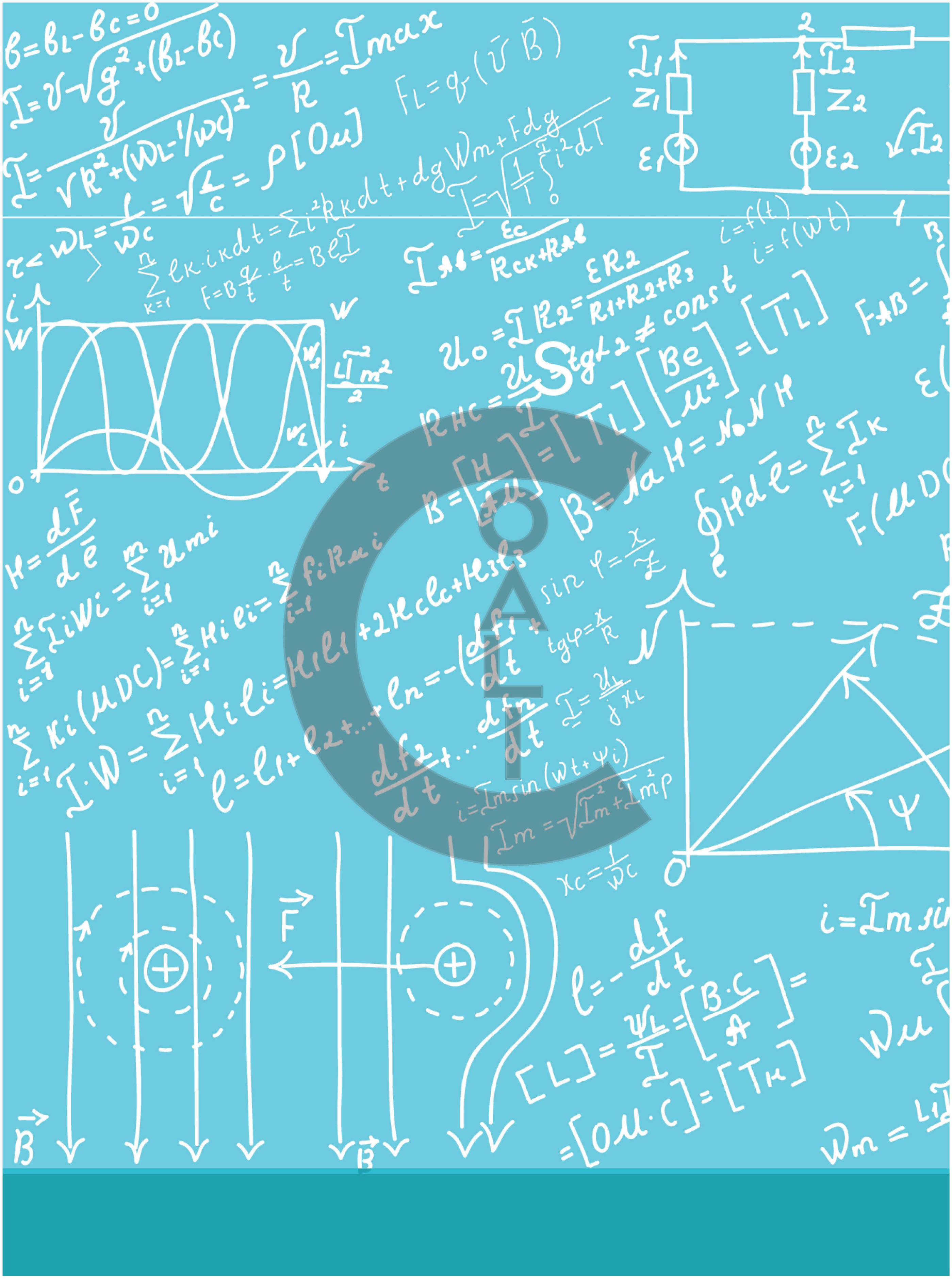
Calculate the maximum radiant power received by an instrument of cross-sectional area 1.0 cm^2 on this space probe. **[3]**

[Total: 5]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic...	Needs more work	Almost there	Ready to move on
understand the terms luminosity and radiant flux intensity	31.1, 31.2			
understand the inverse square law nature of radiant flux intensity	31.2			
understand the meaning of the standard candle	31.1			
use Wien's displacement law and the Stefan-Boltzmann law	31.3			
use $\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$ for receding source (stars and galaxies)	31.4			
use Hubble's law	31.4			
understand that redshift of spectral lines from galaxies provides evidence for the Big Bang theory.	31.4			



〉Chapter P2

Practical skills at A Level

LEARNING INTENTIONS

In this chapter you will learn how to:

- develop a systematic approach to carrying out experiments, including planning, setting up apparatus, investigating and recording results, analysing data and writing conclusions
- plan an investigation to test a relationship or investigate a problem, identifying the dependent, independent and control variables
- use logarithms and logarithmic graphs
- combine uncertainties, extending work from Practical Skills at AS Level
- plot error bars on graphs and find uncertainties in gradients and intercepts.

BEFORE YOU START

Do you know how to:

- estimate an uncertainty?
- tell the difference between systematic, mean and random errors?
- present data in a suitable table?
- draw a simple graph with suitable axes and labels?
- describe a simple experiment giving the steps logically one after the other?
- identify simple problems with an experiment and suggest changes to improve an experiment?
- use logarithms to base e and base 10?

P2.1 Planning and analysis

The practical work in the second year of your A level course builds on what you have covered in the first year. Tests and examinations you may take during your studies will ask you to demonstrate your abilities in two key areas:

- planning experiments
- analysis and evaluation of your results, including any conclusions you can draw.

Why do you think that experimental work is so important and has made such a difference to modern physics? What can you do to improve your experimental technique?

In this chapter, we will look at the different skills that you need to demonstrate your practical abilities.

P2.2 Planning

As you progress through your A level physics studies, you should think about and continually develop your approach to planning experiments. The experiments you will be asked to plan by your teacher will usually provide you with a scenario and sometimes a relationship or an equation that you are to use and test. Often, particular items of apparatus are mentioned and you should use these items, even if you think there is a better method. Sometimes, the experiment will seem familiar to you and sometimes it will be completely new. Before you start, it is important to read the scenario carefully. It is also important to read, understand and re-read any parts of a question you need to answer, before starting on your plan.

In producing your plan, you should draw a diagram showing the actual apparatus to be used, and pay particular attention to the:

- procedure to be followed
- measurements to be taken
- control of variables
- analysis of the data
- safety precautions to be taken.

Defining the problem: identifying the variables

It may seem obvious, but the first thing is to identify the problem. To do that you must identify the:

- **independent variable** in the experiment
- **dependent variable** in the experiment
- **control variables** (the quantities that are to be controlled or kept constant).

It is usually a good idea to start with a clear statement about the variables as the first part of your plan.

Here is an example of the sort of problem you might face in planning an experiment.

The deflection of a balloon by a jet of air is shown in Figure P2.1. You need to plan an investigation to show that $\tan\theta$ is inversely proportional to v^2, where θ is the angle between the ground and the string of the balloon and v is the speed of the air hitting the balloon. You are unlikely to have seen this experiment before, but this should not concern you.

In this example, the speed v of the air is the variable that you will need to alter and so this is the independent variable; the angle θ is the variable that changes as a result and so this is the dependent variable.

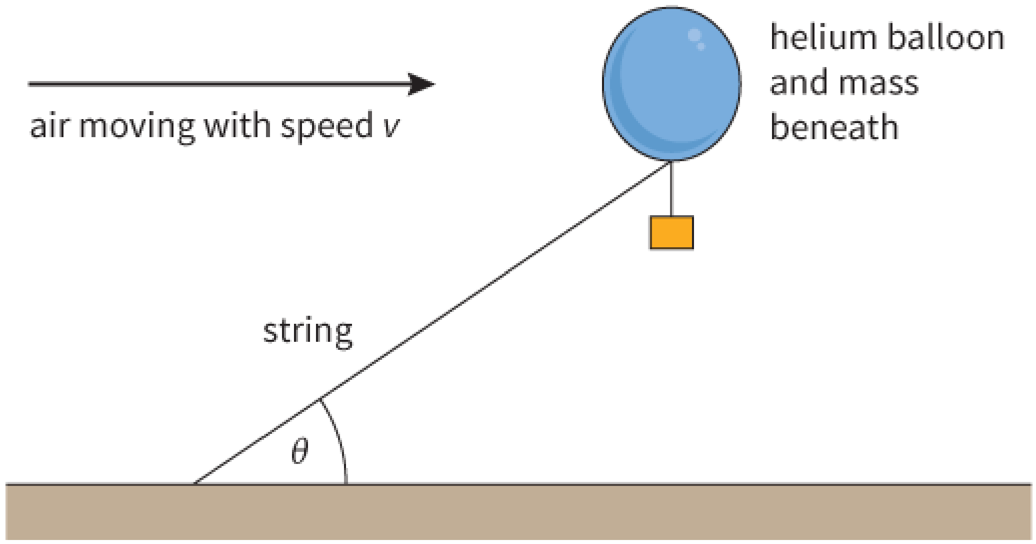


Figure P2.1: A balloon is deflected as the air moves at different speeds.

But what quantities are kept constant? These are the quantities that are controlled. You may be able to think of many, such as the total mass of the balloon and the mass placed underneath it. This total mass is certainly one quantity that should be kept constant, but it is not something that is likely to change during the course of the experiment. In terms of planning the experiment, you need to think about quantities that may easily change during the experiment if care is not taken. For example, you might realise that:

- the balloon may be deflected downwards, particularly if the air blows more strongly; then the air will hit the top, rather than the middle, of the balloon
- the balloon may warm up and expand in size; then more air will hit the balloon.

In either case, the experiment will then not just be testing the effect of the air speed. Your plan should clearly state what you need to keep constant. In our example:

- make sure that the jet of air is always horizontal and hits the middle of the balloon
- keep the temperature of the air inside the balloon constant.

As you can see, you need to think carefully about the experiment. Avoid giving wrong suggestions, for example, keeping the length of the string constant. If the string is longer the balloon may be out of the jet of air, and so it is not entirely a wrong suggestion, but it is not a primary quantity to be kept constant.

Question

1 An experiment is being planned to measure how the resistance of a wire depends on the cross-sectional area of the wire. What are the independent and dependent variables? Suggest three quantities that might be controlled.

Methods of data collection

The next task is to think about how you are going to carry the experiment out. Once you have a method in mind, you should be able to describe:

- the method to be used to vary the independent variable
- how the independent variable is to be measured
- how the dependent variable is to be measured
- how other variables are to be controlled

- the arrangement of apparatus for the experiment and the procedures to be followed, with the aid of a clear, labelled diagram.

It may be worthwhile jotting down your thoughts about the experiment on a rough piece of paper before you start, but do make sure that you write up all your points. It is particularly important to say what you actually measure and how the measurement is made. It may seem obvious to you that a particular quantity is measured, but unless you write it down it is not part of your plan.

Always check that in your account you have clearly said what you will:

- measure and how you will measure it
- change and how you will change it
- keep constant and how this is achieved.

Let's use the example of the balloon deflected by a current of air to show how you could approach this part of the plan.

Describing the experiment

First, describe how to change the independent variable and state what instrument is used to measure it. The apparatus shown in Figure P2.1 does not help very much and you must use your general knowledge and suggest, for example, that a wind fan be used. To change v will mean either changing the distance from the fan to the balloon or adjusting the power supply voltage to the fan.

You will also a need a wind speed indicator, sometimes called an anemometer, to measure the independent variable. Perhaps you have never seen or used a wind speed indicator, but clearly there must be an instrument to actually measure v. You may have to think very carefully to find a sensible instrument when the quantity is unusual.

The instrument to measure the dependent variable is much simpler – a protractor – although it may have to be a large protractor. Alternatively, you could use a ruler to measure the height h from the bench to the top of the string and the length l of the string, and then use $\sin\theta = \frac{h}{l}$ to find θ.

At this stage, try to suggest how to keep at least one of your 'controlled quantities' constant. For example, for the suggestions made earlier:

- compensate for deflection of the balloon downwards by a faster wind by lowering the fan, so that air from the fan stays horizontal and is always aimed at the centre of the balloon.
- keep the temperature of the air inside the balloon constant by leaving the balloon in a room with constant temperature for many hours before the experiment starts, and ensuring that the fan used blows air from the room.

As you can see, you have to think carefully about what happens during the experiment.

As you now have a clear idea of the experiment in your mind, draw a labelled diagram showing everything that you have mentioned. In this example, you could draw the fan, possibly its supply, a protractor, and even an anemometer.

Now describe your planned experiment, making sure that you describe a logical sequence of steps to follow. If you find this difficult, a labelled diagram of each step can sometimes be useful. For example, you might draw a diagram where you remove the balloon and put your wind speed measurement device in place of the balloon to measure v. Did you realise that the reading for v must be made exactly where the balloon was placed?

Additional details

It is also helpful to give additional details. In particular, make sure you suggest anything that needs to be done to ensure there is a large change in the dependent variable.

In the experiment with the balloon, you need a large change in θ as v changes. The readings would not be useful if θ was always very close to one value, for example, 90°. How can this be achieved?

Obviously, the largest air speed must be strong enough to cause a significant deflection. If the deflection is too small, then the mass under the balloon can be decreased; if it is too large, then the mass can be increased. It might be sensible to have the air speed as large as possible and adjust the mass under the balloon until θ is about 30°, and then check that θ varies from 30° to 90° as the fan is slowly moved further away. Of course, the mass under the balloon is then kept constant.

You might also think of any difficulties in carrying out the experiment. For example, draughts must be avoided and you must wait until the balloon has stopped swinging before taking a reading of θ.

Safety

It may seem strange, but you should always comment on safety when asked to carry out any experiment. In some situations, the risks may be unimportant, and it may be sufficient to mention simple ideas such as wearing goggles to protect the eyes when heating liquids or when handling stretched wires, using a safety screen, ensuring that the apparatus is stable and not easily knocked over, using a sand tray under heavy weights to make sure that weights do not fall on your foot and switching off currents when not in use so that wires do not overheat.

In our example with the balloon, keeping away from the rotating blades in the fan and wearing goggles to avoid air blowing into your eye should be sufficient. Do give some detail in your suggestions and do not just say 'use goggles'.

WORKED EXAMPLE

1 Plan an experiment to measure the resistivity ρ of glass, which is about 10^{10} Ω m. You have available a number of sheets of glass of the same size but with different thicknesses. Resistivity ρ is defined as $\rho = \frac{RA}{l}$

Step 1 Identify the variables.

- The independent variable is the thickness *l* of the glass.
- The dependent variable is the resistance *R* of the glass. Finding *R* involves measuring the p.d. across the glass and the current in the glass.
- The control variable is the area *A* of the glass. Since this is mentioned in the question, suggest also that the temperature must be constant.

Step 2 Describe the method of data collection in logical steps.

To alter the independent variable, use glass sheets of different thickness but the same area. The thickness of each piece of glass is measured with a micrometer at several places and averaged.

The area *A* is required. This can be found by measuring the length and breadth of each sheet with a rule and multiplying the values together.

Draw a circuit diagram of an ammeter in series with the glass sheet and a power supply, with a voltmeter across the glass. Connections are made to the large surfaces of the glass. This can be done using aluminium foil, or metal plates as in a capacitor, which closely touch each large face of the glass sheet. Use a diagram to show how this is done.

The logical steps are then to record ammeter and voltmeter readings with one thickness of glass. Then repeat the readings with different thicknesses, suggesting sensible thicknesses of glass, perhaps every mm from 1 mm to 10 mm. If you are going to perform the experiment these thicknesses may be available, but if you are merely planning the experiment then you must suggest sensible values.

Step 3 Add any additional details. How can you obtain reasonable values? Think about the size and thickness of the glass to be used and whether you can detect a reasonable change in the dependent variable, the resistance. You might, for example, suggest using a sheet of glass 1 m^2 in area and 1 mm thick. Its resistance is then:

$$R = \frac{\rho l}{A} = 10^{10} \times 0.001 = 10^{7}\,\Omega$$

Can this be measured with ordinary laboratory apparatus? What voltages and what meters are suitable? A voltage of 10 V produces a current of 1 μA, which is measurable, but 100 V gives a current of 10 μA, which may be easier to measure but more dangerous. With glass of thickness between 1 and 10 mm the current will be 1 to 10 μA and so the ammeter should measure from 1 to 10 μA or up to 10 μA.

As you can see, this means that you need some idea of the size of quantities that can be measured. In this example, you need to know what currents and voltages can be measured with ordinary laboratory equipment.

You may also give additional detail by describing how to attach the metal foil as contacts onto the large faces of the glass sheet with weights on top, or suggest that the glass be cleaned and dried.

Step 4 State any safety points. Glass can cut a person's skin and so gloves should be worn. If voltages above about 50 V are to be used, then use rubber gloves to avoid an electric shock or cover all exposed metal parts with insulation.

Step 5 Give your method of analysis. Remember, every derived quantity must be explained, so do not forget to state that for each thickness the voltage and current readings are used to find the resistance with the formula $R = \frac{V}{I}$.

Since $R = \frac{\rho l}{A}$, choose to plot a graph with *R* on the *y*-axis and *l* on the *x*-axis. The graph should be a straight line through the origin – a diagram may help here.

The gradient of the graph is $\frac{\rho}{A}$, so ρ = gradient × *A*.

Questions

2 What other graph can be plotted in Worked example 1, and how is its gradient used to find ρ?

3 A nail is placed with its sharp end just touching a piece of wood. When a mass falls with a velocity *v* and hits the nail, it drives the nail into the wood. It is suggested that the depth *d* that the nail moves into the wood is related to *v* by the equation $d = kv^2$, where *k* is a constant.

a Suggest:

i the independent, dependent and control variables

ii how the velocity of the falling mass can be measured as it hits the nail

iii sensible values for *d* and how they may be achieved and measured

iv the graph to be plotted and what it shows if the relationship is true.

b Write a logical step-by-step method to test the relationship.

P2.3 Analysis of the data

Whether you are dealing with data you have collected in an experiment, or data provided to you, you will need to analyse it. You need to describe how the data is used in order to reach a conclusion, and give details of any derived quantities that are calculated.

First, look carefully at the quantities in the relationship you have suggested (or at the formula that may be suggested when you are given an experiment to carry out). In our example with the balloon, $\tan\theta$ is inversely proportional to v^2, which means that the formula is $\tan\theta = \frac{k}{v^2}$ where k is a constant.

If possible, you should suggest plotting a graph that you know is a straight line if the equation is correct. In our example, since the equation for a straight line is $y = mx + c$, the y-axis of the graph should be $\tan\theta$ and the x-axis should be $\frac{1}{v^2}$.

You must clearly state:

- what is plotted on each axis of your graph
- that the relationship is valid if the graph gives a straight line through the origin.

You may prefer to draw a sketch graph to show what you mean, but always state clearly what type of graph you are going to use.

More complicated analysis of data

In P1 Practical skills at AS Level, we saw how to interpret equations of the form $y = mx + c$ and how to use a straight-line graph to find the constants m and c. However, you also need to be able to deal with quantities related by equations of the form $y = ax^n$ and $y = ae^{kx}$. For these, you need to be able to use logarithms (logs).

There are two common types of logarithm (see Chapter 20). The first type is sometimes called a natural logarithm, or a logarithm to base e, and is written as ln. The second type is a logarithm to base 10 and is written as lg. The ln type is more useful when dealing with an exponential formula such as e^{kx} but, otherwise, either type may be used. Look closely at any question to see which type is used. Do not mix the different types together in the answer to one question.

The unit of a quantity involving logarithms is specified in an unusual way. For example, the natural logarithm of a quantity s measured in metres is written as ln (s / m) and not as ln (s) / m or ln (s) / ln (m). You can see that the unit is written inside the bracket with the quantity.

You need to be able to take logarithms of equations of the form $y = ax^n$ and $y = ae^{kx}$. (Recall that an equation remains balanced if the same operation is performed on each side.)

Consider the equation: $y = ax^n$

Taking logarithms of both sides gives:

$$\lg y = \lg a + n \lg x$$

$$\ln y = \ln a + n \ln x$$

Now consider the equation: $y = ae^{kx}$

Taking logarithms of both sides gives:

$$\ln y = \ln a + kx$$

(To obtain these results, we have used the rules for logarithms that you ought to know:

log of a product	$\log(ab) = \log(a) + \log(b)$
log of a ratio	$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$

log of a power	$\log(a^n) = n \log(a)$

Questions

4 Calculate:

- **a** lg 10
- **b** ln 10
- **c** lg 100
- **d** lg 5
- **e** the antilogarithm to base 10 of 1 (i.e., find x where lg x = 1)
- **f** the antilogarithm to base e of 0.5 (i.e., find x where ln x = 0.5).

5 The number 48 = 3 × 2^4. Calculate lg 48 and lg 3 + 4 lg 2. Why are they the same?

Which graph to plot?

In handling data, our aim is usually to process the data to obtain a straight line graph. Then we can deduce quantities from the gradient and the intercepts. Table P2.1 shows graphs that can be plotted for different relationships, and the quantities that can be deduced from the graphs.

Relationship	Graph	Gradient	Intercept on *y*-axis	because …
$y = mx + c$	y against x	m	c	
$y = ax^n$	ln y against ln x lg y against lg x	n	ln a lg a	ln y = n ln x + ln a lg y = n lg x + lg a
$y = ae^{kx}$	ln y against x	k	ln a	ln y = kx + ln a

Table P2.1: Choice of axes for straight-line graphs.

A relationship of the form $y = ax^n$

A ball falls under gravity in the absence of air resistance. It falls a distance s in time t. The results are given in the first two columns of Table P2.2. A graph of distance fallen against time gives the curve shown in Figure P2.2.

Time t / s	Distance fallen s / m	ln (t / s)	ln (s / m)
0.20	0.20	−1.61	−1.61
0.40	0.78	−0.92	−0.25
0.60	1.76	−0.51	0.57
0.80	3.14	−0.22	1.14
1.00	4.90	0.00	1.59
1.20	7.05	0.18	1.95

Table P2.2: Results for a ball falling under gravity.

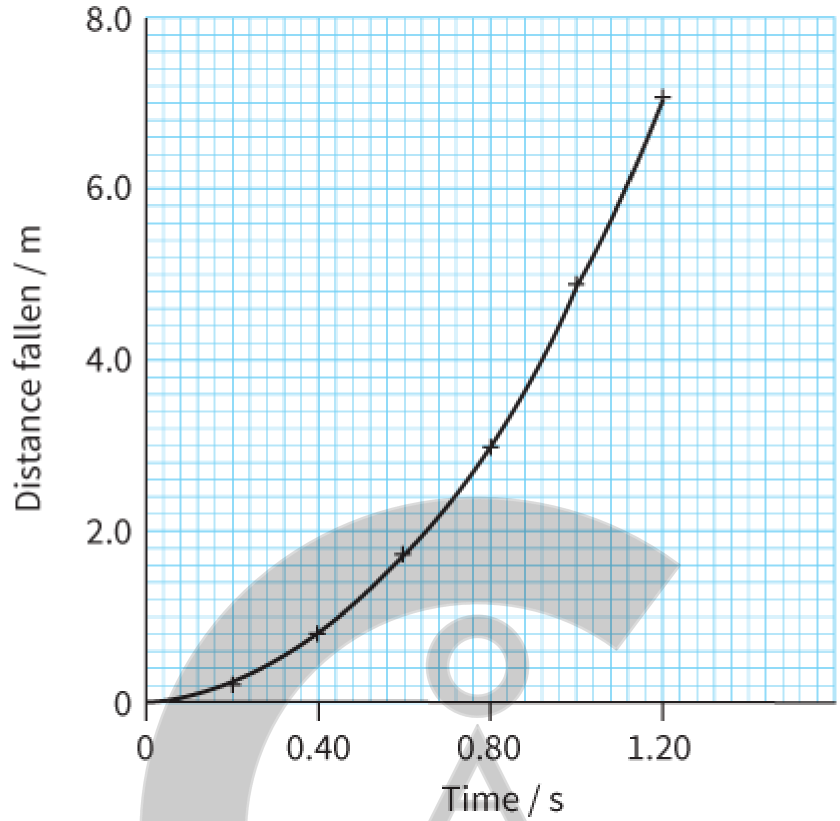


Figure P2.2: A distance–time graph plotted using the data in Table P2.2.

Because this is a curve, it tells us little about the relationship between the variables. If, however, we suspect that the relationship is of the form $y = ax^n$, we can test this idea by plotting a graph of ln *s* against ln *t* (a 'log–log plot'). Table P2.2 shows the values for ln *s* and ln *t*, and the resulting graph is shown in Figure P2.3. (Notice that here we are using natural logs, but we could equally well use logs to base 10.)

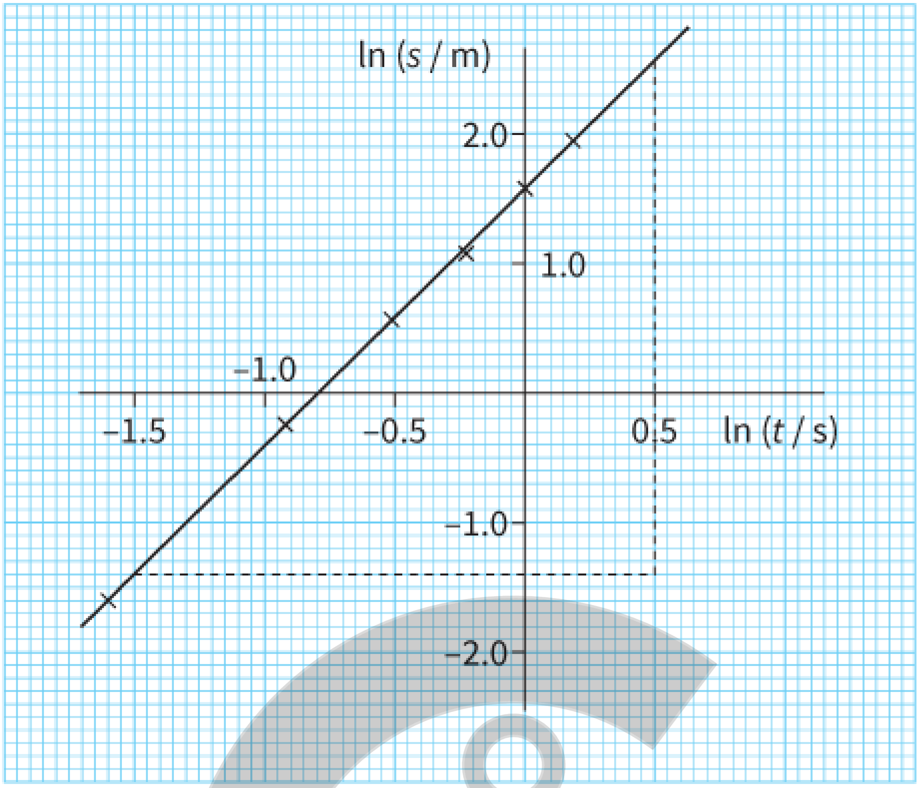


Figure P2.3: A log–log plot for the data shown in Table P2.2.

Because the graph is a straight line, the relationship must be of the form $y = ax^n$ But what are the values for *a* and *n*?

From the graph, the gradient is equal to the value of *n*, the power of *t*:

$$\begin{aligned} n &= \text{gradient} \\ &= \frac{(2.55-(-1.4))}{(0.5-(-1.5))} \\ &= \frac{3.95}{2.0} \\ &= 1.98 \approx 2.0 \end{aligned}$$

So the equation is of the form $s = at^2$. The intercept on the *y*-axis is equal to ln *a*, so:

$$\ln a = 1.6$$

By taking the antilogarithm we get:

$$a = 4.95 \text{ m s}^{-2} \approx 5.0 \text{ m s}^{-2}$$

If we think of the equation for free fall $= \frac{1}{2}gt^2$ the constant $a = \frac{1}{2}g$ But $g = 9.8 \text{ m s}^{-2}$, which is consistent with the value we get for our constant.

A relationship of the form $y = ae^{kx}$

A current flows from a charged capacitor when it is connected in a circuit with a resistor. The current decreases exponentially with time (the same pattern we see in radioactive decay).

Figure P2.4 shows the circuit and Table P2.3 shows typical values of current *I* and time *t* from such an experiment.

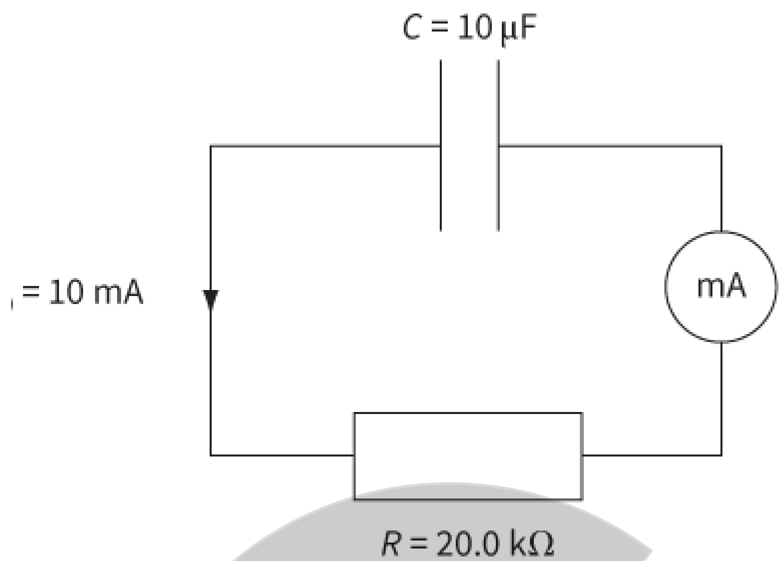


Figure P2.4: A circuit for investigating the discharge of a capacitor.

Current *I* / mA	Time *t* / s	ln (*I* / mA)
10.00	0.00	2.303
6.70	0.20	1.902
4.49	0.40	1.502
3.01	0.60	1.102
2.02	0.80	0.703
1.35	1.00	0.300

Table P2.3: Results from a capacitor discharge experiment.

The graph obtained from these results (Figure P2.5) shows a typical decay curve, but we cannot be sure that it is exponential. To show that the curve is of the form $I = I_0e^{kt}$, we plot ln *I* against *t* (a 'log-linear plot'). Values of ln *I* are included in Table P2.3. (Here, we must use logs to base e rather than to base 10.)

The graph of ln *I* against *t* is a straight line (Figure P2.6), confirming that the decrease in current follows an exponential pattern. The negative gradient shows exponential decay, rather than growth.

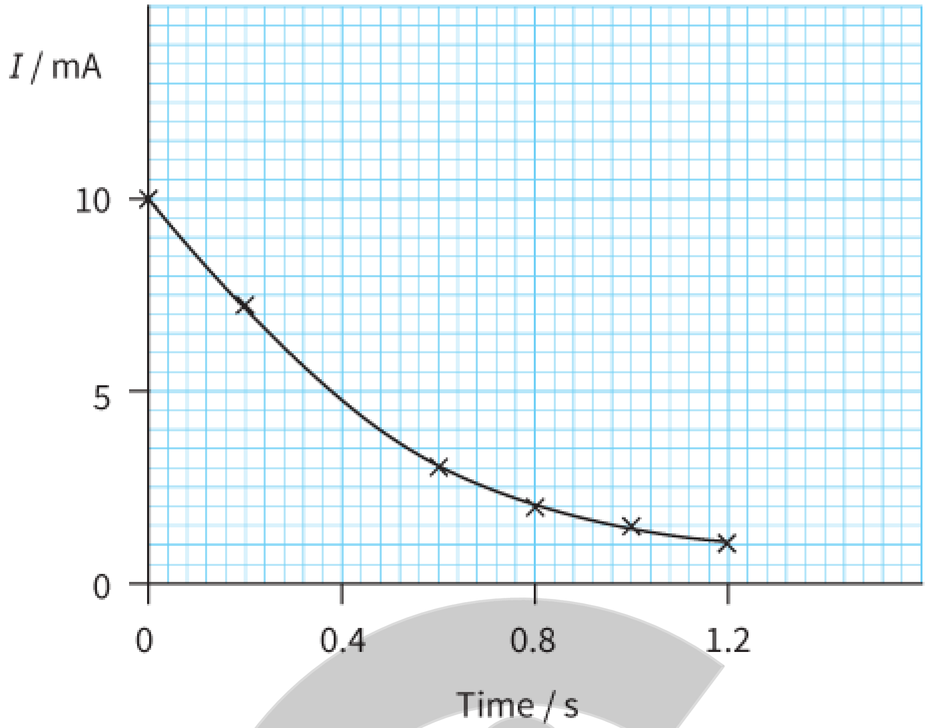


Figure P2.5:

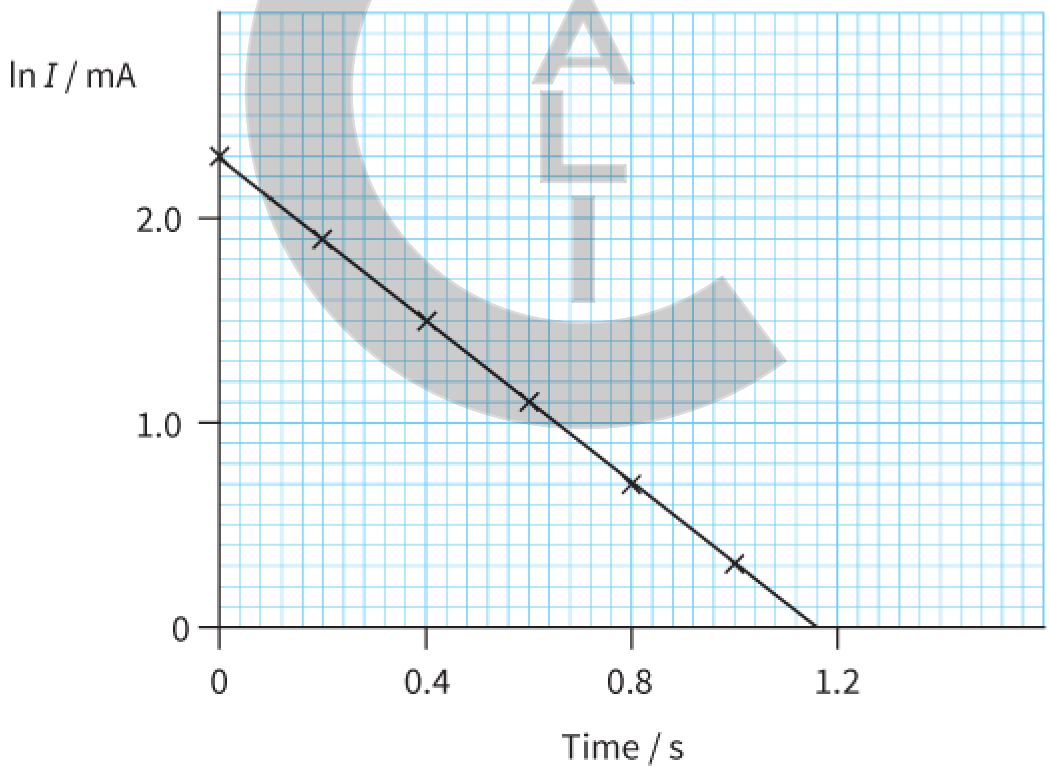


Figure P2.6:

The gradient of the graph gives us the value of the constant *k*:

$$k = \text{gradient} = \frac{(0-2.30)}{(1.16-0)} = -1.98\,s^{-1} \approx -2.0\,s^{-1}$$

From the graph, we can also see that the intercept on the *y*-axis has the value 2.30 and hence (taking the antilogarithm) we have $I_0 = 9.97 \approx 10$ mA. Hence, we can write an equation to represent the decreasing current as follows:

$$I = 10\,e^{-2.0t}$$

We could use this equation to calculate the current at any time *t*.

Questions

6 In the expressions that follow, *x* and *y* are variables in an experiment. All the other quantities in the expressions are constants.

In each case, state the graph you would plot to produce a straight line. Give the gradient of each line in terms of the constants in the expression.

a $y = kx^{\frac{3}{2}}$

b $y = cx^{q}$

c $m = \frac{8x}{By^2}$

d $y = y_0 e^{kx}$

e $R = \frac{(y-y_0)}{x^2}$

7 The period of oscillation *T* of a small spherical mass supported by a length *l* of thread is given by the expression:

$$T = 2\pi\sqrt{\frac{l}{g}}$$

where *g* is the acceleration due to gravity.

Design a laboratory experiment using this expression to determine the acceleration due to gravity. You should draw a diagram showing the arrangement of your equipment. In your account, you should pay particular attention to:

- the procedure to be followed
- the measurements to be taken
- analysis of the data to determine *g*
- any safety precautions that you would take.

P2.4 Treatment of uncertainties

All results should include an estimate of the absolute uncertainty. For example, when measuring the time for a runner to complete the 100 m you may express this as (12.1 ± 0.2) s. This can also be expressed as a percentage **uncertainty** (see Chapter P1); the percentage uncertainty is equal to $\frac{0.2}{12.1} \times 100\% = 1.655$ so we write the value as 12.1 s ± 1.7%, or even 12.1 s ± 2%.

Combining uncertainties

You should read through again the topic in Chapter P1 on combining uncertainties and also how to measure uncertainty if you are not sure.

Uncertainties and graphs

We can use error bars to show uncertainties on graphs. Table P2.4 shows results for an experiment on stretching a spring.

Load / N	Length of spring / cm	Extension / cm
0	12.4 ± 0.2	0.0
1.00	14.0 ± 0.2	1.6 ± 0.4
2.00	15.8 ± 0.2	3.4 ± 0.4
3.00	17.6 ± 0.2	5.2 ± 0.4
4.00	18.8 ± 0.2	6.4 ± 0.4
5.00	20.4 ± 0.2	8.0 ± 0.4

Table P2.4: Results from an experiment on stretching a spring.

When plotting the graph, the points are plotted as usual, and then they are extended to show the maximum and minimum likely values, as shown in Figure P2.7. Then the best fit line is drawn.

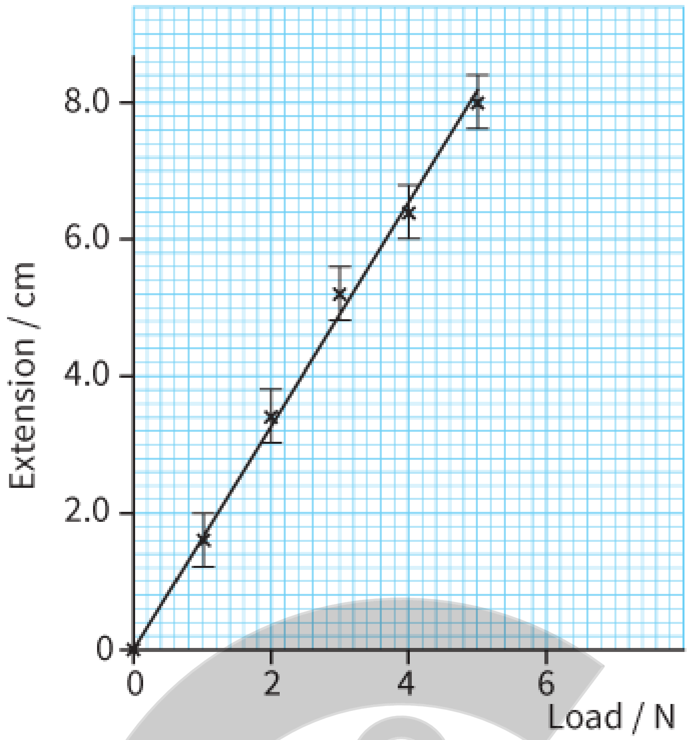


Figure P2.7: A graph representing the data in Table P2.4, with error bars and a line of best fit drawn.

To estimate the uncertainty in the gradient, we draw not only the best fit line but also a **worst acceptable line**, passing through the extremes in the error bars as shown in Figure P2.8.

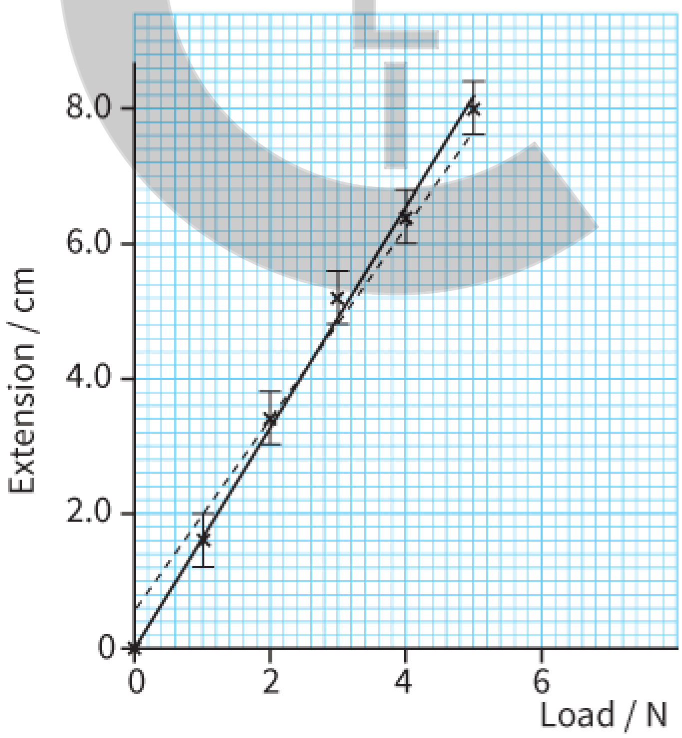


Figure P2.8: The same graph as in Figure P2.7, with a 'worst acceptable' line drawn (dashed).

The gradients for both best fit and worst fit lines are calculated and the uncertainty is the difference in their gradients:

uncertainty = (gradient of best fit line) − (gradient of worst acceptable line)

In our experiment, the gradients are:

line of best fit:

$$\begin{aligned} \text{gradient} &= \left(\frac{8.2-0}{5.0-0}\right)\text{cm N}^{-1} \\ &= 1.64\text{ cm N}^{-1} \approx 1.6\text{cm N}^{-1} \end{aligned}$$

line of worst fit:

$$\begin{aligned} \text{gradient} &= \left(\frac{7.6-0.6}{5.0-0}\right)\text{cm N}^{-1} \\ &= 1.4\text{ cm N}^{-1} \end{aligned}$$

So the uncertainty in the gradient = 1.6 − 1.4 = ± 0.2 cm N^{-1}

The gradient is therefore: 1.6 ± 0.2 cm N^{-1}.

Uncertainties and logarithms

When a log graph is used and we need to include error bars, we must find the logarithm of the measured value and the logarithm of either the largest or the smallest likely value. The uncertainty in the logarithm will be the difference between the two.

WORKED EXAMPLE

2 The resistance of a resistor is given as (47 ± 5) Ω. The value of ln (R / Ω) is to be plotted on a graph. Calculate the value and uncertainty in ln (R / Ω).

Step 1 Calculate the logarithm of the given value:

ln (R / Ω) = ln 47 = 3.85

Step 2 Calculate the logarithm of the maximum likely value:

maximum likely value = 47 + 5 = 52 Ω

ln 52 = 3.95

Step 3 The uncertainty is the difference between the two logarithms:

uncertainty in ln R = 3.95 − 3.85 = 0.10

Thus, ln (R / Ω) = 3.85 ± 0.10

Questions

8 The values of load shown in Table P2.4 are given without any indication of their uncertainties. Suggest a reason for this.

9 A student measures the radius r and the resistance R of several equal lengths of wire. The results are shown in Table P2.5. It is suggested that R and r are related by the equation:

$$R = ar^b$$

where a and b are constants.

a A graph is plotted with ln *R* on the *y*-axis and ln *r* on the *x*-axis. Express the gradient and *y*-intercept in terms of *a* and *b*.

b Values of *r* and *R* measured in an experiment are given in Table P2.5.

***r* / mm**	***R* / Ω**	**ln *r* / mm**	**ln *R* / Ω**
2.0 ± 0.1	175.0		
3.0 ± 0.1	77.8		
4.0 ± 0.1	43.8		
5.0 ± 0.1	28.0		
6.0 ± 0.1	19.4		

Table P2.5: Measurements for Question 9.

Copy and complete the table by calculating and recording values of ln (*r* / mm) and ln (*R* / Ω) and include the absolute uncertainties in ln (*r* / mm).

c Plot a graph of ln (*r* / mm) against ln (*R* / Ω). Include error bars for ln (*r* / mm).

d Draw the line of best fit and a worst acceptable straight line on your graph.

e Determine the gradient of the line of best fit. Include the uncertainty in your answer.

f Using your answer to part **e**, determine the value of *b*.

g Determine the value of *a* and its uncertainty.

P2.5 Conclusions and evaluation of results

In the previous experiment in P2.4, we can conclude that the extension/load for the spring in this example is (1.6 ± 0.2) cm N^{-1}. If a hypothesis is made that the extension is proportional to the load then there is enough evidence here for the conclusion to be supported, as a straight line can be drawn from the origin through all the error bars. If this is not possible then the hypothesis is not validated.

Now, suppose that the hypothesis is that the spring obeys Hooke's law and stretches by 5.0 cm when a load of 2.5 N is applied. The first part is validated for the reasons given. However, an extension of 5.0 cm for a load of 2.5 N gives a value of 2.0 cm N^{-1} for the gradient. This is clearly outside the range allowed for by the uncertainty in our measurements, and therefore the hypothesis is not supported.

REFLECTION

Make a checklist of all the important points that you are likely to forget or which led to you losing marks in any of the exercises.

SUMMARY

Planning includes:

- identifying variables that are independent, dependent and controlled
- the procedure to be followed, including a diagram, where appropriate, and the measurements to be taken
- how the measurements will be analysed, including the graph to be plotted and how the final result is calculated using the graph, for example, how the result is calculated from values of the gradient and the intercept of the graph
- extra detail, for example, how to obtain large changes in the dependent variable, an assessment of risk, a relevant safety precaution and how variables are kept constant.

Analysis of data includes:

- rearranging expressions including taking logarithms to obtain constants in expressions such as:

$$y = mx + c,\ y = ax^n \text{ and } y = ae^{kx}$$

- plotting graphs with error bars and calculating uncertainty by the difference between the gradients of the best fit and worst acceptable lines
- calculating derived quantities with correct units and appropriate number of significant figures.

EXAM-STYLE QUESTIONS

1 Each reading on a thermometer can be made with an uncertainty of ±0.5 °C. The thermometer is used to measure a temperature rise from 20 °C to 80 °C.

What is the percentage uncertainty in the measurement of this temperature rise? **[1]**

a 0.6%

b 0.8%

c 1.7%

d 2.5%

2 The period *T* and the length *l* of a simple pendulum are measured to be *T* = 1.5 ± 0.1 s and *l* = 0.560 ± 0.001 m.

The formula $T = 2\pi\sqrt{\frac{l}{g}}$ is used to find the acceleration of free fall *g*.

What is the best estimate of the uncertainty in the value of *g*? **[1]**

a 0.2%

b 3%

c 7%

d 13%

3 The volume of air inside a bottle affects its resonant frequency.

a What are the dependent and independent variables? **[1]**

b Suggest one quantity to be controlled. **[1]**

c How would you produce sounds of different frequency to show that the bottle resonates? **[1]**

d How would you find the frequency of the sound that makes the bottle resonate? **[1]**

e How would you find the volume of air inside the bottle? **[1]**

f How would you change the volume of air inside the bottle while keeping all other factors constant? **[1]**

g Suggest a safety precaution involving sound. **[1]**

[Total: 7]

4 The terminal velocity of an air bubble that rises in water is affected by the size of the bubble.

a What are the dependent and independent variables? **[1]**

b Suggest a quantity to be controlled. **[1]**

c How would you measure the terminal velocity of an air bubble that rises in water? **[1]**

d How would you generate bubbles of air of different sizes in water? **[1]**

[Total: 4]

5 The count rate from a radioactive source emitting γ-radiation is inversely proportional to the square of the distance from the source. Sources emitting γ-radiation also emit α- and β-radiation and are roughly spherical with a diameter of 2 cm.

a What are the dependent and independent variables? **[1]**

b Suggest a quantity to be controlled. **[1]**

c How could you make sure that only γ-radiation is detected? **[2]**

d How would you measure the count rate? Draw a diagram of the apparatus and explain how it is used. **[2]**

e How would you make the uncertainty in the count rate as small as possible? **[1]**

f Suggest one difficulty in measuring the distance and how this difficulty may be reduced. **[2]**

g Suggest a safety precaution. **[1]**

[Total: 10]

6 The size of a small toy balloon depends on atmospheric pressure.

a What are the dependent and independent variables? **[1]**

b Suggest a quantity to be controlled. **[1]**

c Draw a diagram of an apparatus to investigate the change in size of the balloon as atmospheric pressure changes. **[2]**

d State how the pressure is changed in your apparatus and how it is measured? **[2]**

e Suggest a safety precaution. **[1]**

[Total: 7]

7 Quantities *A* and *B* have the following values: $A = 3.0 \pm 0.2$ cm, $B = 2.0 \pm 0.1$ cm. Find the value of the following expressions and their absolute uncertainties.

a AB **[1]**

b $\frac{A}{B}$ **[1]**

c A^2 **[1]**

d $A - B$ **[1]**

e $A^2 - B^2$ **[1]**

f $\sqrt{A}$ **[1]**

[Total: 6]

8 Explain how you draw the best fit line and the worst fit line on a graph and how you find the uncertainty in the intercept on the *y*-axis. **[2]**

Questions 9 to 13 ask you to design an experiment based on the information given. All these questions have the same marking structure, with marks allocated to the different aspects as shown.

a the procedure to be followed **[3]**

b the measurements to be taken **[5]**

c the analysis of data **[2]**

d the safety precautions to be taken **[1]**

e additional detail. **[4]**

9 The resistance *R* of a light-dependent resistor (LDR) varies with the distance *d* from a very bright source of light. It is suggested that *R* and *d* are related by the formula $R = kd^n$, where *k* and *n* are constants. Design a laboratory experiment to test this relationship. The LDR has a resistance of 50 Ω in bright light and 200 kΩ in the dark. **[15]**

10 A ruler with a small mass at one end is clamped at the other end, as shown, and oscillates up and down when plucked by hand.

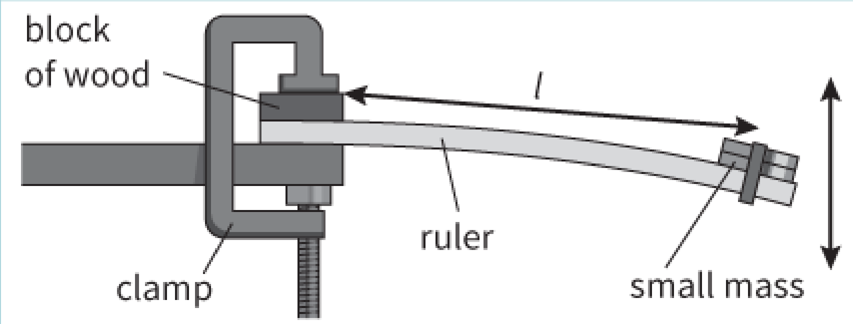


Figure P2.9

It is suggested that the period of oscillation T of the ruler is related to the length l by the relationship $T = kl^n$, where k and n are constants. Design a laboratory experiment to test this relationship and to find the value of n. **[15]**

11 A current-carrying coil produces a magnetic field. It is suggested that the magnetic field strength B at the centre of the coil is proportional to the current I in the coil. Design a laboratory experiment that uses a Hall probe to test this relationship. **[15]**

12 A bar magnet dropped into a coil induces an e.m.f. in the coil. It is suggested that E, the maximum induced e.m.f., is proportional to v, the speed of the magnet. Design a laboratory experiment to test this relationship. You might like to look at Figure 26.24 in Chapter 26. **[15]**

13 A student has a number of different transformers of varying numbers of turns. An alternating input current to the transformer induces an output e.m.f. It is suggested that the output e.m.f. V_s is directly proportional to the frequency f of the applied current. Design a laboratory experiment to test this relationship. **[15]**

14 The period T of a simple pendulum is related to its length l by the equation:

$$T = 2\pi\sqrt{\frac{l}{g}}$$

where g is the acceleration of free fall.

a A graph is plotted with T^2 on the y-axis and l on the x-axis. Express the gradient in terms of g. **[1]**

b A student measures the time t for 10 oscillations for different lengths l. This table shows her data.

l / m	t / s	T	T^2
0.300	11.1 ± 0.1		
0.400	12.8 ± 0.1		
0.500	14.2 ± 0.1		
0.600	15.8 ± 0.1		

l / m	*t* / s	*T*	T^2
0.700	16.9 ± 0.1		
0.800	18.1 ± 0.1		

Table P2.6

i Calculate and record values of *T* and T^2, including the absolute uncertainties in *T* and T^2. **[3]**

ii Plot a graph of T^2 / s^2 against *l* / m including error bars for T^2. **[2]**

iii Draw a straight line of best fit and a worst acceptable line on your graph. **[2]**

iv Determine the gradient of your line and include the uncertainty in your answer. **[2]**

v Use your value of the gradient to determine *g* and include the absolute uncertainty in your value. **[2]**

vi Using your value of *g* and its uncertainty, calculate the value of *t* when the length *l* is 0.900 m. Include the absolute uncertainty in your answer. **[2]**

[Total: 14]

15 Readings are taken of the resistance *R* of a thermistor at different temperatures *T.* It is suggested that the relationship between *R* and *T* is $R = kT^n$, where *k* and *n* are constants.

a A graph is plotted with lg *R* on the *y*-axis and lg *T* on the *x*-axis. State the value of the gradient and the *y*-intercept in terms of *k* and *n*. **[2]**

b Values for *T* and *R* are shown in this table. **[4]**

T / K	*R* / Ω	lg (*T* / K)	lg (*R* / Ω)
273	550 ± 10		
283	480 ± 10		
293	422 ± 10		
303	370 ± 10		
313	330 ± 10		

Table P2.7

Complete the table and include absolute uncertainties in lg (*R* / Ω).

c **i** Plot a graph of lg (*R* / Ω) against lg (*T* /K)*.* Include error bars. **[2]**

ii Draw the line of best fit and a worst acceptable line on your graph. **[2]**

iii Determine the gradient of your line of best fit and the uncertainty in your value. **[2]**

iv Determine the *y*-intercept of your graph (this is where the *x*-value, in this case, lg (*T* /K), is zero). Give the uncertainty in your value. **[2]**

v Determine values for *n* and *k* and the uncertainties in your answers. **[3]**

[Total: 17]

SELF-EVALUATION CHECKLIST

After studying the chapter, complete a table like this:

I can	See topic…	Needs more work	Almost there	Ready to move on
identify independent variables, dependent variables and variables to be kept constant	P2.2			
describe methods and procedures to vary, measure and keep variables constant	P2.2			
assess safety risks and describe precautions to reduce risk	P2.2			
understand the type of graphs to plot to produce straight lines for relationships of the form: $y = mx + c$, $y = ax^n$ and $y = ae^{kx}$	P2.3			
plot lines of best fit and a worst acceptable line on a graph including the use of error bars	P2.4			
convert absolute into percentage uncertainty and vice versa	P2.4			
calculate uncertainty estimates in derived quantities and in the gradient of a graph	P2.4			
express a quantity as a value, an uncertainty estimate and a unit.	P2.4			

Appendix 1: Physical quantities and units

Physical quantities have a numerical value and a unit. In physics, it is essential to give the units of physical quantities. For example, mass can be measured in kilograms. Hence you might write the mass of the trolley as:

mass of trolley = 0.76 kg

It would be a serious error to omit the unit kg at the end of the numerical value.

The scientific system of units is called the Système Internationale d'Unités (or SI system). The seven base units of this system are listed in Table 1. Each of the units is carefully defined, but the definitions need not concern us here.

All other units can be derived from the seven base units. For example:

- volume is measured in cubic metres (m^3)
- velocity is measured in metres per second ($m\ s^{-1}$)
- density is measured in kilograms per cubic metre ($kg\ m^{-3}$).

Physical quantity	Unit
mass	kilogram, kg
length	metre, m
time	second, s
temperature	kelvin, K
electric current	ampere, A
amount of substance	mole, mol
luminous intensity	candela, cd

Table 1: The seven base units of the SI system. (Note that you are not required to use the candela in this book.)

Prefixes

In physics, you will have to cope with very small and very large numbers. Numbers are written using powers of 10 to make them less awkward. This is known as scientific notation. Prefixes are used as an abbreviation for some of the powers of 10. For example, the height of a 5400 m high mountain may be written as either 5.4times 10^3 m or 5.4 km. The prefixes you will need most often are shown in Table 2.

Prefix	Symbol	Value
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}

Prefix	Symbol	Value
centi	c	10^{-2}
deci	d	10^{-1}
kilo	k	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}

Table 2: Some of the prefixes used in the SI system.

Estimation

When you carry out an experiment or a calculation, it is sensible to look at the answer that you get (and the results of intermediate calculations) to see if they seem reasonable. The only way you can know if an answer is absurd is if you are aware of some benchmarks. Some suggestions are given below. Try to add to this list as you go through your physics course.

mass of a person	70 kg
height of a person	2.0 m
walking speed	1 m s^{-1}
speed of a car on the motorway	30 m s^{-1}
volume of a can of drink	300 cm^3
density of water	1000 kg m^{-3}
weight of an apple	1 N
typical current in domestic appliance	13 A
e.m.f. of a car battery	12 V

Appendix 2: Data and formulae

Data

acceleration of free fall*	g	$9.81\ \text{m s}^{-2}$
speed of light in free space	c	$3.00 \times 10^{8}\ \text{m s}^{-1}$
elementary charge	e	$1.60 \times 10^{-19}\ \text{C}$
unified atomic mass unit		$1\ \text{u} = 1.661 \times 10^{-27}\ \text{kg}$
rest mass of proton	m_p	$1.67 \times 10^{-27}\ \text{kg}$
rest mass of electron	m_e	$9.11 \times 10^{-31}\ \text{kg}$
Avogadro constant	N_A	$6.02 \times 10^{23}\ \text{mol}^{-1}$
molar gas constant	R	$8.31\ \text{J K}^{-1}\ \text{mol}^{-1}$
Boltzmann constant	k	$1.38 \times 10^{-23}\ \text{J K}^{-1}$
gravitational constant	G	$6.67 \times 10^{-11}\ \text{N m}^{2}\ \text{kg}^{-2}$
permittivity of free space	ε_0	$8.85 \times 10^{-12}\ \text{C}_2\ \text{N}^{-1}\ \text{m}^{-2}\ (\text{F m}^{-1})$
Planck constant	h	$6.63 \times 10^{-34}\ \text{J s}$
Stefan-Boltzmann constant	σ	$5.67 \times 10^{-8}\ \text{W m}^{-2}\ \text{K}^{-4}$

*Note that this is the value of g that you should use in answering questions; g varies significantly over the Earth's surface, with values ranging from $9.78\ \text{m s}^{-2}$ at the equator to $9.83\ \text{m s}^{-2}$ at the poles.

Formulae

uniformly accelerated motion	$s = ut + \frac{1}{2}at^2$ $v^2 = u^2 + 2as$
hydrostatic pressure	$\Delta p = \Delta\rho g h$
upthrust	$F = \rho g V$
Doppler effect	$f_0 = \frac{f_s v}{v \pm v_s}$
electric current	$I = nAvq$
resistors in series	$R = R_1 + R_2 + \ldots$
resistors in parallel	$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots$
gravitational potential	$\varphi = \frac{-Gm}{r}$
gravitational potential energy	$E_p = -\frac{GMm}{r}$
pressure of an ideal gas	$p = \frac{1}{3}\frac{Nm}{V} < c^2 >$
simple harmonic motion	$a = -\omega^2 x$
velocity of a particle in s.h.m.	

	$v = v_0 \cos\omega t$
	$v = \pm\omega\sqrt{(x_0^2 - x^2)}$
electric potential	$V = \frac{Q}{4\pi\pm\varepsilon_0 r}$
electric potential energy	$E_p = \frac{Qq}{4\pi\varepsilon_0 r}$
capacitors in series	$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots$
capacitors in parallel	$C = C_1 + C_2 + \ldots$
discharge of a capacitor	$I = I_0 \exp^{-\frac{t}{RC}}$
Hall voltage	$V_H = \frac{BI}{ntq}$
alternating current or voltage	$x = x_0 \sin \omega t$
radioactive decay	$x = x_0 \exp^{(-\lambda t)}$ or $x = x_0 e^{-\lambda t}$
decay constant	$\lambda = \frac{0.693}{t_{\frac{1}{2}}}$
intensity reflection coefficient	$\frac{I_r}{I_0} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2}$
The Stefan-Boltzmann law	$L = 4\pi\sigma r^2 T^4$
Doppler redshift	$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta f}{f} \approx \frac{v}{c}$
work done on or by a gas	$W = p\Delta V$
energy of charged capacitor	$W = \frac{1}{2}QV$

Appendix 3: Mathematical equations and conversion factors

Conversion factors

electronvolt	1 eV = 1.60×10^{-19} J
Day	1 day = 8.64×10^{4} s
Year	1 year ≈ 3.16×10^{7} s
Light year	1 light year ≈ 9.5×10^{15} m

Mathematical equations

arc length = $r\theta$

circumference of circle = $2\pi r$

area of circle = πr^2

curved surface area of cylinder = $2\pi rh$

volume of cylinder = $\pi r^2 h$

surface area of a sphere = $4\pi r^2$

volume of sphere = $\frac{4}{3}\pi r^3$

Pythagoras' theorem: $a^2 = b^2 + c^2$

cosine rule: $a^2 = b^2 + c^2 - 2bc\cos A$

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

for small angle θ : $\sin\theta \approx \tan\theta \approx \theta$ and $\cos\theta \approx 1$

$\lg(AB) = \lg(A) + \lg(B)$

$\lg(\frac{A}{B}) = \lg(A) - \lg(B)$

$\ln(x^n) = n\ln(x)$

$\ln(e^{kx}) = kx$

Appendix 4: The Periodic Table

Key

atomic number
atomic symbol
name
relative atomic mass

Group																	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 H hydrogen 1.0																	2 He helium 4.0
3 Li lithium 6.9	4 Be beryllium 9.0											5 B boron 10.8	6 C carbon 12.0	7 N nitrogen 14.0	8 O oxygen 16.0	9 F fluorine 19.0	10 Ne neon 20.2
11 Na sodium 23.0	12 Mg magnesium 24.3											13 Al aluminium 27.0	14 Si silicon 28.1	15 P phosphorus 31.0	16 S sulphur 32.1	17 Cl chlorine 35.5	18 Ar argon 39.9
19 K potassium 39.1	20 Ca calcium 40.1	21 Sc scandium 45.0	22 Ti titanium 47.5	23 V vanadium 50.9	24 Cr chromium 52.0	25 Mn manganese 54.9	26 Fe iron 55.8	27 Co cobalt 58.9	28 Ni nickel 58.7	29 Cu copper 63.5	30 Zn zinc 65.4	31 Ga gallium 69.7	32 Ge germanium 72.6	33 As arsenic 74.9	34 Se selenium 79.0	35 Br bromine 79.9	36 Kr krypton 83.8
37 Rb rubidium 85.5	38 Sr strontium 87.6	39 Y yttrium 88.9	40 Zr zirconium 91.2	41 Nb niobium 92.9	42 Mo molybdenum 95.9	43 Tc technetium –	44 Ru ruthenium 101.1	45 Rh rhodium 102.9	46 Pd palladium 106.4	47 Ag silver 107.9	48 Cd cadmium 112.4	49 In indium 114.8	50 Sn tin 118.7	51 Sb antimony 121.8	52 Te tellurium 127.6	53 I iodine 126.9	54 Xe xenon 131.3
55 Cs caesium 132.9	56 Ba barium 137.3	57–71 La lanthanum	72 Hf hafnium 178.5	73 Ta tantalum 180.9	74 W tungsten 183.8	75 Re rhenium 186.2	76 Os osmium 190.2	77 Ir iridium 192.2	78 Pt platinum 195.1	79 Au gold 197.0	80 Hg mercury 200.6	81 Tl thallium 204.4	82 Pb lead 207.2	83 Bi bismuth 209.0	84 Po polonium –	85 At astatine –	86 Rn radon –
87 Fr francium –	88 Ra radium –	89–103 actinoids	104 Rf rutherfordium –	105 Db dubnium –	106 Sg seaborgium –	107 Bh bohrium –	108 Hs hassium –	109 Mt meitnerium –	110 Ds darmstadtium –	111 Rg roentgenium –	112 Cn copernicium –		114 Fl flerovium –		116 Lv livermorium –		

lanthanoids	57 La lanthanum 140	58 Ce cerium 140.1	59 Pr praseodymium 140.9	60 Nd neodymium 144.4	61 Pm promethium –	62 Sm samarium 150.4	63 Eu europium 152.0	64 Gd gadolinium 157.3	65 Tb terbium 158.9	66 Dy dysprosium 162.5	67 Ho holmium 164.9	68 Er erbium 167.3	69 Tm thulium 168.9	70 Yb ytterbium 173.1	71 Lu lutetium 175.0
actinoids	89 Ac actinium –	90 Th thorium 232.0	91 Pa protactinium 231.0	92 U uranium 238.0	93 Np neptunium –	94 Pu plutonium –	95 Am americium –	96 Cm curium –	97 Bk berkelium –	98 Cf californium –	99 Es einsteinium –	100 Fm fermium –	101 Md mendelevium –	102 No nobelium –	103 Lr lawrencium –

Acknowledgements

The authors and publishers acknowledge the following sources of copyright material and are grateful for the permissions granted. While every effort has been made, it has not always been possible to identify the sources of all the material used, or to trace all copyright holders. If any omissions are brought to our notice, we will be happy to include the appropriate acknowledgements on reprinting.

Thanks to the following for permission to reproduce images:

Cover Image Scanrail/Getty Images

Unit 1: Liufuyu/GI; Edward Kinsman/SPL; Lucas Oleniuk/Toronto Star/GI; **Unit 2:** Vladru/GI; Valery Sharifulin/GI; David Scharf/SPL; Science Source/GI; Cylonphoto/GI; Loren Winters, Visuals Unlimited/SPL; **Unit 3:** Comstock Images/GI; Erik Simonsen/GI; Tim Hughes/GI; David Madison/GI; Dzphotovideo/GI; Odd Andersen/GI; Fat camera/GI; **Unit 4:** Stockbyte/GI; Gaspr13/GI; Andrew Lambert Photography/SPL; Khalil Mazraawi/GI; Guy Cali/GI; **Unit 5:** 123ArtistImages/GI; George Steinmetz/GI; Bim/GI; Mike Clarke/GI; Anadolu Agency/GI; Kevin Morris/GI; Daniel Sambraus/GI; **Unit 6:** F9photos/GI; Trl Ltd/SPL/GI; Baona/GI; Andrew Lambert Photography/SPL(x3); Motoring Picture Library/Alamy; Ultra.f/GI; Science Photo Library/GI; **Unit 7:** Ser-y-star/GI; Ullstein bild/GI; B2M Productions/GI; **Unit 8:** Gyn9038/GI; Adam Hart-Davis/SPL(x4); Andrew Lambert Photography/SPL; Acilo/GI; **Unit 9:** Janaka Maharage Dharmasena/GI; Visual Field/GI; Age fotostock/Alamy; Marwood Jenkins/GI; Stephen Smith/GI; Andrew Lambert Photography/SPL; **Unit 10:** Dant-/GI; Kurita Kaku/GI; Philippe Huguen/GI; Nicholas Eveleigh/GI; **Unit 11:** Janaka Maharage Dharmasena/GI; Science Photo Library; Sheila Terry/SPL; **Unit 12:** EpicStockMedia/GI; Boris Horvat/GI; Doug Johnson/SPL; Filmfoto/GI; Sciencephotos/Alamy; Mgallar/GI; Onfokus/GI; Emilio Segre Visual Archives/American Institute of Physics/SPL; Rainervon Brandis/GI; **Unit 13:** Dhoxax/GI; Klaus Vedfelt/GI; Phil Ashley/GI; Sciencephotos/Alamy; Turk_Stock_Photographer/GI; Avalon/Bruce Coleman Inc/Alamy; Giphotostock/SPL; Zocha_K/GI; **Unit 14:** Goodshoot/GI; TongRo Images Inc/GI; Bettmann/GI; Tim Ridley/GI; Andrew Lambert Photography/SPL; Stickney Design/GI; **Unit 15:** H. Mark Weidman Photography/Alamy; Robert Gilhooly/Alamy; Prof. Peter Fowler/SPL; Jim Miller/GI; Dea/Biblioteca Ambrosiana/GI; Science & Society Picture Library/GI; Lawrence Berkeley Laboratory/SPL; Adam Hart-Davis/SPL; Janaka Maharage Dharmasena/GI; **Unit 16:** Ssuaphoto/GI; Massimo Bettiol/Stringer/GI; Elen11/GI; Gray Wall Studio/Shutterstock; **Unit 17:** Nasa/GI; NASA/Roger Ressmeyer/Corbis/VCG via Getty Images; CORBIS/Corbis via Getty Images; Nasa/GI; **Unit 18:** Argus/Shutterstock; S. Greg Panosian/GI; Andrew Lambert Photography/SPL(x2); Pgiam/GI; Alan Copson/GI; Andrew Lambert Photography/SPL; Fairfax Media/GI; Simon Fraser/GI; **Unit 19:** Janaka Maharage Dharmasena/GI; Martin Rietze/GI; Martyn F. Chillmaid/SPL; Andrew Lambert Photography/SPL; **Unit 20:** TongRo Images Inc/GI; Patrick Dumas/Eurelios/SPL; **Unit 21:** Miluxian/GI; Ralph H Wetmore II/GI; Andrew Lambert Photography/SPL(x2); **Unit 22:** Satori13/GI; Public Health England/SPL; **Unit 23:** Kilukilu/Shutterstock; Mark Williamson/GI; Yurazaga/GI; **Unit 24:** Rainer Plendl/GI; Alvarez/GI; Martyn F. Chillmaid/SPL; **Unit 25:** Serg_Aurora/GI; Omikron/SPL; Andrew Lambert Photography/SPL(x2); Mondadori Portfolio/GI; **Unit 26:** National Geographic Image Collection/Alamy; Sean Gallup/GI; Tim Wright/GI; **Unit 27:** Stockbyte/GI; Shinyfamily/GI; Science Photo Library-Adam Gault/GI; Andrew Lambert Photography/SPL; **Unit 28:** Tee Photolive/GI; bojanstory/GI; ST photography/Alamy; Paul Broadbent/Alamy; Volker Steger/SPL; John Thomas/SPL/GI; Science & Society Picture Library/GI; Andrew Lambert Photography/SPL; Professor Doctor Hannes Lichte, used by permission; Dr David Wexler, Coloured by Dr Jeremy Burgess/SPL; Dr Tim Evans/SPL; **Unit 29:** Ktsimage/GI; Natalie Board/EyeEm/GI; Andrew Lambert Photography/SPL; **Unit 30:** pixologicstudio/GI; Mauro Fermariello/SPL; AJ Photo/HOP Americain/SPL; Edward Kinsman/SPL; Zephyr/SPL; Sami Sarkis/GI; Michelle Del Guercio/SPL; Scott Camazine/SPL; BSIP/GI; Gustoimages/Science Photo Library/GI; Dr Najeeb Layyous/SPL; Science & Society Picture Library/GI; Skynesher/GI; **Unit 31:** Nasa/Handout/GI(x2); Alan Dyer/Stocktrek Images/GI; Balzs Ujhelyi/EyeEm/GI; Carlos Clarivan/SPL; Juliann/Shutterstock

Key: GI= Getty Images, SPL= Science Photo Library, Alamy = Alamy Stock Photo

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108859035

First edition 2010
Second edition 2014
Third edition 2020

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

A catalogue record for this publication is available from the British Library

ISBN 978-1-108-85903-5 Coursebook with Digital Access
ISBN 978-1-108-79652-1 Digital Coursebook
ISBN 978-1-108-79655-2 Coursebook – eBook

Additional resources for this publication at www.cambridge.org/9781108859035

Exam-style questions and sample answers have been written by the authors. In examinations, the way marks are awarded may be different. References to assessment and/or assessment preparation are the publisher's interpretation of the syllabus requirements and may not fully reflect the approach of Cambridge Assessment International Education.

Cambridge International recommends that teachers consider using a range of teaching and learning resources in preparing learners for assessment, based on their own professional judgement of their students' needs.